11th International Conference on Computer and Knowledge Engineering (ICCKE 2021), October 28-29, 2021, Ferdowsi University of Mashhad

ISUR: Iris Segmentation based on UNet and ResNet

Ali Khaki
Faculty of Computer and IT Engineering
Mazandaran University of Science and Technology
Babol, Iran
alikhaki@ustmb.ac.ir

Ali Aghagolzadeh
Faculty of Electrical and Computer Engineering
Babol Noshirvani University of Technology
Babol, Iran
aghagol@nit.ac.ir

Bagher Rahimpour Cami
Faculty member of Computer and IT Engineering
Mazandaran University of Science and Technology
Babol, Iran
b.rahimpour@ustmb.ac.ir

2021 11th International Conference on Computer Engineering and Knowledge (ICCKE) | 978-1-6654-0208-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCKE54056.2021.9721475

***Abstract*—Iris segmentation is an important research subject. There are many studies for iris segmentation. Deep learning-based methods have improved iris segmentation accuracy greatly. Now, the main challenge for researchers is to segment the noisy iris images captured in non-ideal environments created by visible light and user non-cooperation. This study proposes a robust method based on deep learning to segment noisy iris images called ISUR that includes two stages: detection and segmentation. First, an attention mask is created via a Convolutional Neural Network (CNN) based detector. Second, the eye image is segmented using a new CNN, and the obtained attention mask is used to focus on the iris region in this network. The proposed method has been tested on well-known iris datasets: UBIRIS and Iris Distance subset from CASIA, and has shown promising results compared to UNet.**

detection ve segmentation

***Keywords*—Iris Detection, Iris Segmentation, CNN, ResNet, Squeeze and Excitation Network**

ilk olarak CNN ile attention mask oluşturuldu.
Farklı bir CNN ile göz resimleri segmente edildi.
Çıkarılan attention mask iris bölgesine odaklandı.
Amaç:attention mask ile iris segmentasyonu gerçekleştirme

I. INTRODUCTION

Iris segmentation is a step of iris recognition system and plays a key role in this system. Its objective is to obtain the iris area. The error in this step will be transferred into other steps of the system and will reduce system performance. There are many studies for iris segmentation. In the recent years, the novel iris segmentation methods have been based more on deep learning with CNNs. Although these methods are robust and accurate but improving segmentation accuracy on noisy iris images is a main challenge for these methods. Therefore, this paper studies the accurate iris segmentation based on CNNs on noisy iris images.

This study proposes a novel and robust method based on deep learning with CNNs for better segmentation than the other methods and includes two stages: detection and segmentation. This method is focused on segmenting the iris area using an attention mask. ResNet-18 is also used in combination with the SE block as the backbone by the proposed method.

II. RELATED WORKS

Deep learning-based Iris segmentation methods segment the iris images using CNNs. Some studies of these methods are briefly explained in the following. Zhao and Kumar accurately detected and segmented the iris using Mask RCNN [1]. ResNet-101 is the backbone of this network. So, this network needs high memory. FCDNN network was proposed to segment

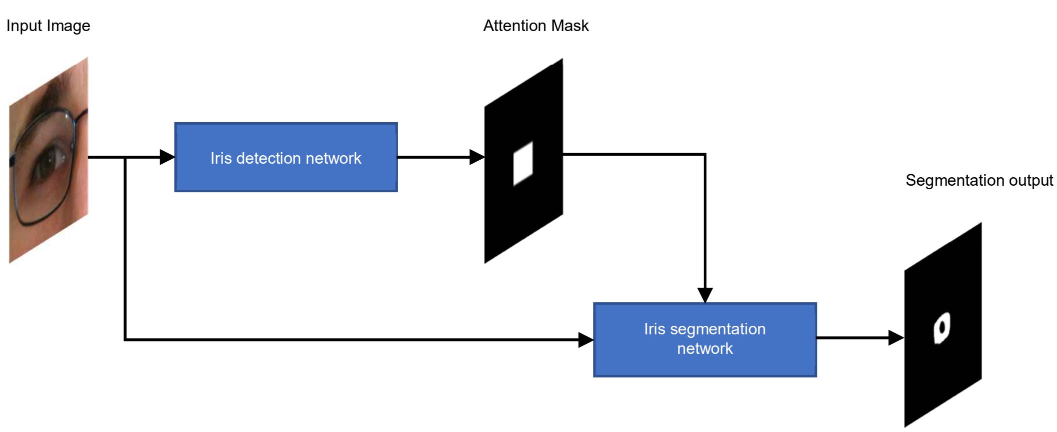


Fig. 1. The proposed method.

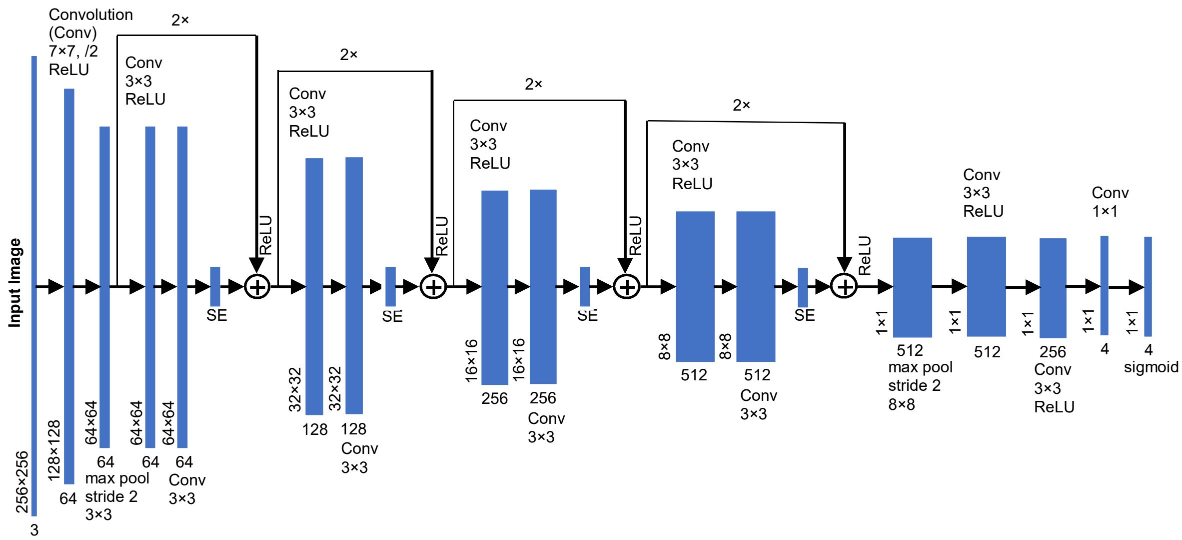


Fig. 2. IrisAttention.

the noisy iris images by Bazrafkan et al., and is a combination of four networks [2]. While the performance is improved but a vast number of parameters and a large model are produced. Reference [3] proposed a CNN based on SegNet called FRED-Net. Residual learning was used for improving the network. FRED-Net accurately segments the iris images. However, this model is large, and a vast number of training images is needed for increasing the network accuracy. Li et al. proposed a robust method based on UNet for iris segmentation, but this method uses a vast amount of memory [4]. IrisParseNet was developed from UNet for accurate iris segmentation [5]. This method also takes much memory. ATT-UNet was proposed for better iris segmentation using UNet and an attention mechanism[6]. This model is huge, and its training needs a long time (for large datasets). FD-UNet is proposed for accurate iris segmentation by Zhang et al., but it has many parameters and is large [7]. A fully convolutional network is used along with data augmentation by Varkarakis et al., which improved the accuracy, but the model is large and uses many images [8]. Although these methods have improved the iris segmentation accuracy greatly, there are still different challenges for increasing the accuracy on the noisy iris images.

III. CONTRIBUTION

This study proposes a CNNs-based robust method for the accurate segmentation called ISUR that is developed for noisy iris images captured in non-ideal environments created by visible light and user non-cooperation which is a serious challenge and the existing methods cannot segment accurately. This method consists of a detection stage and, then, segmentation. The following are the novelties of this study:

1. This method does not use any traditional image pre-processing algorithm.
2. The proposed method eliminates noise such as eyelids, eyelashes, and sclera, using an attention mask as much as possible.
3. This method is tested with UBIRIS and Iris Distance subset from CASIA, which these datasets include visible light and NIR light environments.

IV. PROPOSED METHOD

A. ResNet

ResNet is a CNN, which was proposed by He et al. [9]. This network solved the degradation problem by introducing the residual building block. This block is for the training of deep networks.

B. SE block

SE block is an architectural unit [10]. It improved performance for CNNs by selecting important feature maps and stopping mapping insignificant features [10]. This block is integrated with the residual building block, and a SE Residual block is created. In addition, residual building blocks from ResNet-18 is replaced with SE-Residual block and increased the network efficiency.

C. Method structure

The proposed method is built in two stages: detection and segmentation. Fig. 1 shows this method.

First, this method obtains the coordinates of the iris region (bounding box) using a detector based on a CNN and then produces an attention mask. In the next stage, this mask will be used in the segmentation network to focus on the iris region. Also, in this method, these two networks are trained separately. This method is coded with Python (3.7), Keras (2.1.0), and Tensorflow-gpu (1.14.0), and it is trained on an operating system of Windows 10, the GPU of 940MX (with 2GB of memory), and the CPU of Core i7 8550U with 8 GB RAM.

1) Iris detection

In this stage, the iris region is obtained by proposing a detector called irisAttention. irisAttention is a CNN that uses ResNet-18 as a backbone. As mentioned earlier, the SE block was integrated with residual building blocks to increase the ResNet-18 efficiency. Now ResNet-18 will be examined in more detail. The first layer of this network is a 7×7 convolution layer with stride 2 that followed by a ReLU function. In the next layer, this network uses a Max pooling (max pool) with pool size 3×3 and stride 2 for downsampling. Max pooling is followed by the four levels of SE-Residual blocks with 64, 128, 256, and 512 filters; each includes 2 SE-Residual blocks. All convolution layers at these levels are 3×3 and are followed by a ReLU. The first convolution layer of these levels has stride 2 except for the first level that the stride is 1. After the ResNet-18, the irisAttention network uses a 2×2 Max pooling (stride=8), and two 3×3 convolutions. These convolutions have 512 and 256 filters. Also, a ReLU is used after each convolution. In the end, this network utilizes a 1×1 convolution with 4 filters and a sigmoid function. Fig. 2 shows this network.

It is essential that the last Max Pooling reduces the feature map size to 1×1. Therefore, the image size does not matter; Max pooling with the appropriate stride (here is 8) should be used to reach the size 1×1 of the feature map. Also, four channels in the last layer are used for the coordinates of start and end points of the iris region.

There are many optimization methods to update the parameters of the model, but this model is trained with Adam optimizer. The learning rate is 0.0001 and 0.00001. Also, the network used to train a batch size of 2. The loss is computed using the Mean Square Error. A mask is created using the predicted bounding box for focusing on the iris region called the attention mask, which will be used in the next stage.

2) Iris segmentation

In this stage, a novel and robust CNN is proposed for iris segmentation that is modified and extended from UNet. The architecture, like UNet, includes a contracting expansive path that segments the iris image. On the contracting side, ResNet-18 is utilized to extract features as a backbone. Since this network was explained in the previous section in details, it is only mentioned that the 3×3 Max pooling is replaced with a 2×2 Max pooling. At the end of the contracting path, a Max pooling is placed with pool size 2×2 and stride 2 after the ResNet-18. This model is also divided into six levels on the expansive side. Each level uses an upsampling operation, and two 3x3 convolutions with stride 1, each followed by a ReLU. The channels are 512, 256, 128, 64, 64, and 32 in these levels, respectively. Also, all operations are padded. In the first four levels, a SE block is placed in each after the last convolution layer. Each SE is followed by a concatenation operation. This operation concatenates the feature maps in SE output and the output of the corresponding levels of the ResNet-18. At the fifth level, an element-wise multiplication is used after the last convolution layer. This operation multiplies the feature maps and the attention mask. In addition, there is a 2×2 Max pooling with stride 2 between the attention mask and the multiplication operation. After the sixth level, a concatenation operation is used, which combines together the feature maps at the end of this level, the fifth level before the multiplication operation, and the fourth level before the concatenation operation. Also, this concatenation uses a 1×1 convolution with 32 filters for decreasing the number of channels and an upsampling (size 4 for the fourth level and size 2 for the fifth level) for increasing the feature map size. At the end of the network, a 1×1 convolution is placed with the sigmoid function for mapping each 96-component feature vector to a class (binary). It is essential that each 3x3 convolution is followed by a batch normalization (BN). Fig. 3. shows the complete details of this network. Like irisAttention, Adam optimizer is chosen for training the network. The learning rate is 0.0001 and 0.00001. The model is trained using batch size 2.

The combination of Dice and BCE (Binary Cross Entropy) is utilized for computing the loss. Dice is a statistical method to examine the similarity between two sets. Also, BCE is a common method for segmentation tasks. Formulation of combined method is as follows:

$$Loss = (0.7 * BCE) + (0.3 * Dice) \qquad (1)$$

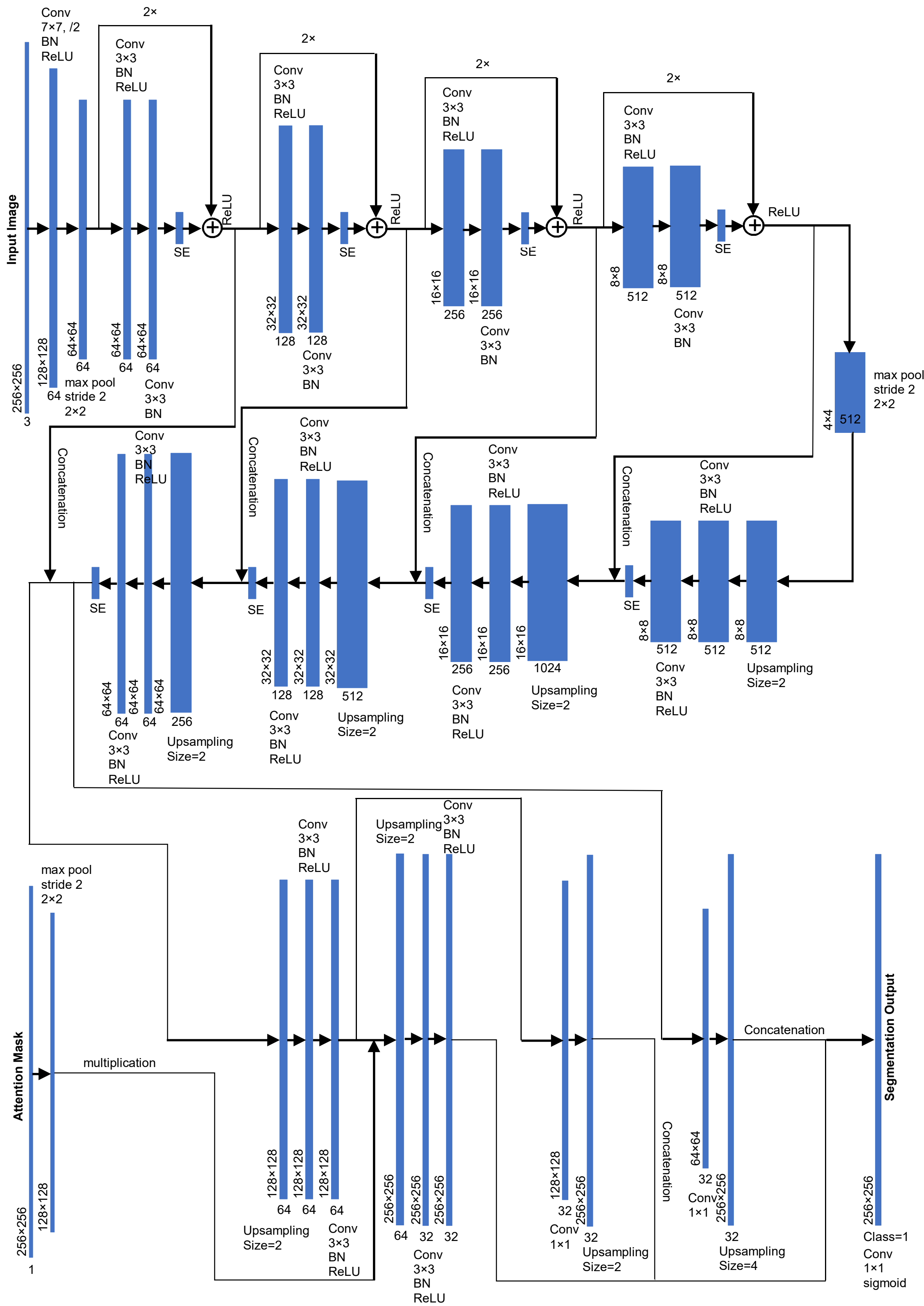
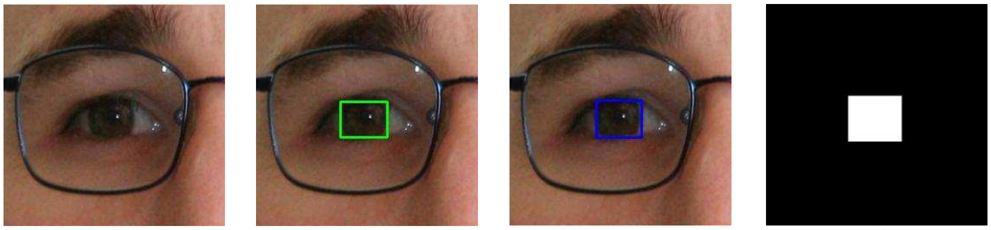


Fig. 3. The segmentation network architecture.



(a) (b) (c) (d)

Fig. 4. Sample result for the iris detection stage
(a) Eye Image (b) Ground truth (c) Predicted Bounding Box (d) Attention mask.

V. EXPERIMENTS

The proposed method will be evaluated in this section. For this purpose, experiments performed will be described in detail.

A. Dataset

There are many datasets for iris segmentation. The following datasets are chosen for evaluating the proposed method. The images are resized to 256×256 pixels. The bounding boxes required for iris detection are annotated by the authors. Bounding box annotation contains coordinates that have information of where exactly the iris region resides in the image. Also, the attention mask is produced by adding 5 pixels to the predicted coordinates. It is due to possible error.

- CASIA-Iris-Distance from CASIA Version 4

CASIA-Iris-Distance [11] includes 2,567 iris images from 142 subjects. The images were captured under near-infrared light illumination by a high-resolution camera. This subset was processed and labeled by the author [12]. The images of this subset were divided into two sets, training with 296 images and evaluation with 99 images. Also, the images are 480×640 pixels.

- UBIRIS Version 2

UBIRIS [13] includes 11,102 samples from 261 subjects that only 2,250 images are used in two sets, training with 1,687 images and evaluating with 563 images. The images were collected on non-ideal conditions by the camera of Canon EOS 5D. The images are 300x400 pixels. The color space is RGB. The required masks are provided by [14], [15].

B. Proposed method's accuracy

The proposed method is evaluated in the following. Here, the meaning of accuracy (error rate) is the segmentation accuracy, which the detection accuracy can also be checked.

1) Iris detection

The detection stage result is shown in this subsection. Fig. 4 shows the result from this stage.

2) Iris segmentation

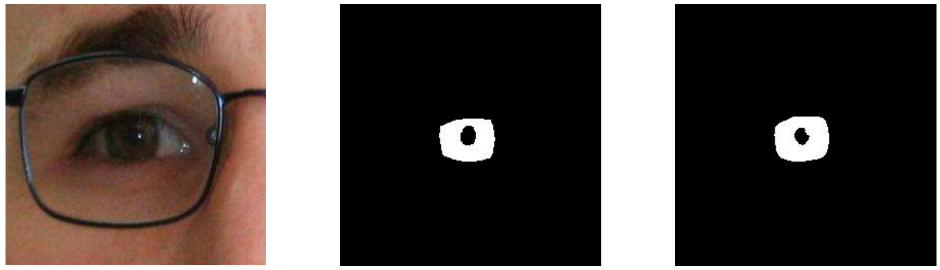
The results of the iris segmentation stage are provided using the proposed network. Fig. 5 and Fig. 6 show the sample of iris image segmented by the proposed network and UNet. UNet [16] is a CNN. This network is a robust and popular method for image segmentation, and its accuracy is high. Therefore, it was chosen for the comparison with our proposed network.

This network performs good segmentation on noisy iris images. The samples of extremely noisy images are selected from the UBIRIS dataset and are shown in Fig. 7.

NICE.I protocol [17] is a suitable protocol for evaluating the error rate on iris segmentation, which is

(a) (b) (c) (d)

Fig. 5. The predicted mask by the proposed network
(a) Eye image (b) Attention mask (c) Ground truth image (d) Predicted mask.



(a) (b) (c)

Fig. 6. The predicted mask by UNet
(a) Eye image (b) Ground truth image (c) Predicted mask.

explained in the following:

$$E = \frac{1}{N}\sum_{i=1}^{N} E_i \quad (2)$$

$$E_i = \frac{1}{W_i \times H_i}\sum_{x=1}^{W_i}\sum_{y=1}^{H_i} O_i[x,y] \oplus G_i[x,y] \quad (3)$$

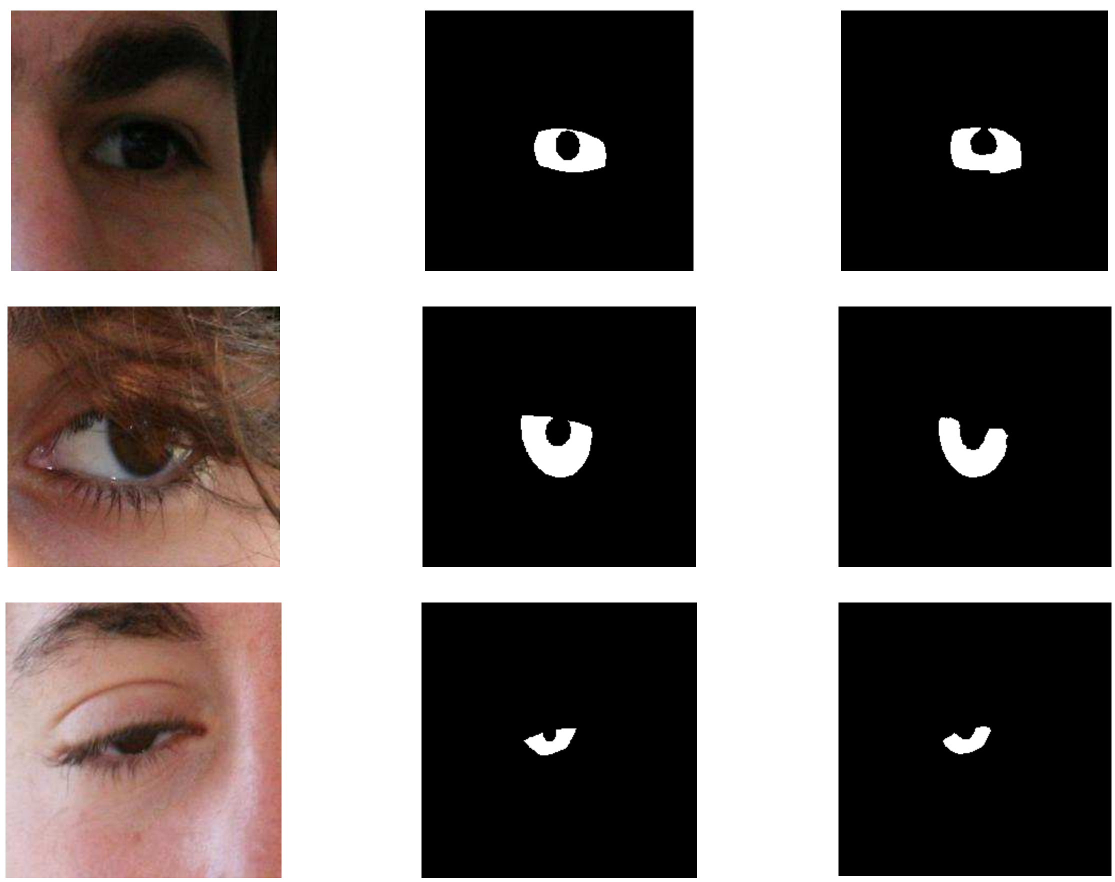
where G and O are denoted for the ground truth binary mask and the predicted mask. W and H are image sizes. $\oplus$ is used for the exclusive-or operation. N of Eq.2 is the number of samples.

The segmentation results on UBIRIS and CASIA-Iris-Distance datasets are shown in TABLE I.

TABLE I shows that the proposed network is able for improving the iris segmentation accuracy on the iris datasets with the visible light illumination and the near-infrared light illumination. The results show that the proposed network performs better.

VI. CONCLUSION

This study presented a robust method based on deep learning with CNNs called ISUR for accurate iris segmentation. In this method, a CNN-based detector is



(a) (b) (c)

Fig 7. Samples of extremely noisy images from UBIRIS dataset
(a) Eye images (b) Ground truth images (c) Predicted masks.

TABLE I
COMPARISON OF ERROR RATE OF THE PROPOSED NETWORK AND UNet

	Error Rate	
	UBIRIS	**CASIA-Iris-Distance**
UNet [16]	0.006764	0.005224
Proposed Network	0.006461	0.004778

used to focus on the iris region and eliminate noise, and finally, a new CNN is proposed using the attention mask for the iris segmentation, which in general was able to show promising performance.

REFERENCES

[1] Z. Zhao and A. Kumar, "A deep learning based unified framework to detect, segment and recognize irises using spatially corresponding features," *Pattern Recognit.*, vol. 93, pp. 546–557, Sep. 2019.

[2] S. Bazrafkan, S. Thavalengal, and P. Corcoran, "An end to end Deep Neural Network for iris segmentation in unconstrained scenarios," *Neural Networks*, vol. 106, pp. 79–95, Oct. 2018.

[3] M. Arsalan, D. S. Kim, M. B. Lee, M. Owais, and K. R. Park, "FRED-Net: Fully residual encoder–decoder network for accurate iris segmentation," *Expert Syst. Appl.*, vol. 122, pp. 217–241, 2019.

[4] Y.-H. H. Li, W. R. Putri, M. S. Aslam, and C.-C. C. Chang, "Robust Iris Segmentation Algorithm in Non-Cooperative Environments Using Interleaved Residual U-Net," *Sensors*, vol. 21, no. 4, pp. 1–21, 2021.

[5] C. Wang, J. Muhammad, Y. Wang, Z. He, and Z. Sun, "Towards Complete and Accurate Iris Segmentation Using Deep Multi-Task Attention Network for Non-Cooperative Iris Recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 2944–2959, 2020.

[6] S. Lian, Z. Luo, Z. Zhong, X. Lin, S. Su, and S. Li, "Attention guided U-Net for accurate iris segmentation," *J. Vis. Commun. Image Represent.*, vol. 56, no. October, pp. 296–304, 2018.

[7] W. Zhang, X. Lu, Y. Gu, Y. Liu, X. Meng, and J. Li, "A Robust Iris Segmentation Scheme Based on Improved U-Net," *IEEE Access*, vol. 7, pp. 85082–85089, 2019.

[8] V. Varkarakis, S. Bazrafkan, and P. Corcoran, "Deep neural network and data augmentation methodology for off-axis iris segmentation in wearable headsets," *Neural Networks*, vol. 121, pp. 101–121, 2020.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 7132–7141.

[11] "CASIA-Iris version 4 dataset." http://biometrics.idealtest.org/.

[12] 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 *2016 International Conference on Biometrics (ICB)*, 2016, pp. 1–8.

[13] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The {UBIRIS.v2}: A Database of Visible Wavelength Images Captured On-The-Move and At-A-Distance," *IEEE Trans. PAMI*, vol. 32, no. 8, pp. 1529–1535, Aug. 2010.

[14] H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, and A. Uhl, "A ground truth for Iris segmentation," *Proc. - Int. Conf. Pattern Recognit.*, pp. 527–532, 2014.

[15] F. Alonso-Fernandez, J. Bigun, H. Hofbauer, A. Uhl, and P. Wild, "E valuation of the IRISSEG datasets," no. January, 2014.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *CoRR*, vol. abs/1505.0, May 2015.

[17] "NICE.I - Noisy Iris Challenge Evaluation, Part I." http://nice1.di.ubi.pt/index.html (accessed Apr. 01, 2020).