

1-1-2020

## Iris Segmentation Using Interactive Deep Learning

Mousumi Sardar

*Indian Statistical Institute, Kolkata*

Subhashis Banerjee

*Indian Statistical Institute, Kolkata*

Sushmita Mitra

*Indian Statistical Institute, Kolkata*

Follow this and additional works at: <https://digitalcommons.isical.ac.in/journal-articles>

---

### Recommended Citation

Sardar, Mousumi; Banerjee, Subhashis; and Mitra, Sushmita, "Iris Segmentation Using Interactive Deep Learning" (2020). *Journal Articles*. 461.

<https://digitalcommons.isical.ac.in/journal-articles/461>

This Research Article is brought to you for free and open access by the Scholarly Publications at ISI Digital Commons. It has been accepted for inclusion in Journal Articles by an authorized administrator of ISI Digital Commons. For more information, please contact [ksatpathy@gmail.com](mailto:ksatpathy@gmail.com).

Received September 2, 2020, accepted October 26, 2020, date of publication December 1, 2020, date of current version December 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3041519

# Iris Segmentation Using Interactive Deep Learning

**MOUSUMI SARDAR<sup>ID</sup>, SUBHASHIS BANERJEE<sup>ID</sup>, (Student Member, IEEE),  
AND SUSHMITA MITRA<sup>ID</sup>, (Fellow, IEEE)**

Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700108, India

Corresponding author: Mousumi Sardar (mousumi.sardar02@gmail.com)

The work of Mousumi Sardar was supported by the University Grants Commission for providing financial assistance through the Rajiv Gandhi National Fellowship (RGNF) Scheme. This work was supported by the Intel Corporation for providing access to the Intel AI DevCloud.

**ABSTRACT** Automated iris segmentation is an important component of biometric identification. The role of artificial intelligence, particularly machine learning and deep learning, has been considerable in such automated delineation strategies. Although the use of deep learning is a promising approach in recent times, some of its challenges include its high computational requirement as well as availability of large annotated training data. In this scenario, interactive learning offers a cost-effective yet efficient alternative. We introduce an interactive variant of UNet for iris segmentation, including Squeeze Expand modules, to lower training time while improving storage efficiency through a reduction in the number of parameters involved. The interactive component helps in generating the ground truth for datasets having insufficient annotated samples. The effectiveness of the model ISqEUNet is illustrated through the use of three publicly available iris databases, along with comparisons involving existing state-of-the-art methodologies.

**INDEX TERMS** Active learning, biometrics, deep learning, fine tuning, iris segmentation.

## I. INTRODUCTION

Iris recognition is one of the most trusted approaches for automated biometric identification, and important for security and authentication systems. This is mainly due to the complexity, uniqueness and stability of the human iris. Inaccurate iris segmentation can cause failure in its recognition [1]; with error rates rising as inaccuracy in the segmentation task increases [2]. Segmentation plays a significant role in medical applications also.

Segmentation of iris images captured under ideal conditions constitutes a comparatively simpler image processing task [3], since the iris region demonstrates clear distinction between the sclera and pupil. In the unconstrained scenario, on the other hand, segmentation becomes more challenging as the acquisition of images no longer remains ideal due to factors like occlusion (caused by eyelids/eyelashes), poor (or overexposed) illumination, blurring, user noncooperation, difference in imaging equipments, etc.

In this scenario the accuracy of iris segmentation assumes utmost importance. Starting from a correct delineation of the iris region, one can proceed to extract valuable information

from the iris image to further improve upon the accuracy of the iris recognition system.

## II. LITERATURE REVIEW

Existing iris segmentation approaches can be roughly divided into three categories, based on their use of boundary, pixel or deep learning [4]. While the first kind localizes the iris region by determining the pupil and sclera boundaries relying on the contrast, the second approach classifies the pixels belonging to the iris region based on discriminative iris features. On the other hand, the deep learning approaches are similar to pixel-based methods but with improved performance and automated feature extraction.

Daugman's early research [5] assumed the iris region to be circular and bounded by the pupil and sclera on either side. An integro-differential operator searched over the image domain to detect the iris-sclera boundary, followed by the pupil-iris boundary. A gradient based edge detection was employed [6] to locate the two boundaries through Hough transform. Noise and occlusion from the eyelids and eyelashes were parameterized as parabolic arcs and located by a gradient-based edge detector. Active contour based methods [1], [7] considered iris boundaries to be non-circular, while handling occlusion and noise. Several other approaches in

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Oliva<sup>ID</sup>.

the boundary-based category include boundary fitting [8], illumination normalization with coarse iris localization [9] and reflection removal [10], to improve the accuracy of the segmentation.

The pixel-based algorithms build a classifier for detecting pixels of the iris region. A step-wise segmentation approach was adopted in Ref. [11] based on the image intensities. The eyelashes were partitioned from the input image using texture. Next the iris was delineated through gray scale information, followed by a post-processing step that utilized eye geometry to refine the output. Graph-cut energy minimization helped in optimally determining the eyelashes, pupil and iris. Zernike moment features were computed at different radii for classifying pixels from the iris region, with the help of support vector machine [12]. A random walker algorithm was developed [13] to efficiently estimate coarse segmentation of distantly acquired iris images, in a constrained environment. Post processing provided enhanced segmentation accuracy. A graph-based modeling of the segmentation mapped each pixel to a vertex (node), with the linkage between any two pixels corresponding to the edge between them. A multilayer perceptron, with a single hidden layer, was also used [14] for pixel classification to distinguish between the sclera and iris regions.

#### A. ROLE OF DEEP LEARNING

Traditional segmentation algorithms, based on image histograms, edges and other clustering techniques, can be simple and fast; but these typically require significant use-case specific tuning, with limited accuracy on complex scenarios. Both boundary- and pixel-based approaches require prior domain knowledge along with extensive pre- and post-processing. Moreover, their use of hand-crafted features lacks flexibility to search for optimal descriptors. In this scenario, deep learning-based methods have dramatically boosted the field of image segmentation [15], [16] by automatically learning the features. Convolutional Neural Networks [17] (CNNs or ConvNets) being one of the most commonly used deep learning models for classification [18], researchers in computer vision incorporated simple modifications to make them amenable to segmentation. The first layer in a CNN is always a Convolutional Layer which extracts the Low-Level features (edges, color, gradient orientation, etc.) from an input image. The Pooling layer reduces the spatial size of the convolved feature space to decrease the number of parameters. It also helps in extracting dominant features which are rotation and position invariant. The last fully connected layer generates the image class output labels.

Iris-related applications are typically sensitive due to the very dense and complex nature of the iris texture. Inspired by the success of CNNs, a few researchers focused on its application towards iris recognition and segmentation. Typically it consists of generating a binary mask to separate the pixels of the iris from those of the non-iris region. A segmentation network is trained on manually segmented annotations (ground truth), and evaluated on a separate

ground-truth augmented database. DeepIris [19] was used to solve intra-class variation of heterogeneous iris images. Relational features were extracted by a CNN to measure the similarity between two candidate irises during the verification process. A deep convolutional neural network (DCNN) based iris segmentation model proposed in [20] to extract highly irregular iris texture areas specifically in post-mortem iris images. A capsule network architecture with modified routing algorithm based on the dynamic routing between two capsule layers used for iris recognition [21].

A parameterization of the iris was created for CNN based segmentation [22], to bridge the gap between traditional CNN based segmentation and the rubbersheet-transform of the iris. This helped supplement the transformation of the iris from the polar to cartesian coordinates during normalization. IrisDenseNet [23] consists of a densely connected encoder and a SegNet decoder. While the network exhibited good performance, yet it was computationally intensive to train on large datasets due to its dense connectivity.

Two-stage CNNs were employed for iris segmentation [24], from images captured in the visible spectrum. A pre-trained VGG-face model was fine-tuned with transfer learning for finer adjustment of a rough iris boundary, as extracted by circular Hough transform. Transfer learning utilizes the knowledge learned from larger datasets towards solving different but related problems involving smaller data. DeepIrisNet [25] was designed for optimal iris representation and its cross-sensor recognition. It comprised of a large number of convolution/inception layers for handling large-scale iris data with complex distributions, entailing better utilization of computing resources.

Unlike CNN models, which use fully connected layer after the convolutional layer to get a fixed length feature vector, a fully convolutional network (FCN) [26] transforms the height and width of the intermediate layer feature map back to the size of the input image through the transposed convolution layer. This enables the prediction to have a one-to-one correspondence with the input image in the spatial dimension (height and width).

Accurate iris boundaries were identified [27], in non-cooperative environments, using hierarchical CNNs (HCNNs) and multi-scale FCNs (MFCNs). Iris segmentation on lower quality images, including augmentation, was performed using Fully convolutional deep neural network (FCDNN) [28].

However the segmentation results with FCNs still lacked perfection, as the feature maps for up-sampling were too coarse. Unlike FCN, which up-samples different sizes of coarse feature maps to the target resolution, the UNet [29] reformulates the up-sampling stage by including skip connections between the down-sampling and up-sampling paths. Hence the UNet generates more precise segmentation.

Although these CNN variants have demonstrated good performance for iris segmentation, they require millions of parameters and large volume of data for proper training. SqueezeNet [30], a smaller deep neural network, circumvents this problem with fewer parameters but with similar level of

accuracy. A SqueezeNet stacks a bunch of squeeze-expand (SqE) modules with a few pooling layers.

Typically CNNs do not generalize well to previously unseen object classes which may not have been present during training. Therefore labeled instances of each object class need to be available in the training set. Given that the expertise and time required to get correct annotations of all data is often very expensive, this severely constrains the performance of CNNs to segment objects. The concept of active learning, incorporating interactive refinement of intermediate results, helps overcome this problem. Active learning is a technique which enables learning from limited annotated examples by detecting, and asking the user to update, the most uncertain part(s) of the training data. In recent years interactive segmentation algorithms have become very popular, particularly in the field of medical image analysis [31], [32], by providing faster inference with improved storage efficiency.

### III. CONTRIBUTION

This article presents a novel interactive variant of UNet, incorporating SqE modules, for efficient iris segmentation. It enables boosting of speed while significantly improving the segmentation accuracy in the presence of limited annotated samples. Incorporation of active learning for interactive refinement is new for iris segmentation. The contribution of this research is summarized below.

- Reduction in number of parameters in the encoder part of the UNet by introducing the Squeeze-Expand (SqE) module, which replaces the existing sequence of  $3 \times 3$  convolution operations. Fewer number of trainable parameter in the SqE module enables reduction in training time with improved storage efficiency. Moreover, the less complex model helps avoid over-fitting during training of the relatively smaller iris datasets available.
- Interactive learning is incorporated to circumvent the problem of limited annotations in the publicly available iris datasets. The model is able to utilize image-specific information for robust handling of large context variations among different images. Interactive and automated generation of ground truth helps reduce time expense of experts while producing accurate segmentation.

The remaining part of the article is organized as follows. The UNet and SqueezeNet, used in this research, are described in Sec. IV-A and IV-B respectively. The proposed Interactive Squeeze-Expand UNet (ISqEUNet) model, involving active learning for interactive segmentation of the iris, is introduced in Sec. V. The experimental results on three iris datasets, viz. CASIA-IrisV4-Interval [33], IITD [34] and UBIRIS.v2 [35], demonstrating the superiority our model over existing state-of-the-art related literature, are provided in Sec. VI. Finally, Sec. VII concludes the article.

### IV. PRELIMINARIES

The main focus of this research is on designing a novel interactive deep learning methodology, which builds by incorporating the SqE module into the UNet for efficient segmentation of the iris. In this context it is imperative

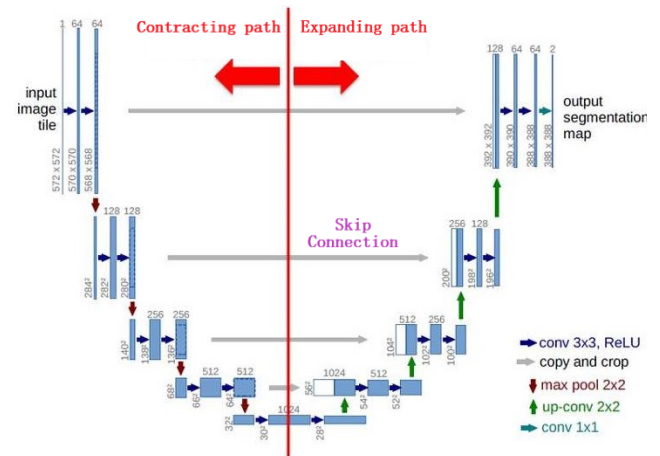


FIGURE 1. The UNet architecture [29].

to introduce the readers to the preliminaries of UNet and SqueezeNet.

#### A. UNet

The UNet was developed by Ronneberger *et al.* [29] for biomedical image segmentation. It has a CNN-like architecture for fast and precise segmentation of images. Compared to FCN, the two main differences are (1) the UNet is symmetric, and (2) the skip connections between the contracting and expanding paths apply the operator concatenation (instead of aggregation). This is illustrated in Fig. 1. The contracting path of UNet contains a series of convolution layers and pooling layers. The model learns global features by gradually reducing the feature map size and mean, while increasing the number of feature channels. The expanding path, on the other hand, contains a series of convolution and deconvolution operations. These up-sample the feature maps in incremental steps to the input size while reducing the feature channels. The skip connections supply additional local information to the up-sampling path to enable more accurate segmentation.

The Attention UNet (*AttUNet*) [4] employs “attention” in the framework of UNet for accurate iris segmentation. It regresses a bounding box of the potential iris region, followed by the generation of an attention mask to be used as a weighted function on the discriminative feature maps. This bounding box regression module consists of a pooling layer and a fully connected layer, added at the end the contracting path, and generates a rectangular coordinates based attention mask. Thereby the segmentation model is made to pay more attention to the iris region. Unlike conventional deep learning, which considers the whole eye image as input, here the attention component helps in estimating the position of the iris at the end of the contracting path with improved performance.

#### B. SqueezeNet

SqueezeNet [30] is a small CNN architecture which achieves accuracy of the level of AlexNet [18] on the ImageNet database with 50 times fewer parameters. The building block

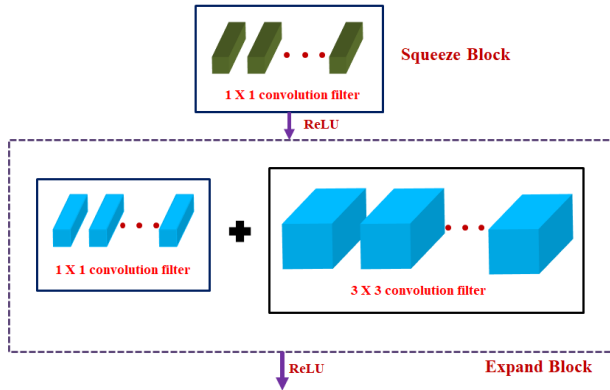


FIGURE 2. The squeeze expand module.

of SqueezeNet consists of a squeeze layer with an  $1 \times 1$  convolution layer, followed by an expand layer that has a combination of  $1 \times 1$  and  $3 \times 3$  convolution layers. A concatenation operation is performed in the expand layer to combine those two convolution layers. This is depicted in Fig. 2. Note that a squeeze layer with only  $1 \times 1$  filters, although having 9 times less parameters as compared to the typical  $3 \times 3$  filters of CNNs, can still function as a fully-connected layer working on feature points at the same position. The main objective of using the squeeze layer is to reduce the depth of feature maps; since it is often very time consuming to multiply volumes having extremely large depths. The total number of parameters in the  $3 \times 3$  convolutional layer is  $(input\_depth) \times (number\_of\_filters) \times (3 \times 3)$ .

In order to maintain a small number of parameters, the number of  $3 \times 3$  filters need to be reduced along with the depth of the input volume. As the number of input channels to these filters is reduced in the squeeze layer, it results in fewer computations at the expand layer. The squeeze layer and expand layer work on the same feature map size, with the former reducing the depth and the latter increasing it.

## V. ISqEUNet FOR IRIS SEGMENTATION

Here we describe the proposed Interactive Squeeze Expand UNet (ISqEUNet) for addressing the task of iris segmentation.

### A. NETWORK ARCHITECTURE

The overall structure of the proposed deep network is depicted in Fig. 3. It has both encoding and decoding paths. The contracting path extracts higher level features through a repeated use of the Squeeze-Expand module of Fig. 2. The number of kernels gradually increases to enable the architecture effectively learn the complex structures.

The expanding path consists of an up-sampling of the feature map, followed by a  $2 \times 2$  up-convolution that halves the number of feature channels, a concatenation with the corresponding cropped feature map from the contracting path, and two  $3 \times 3$  convolutions. The feedback connections between the encoding and decoding paths, concatenating features from both, enables the model to simultaneously utilize both local and global information. An  $1 \times 1$  convolutional layer is

adopted at the final layer in order to map the feature vector to the desired number of classes. The expanding path decodes the feature map to reconstruct the output segmentation mask. The binary cross entropy loss between predicted output ( $P$ ) and its corresponding ground truth ( $G$ ) is calculated as

$$L_B(P, G) = -\frac{1}{n} \sum_{i=1}^n g_k \log p_k + (1 - g_k) \log(1 - p_k), \quad (1)$$

where  $n$  represents the total number of pixels in an image, with  $g_k$  and  $p_k$  indicating the  $k$ th pixel values in  $G$  and  $P$  respectively.

The Squeeze-Expand (SqE) module is introduced in the contracting path of the UNet by replacing the series of  $3 \times 3$  convolution operations. This improves accuracy, as compared to the conventional UNet, while also reducing computation time with fewer number of trainable parameters. For example, with an image size of  $256 \times 256 \times 3$ , the number of parameters required to train the original UNet model was  $3.1 \times 10^7$ . In our proposed ISqEUNet this reduces to  $1.6 \times 10^7$  parameters. A structural comparison of the contracting path of *UNet*, *AttUNet*, and *ISqEUNet* is provided in Table 1 to highlight this.

The ISqEUNet combines the location information from the down-sampling path with the contextual information in the up-sampling path to generate a superposition of local and global information, in order to efficiently predict a good segmentation map.

### B. INCORPORATING INTERACTIVE LEARNING

Interactive learning allows the algorithm to actively query the user to obtain the desired output for the unknown test samples.

Interactive segmentation integrates the user's domain knowledge with the application-requirements for more robust performance. Here we introduce a novel interactive image-specific fine tuning strategy to circumvent, to some extent, the requirement of large number of annotated images during training. It is particularly suitable in segmentation of images when the availability of annotations is scarce.

Unlike Ref. [36], which uses bounding box and scribble-based segmentation with CNNs, we incorporate the SqE module into the conventional UNet structure with active learning. Architecture of the ISqEUNet model is depicted in Fig. 3. The details of interactive learning is outlined as Algorithm 1.

The pre-trained ISqEUNet was presented with a smaller dataset containing few annotated images, and the corresponding predicted segmentation obtained at output. The images which are incorrectly predicted, get refined manually by the user in an interactive mode.

Next the user-refined images are used as additional ground truth (in lieu of the corresponding incorrectly predicted images), and the model fine-tuned through function call to Algorithm 2 for updating weights. Depending on the value of the threshold  $c$  the algorithm decides whether to proceed with further fine tuning, using a *Flag* to mark user interaction.



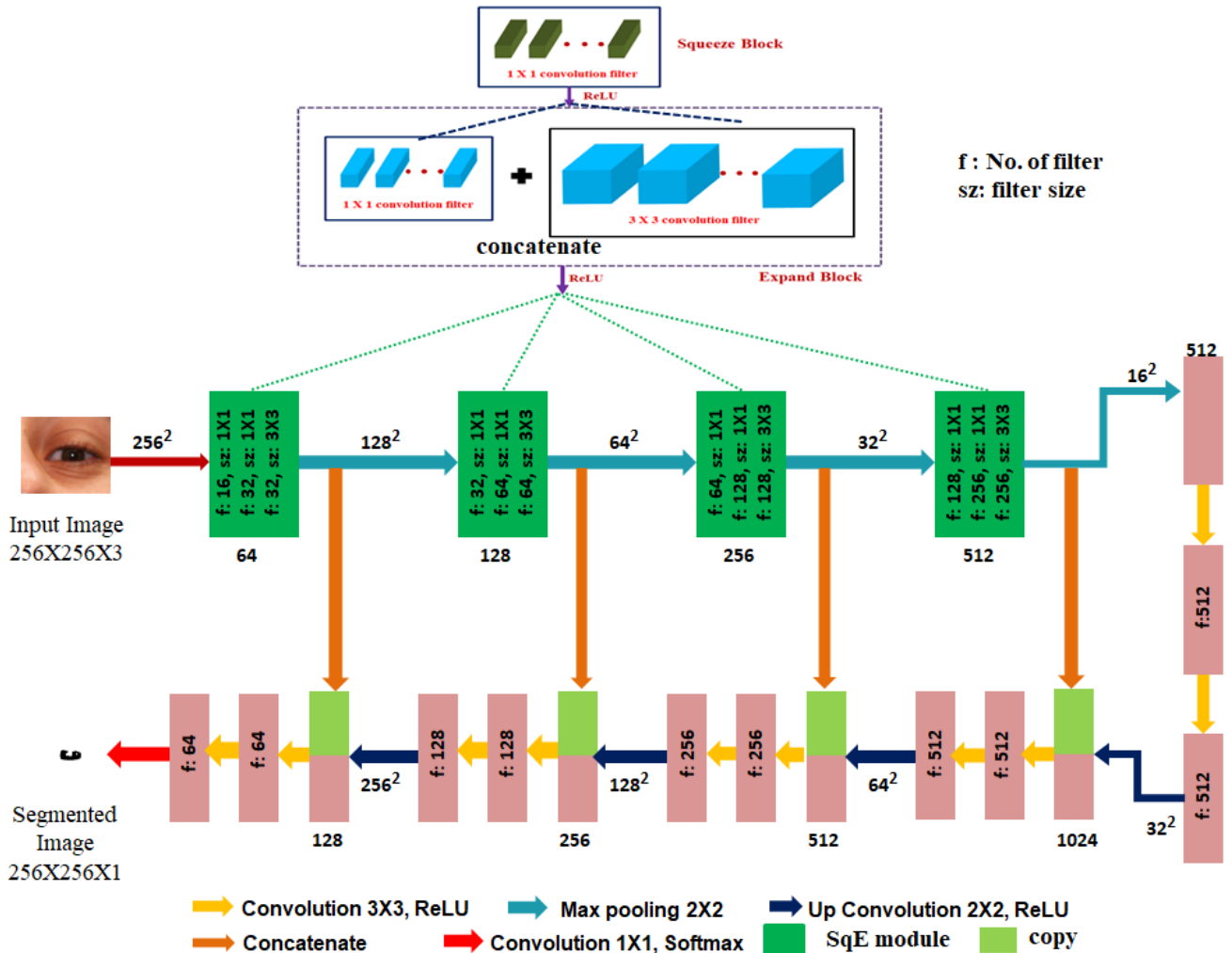


FIGURE 3. Model ISqEUNet with SqE module.

Thus the model interactively learns and updates the weights, in the absence of sufficient annotations, and improves its prediction accuracy while helping generate ground truth annotations for the unseen images.

It is to be noted that in most of these cases no interaction is required. For some of the noisy images minimal user interaction is required, involving manual correction of only 2-3% pixels. Bypassing the manual annotation of the entire iris region, interactive learning allows refinement of just a few incorrectly predicted pixels of the noisy images.

## VI. EXPERIMENTAL RESULTS

In this section we present the performance of ISqEUNet on the three publicly available databases. Implementation was made in Keras. Both qualitative and quantitative evaluation of iris segmentation is provided. Comparative study with state-of-the-art literature establishes the effectiveness of our model.

### A. DATASETS

Three publicly available well-known iris databases, viz. CASIA-IrisV4-Interval [33], IITD [34], [37] and UBIRIS.v2 [35] were used to evaluate the performance of ISqEUNet.

CASIA-Iris-V4-Interval is a subset of CASIA-IrisV4, collected by the Chinese Academy of Sciences' Institute of Automation (CASIA). The CASIA-Iris-V4-Interval database contains left and right eye images of 249 (mostly Chinese) subjects, each with approximately 1-10 images, with the total number of images being 2639 of size  $320 \times 280$  and 256 grey levels. The IITD database has 224 subjects (students and staff of IIT Delhi, India), with 10 (left and right) iris images of size  $320 \times 240$  from each.

UBIRIS.v2 is an iris database which contains visible wavelength iris images captured on-the-move and at a distance, with more realistic noise factors. This database, with images of size  $400 \times 300$ , encompasses a total of 261 subjects. A subset of 500 images from UBIRIS.v2 (as obtained from [38], with ground truth) was used for training, along with the remaining 500 samples for testing. This was called NICE.I competition challenge [38], and was used in our experiments.

### B. EVALUATION METRICS

The Mean Error Rate (MER) [4] is a widely used evaluation metric for the task of binary image segmentation. It provides

**TABLE 1.** Contracting path architecture of *UNet*, *AttUNet* and *ISqEUNet*.

	<i>UNet</i>	<i>AttUNet</i>	<i>ISqEUNet</i>
Input size	$256 \times 256 \times 3$	$512 \times 512 \times 3$	$256 \times 256 \times 3$
layers in contracting path	64,conv1,3 × 3,ReLU	64,conv1,3 × 3,ReLU	16,conv1,1 × 1,ReLU
	64,conv2,3 × 3,ReLU	64,conv2,3 × 3,ReLU	32,conv2,1 × 1,ReLU
	-	-	32,conv3,3 × 3,ReLU
	pool1, 2 × 2	pool1, 2 × 2	pool1, 2 × 2
	128,conv3,3 × 3,ReLU	128,conv3,3 × 3,ReLU	32,conv4,1 × 1,ReLU
	128,conv4,3 × 3,ReLU	128,conv4,3 × 3,ReLU	64,conv5,1 × 1,ReLU
	-	-	64,conv6,3 × 3,ReLU
	pool2, 2 × 2	pool2, 2 × 2	pool2, 2 × 2
	256,conv5,3 × 3,ReLU	256,conv5,3 × 3,ReLU	64,conv7,1 × 1,ReLU
	256,conv6,3 × 3,ReLU	256,conv6,3 × 3,ReLU	128,conv8,1 × 1,ReLU
layers in upsampling path	-	256,conv7,3 × 3,ReLU	128,conv9,3 × 3,ReLU
	pool3, 2 × 2	pool3, 2 × 2	pool3, 2 × 2
	512,conv7,3 × 3,ReLU	512,conv8,3 × 3,ReLU	128,conv10,1 × 1,ReLU
	512,conv8,3 × 3,ReLU	512,conv9,3 × 3,ReLU	256,conv11,1 × 1,ReLU
	-	512,conv10,3 × 3,ReLU	256,conv12,3 × 3,ReLU
	pool4, 2 × 2	pool4, 2 × 2	pool4, 2 × 2
	1024,conv9,3 × 3,ReLU	512,conv11,3 × 3,ReLU	512,conv13,3 × 3,ReLU
	1024,conv10,3 × 3,ReLU	512,conv12,3 × 3,ReLU	512,conv14,3 × 3,ReLU
	-	512,conv13,3 × 3,ReLU	-
	up-sampling, 2 × 2	up-sampling, 2 × 2	up-sampling, 2 × 2
No. of parameters	$3.1 \times 10^7$	$2.6 \times 10^7$	$1.6 \times 10^7$

<sup>a</sup> Layers in expanding path are the same as shown in Fig.1 for *UNet*, *ATTUNet* and *ISqEUNet*.

the ratio of all false pixel prediction in the whole image. Let  $Im$  be the input image of size  $r \times c$ , with  $P(i, j)$  the predicted binary image and  $G(i, j)$  its ground truth mask (annotation). Here  $N$  is the total number of test images and  $\oplus$  denotes the binary XOR operation. We have

$$MER = \frac{1}{(N \times r \times c)} \times \sum_{i=1}^r \sum_{j=1}^c P(i, j) \oplus G(i, j). \quad (2)$$

Dice similarity coefficient (*DSC*) is a spatial overlap index, which provides a measure of similarity between the prediction and the ground truth. The range of this matrix is  $[0, 1]$ , with higher values indicating better segmentation. It is defined as

$$DSC = \frac{2 \times |P \cap G|}{|P| + |G|}. \quad (3)$$

The Mean True Positive Rate (*mTPR*) is another common measure used in image segmentation. It computes the average ratio of predicted ground truth pixels w.r.t. the total ground truth-foreground pixels. It is expressed as

$$mTPR = \frac{1}{N} \times \frac{TP}{TP + FN}, \quad (4)$$

where  $TP$  and  $FN$  denote the numbers of true and false positives, respectively.

### C. SEGMENTATION

The model was trained on 70% of the data, for classification into iris and non-iris pixels; the remaining data constituted the test set. Each set consisted of discrete images from independent individuals, such that there were no images from the same person in more than one set. Training the

### Algorithm 1 Interactive Learning

**Input:** Data set  $X \in \{X_i\}_{i=1}^k$ , where  $k = \#$  images in  $X$ , Pre-trained ISqEUNet with frozen weights  $W$ , Threshold  $c \in [0, 1]$ ;

**Output:** Refined Ground Truth  $Y$

```

1: Input  $c$  /* user-defined threshold to control fine tuning */
2: repeat
3:   Initialize  $Flag \leftarrow 0$ ; /* Flag to determine if refinement needed */
4:   Initialize  $Counter \leftarrow 0$ ;
5:   for each sample image  $\{X_i, i = 1, \dots, k\}$  do
6:     Input  $X_i$  to the ISqEUNet with weight  $W$  and generate predicted segmentation output  $Y_i$ ;
7:     if user detects any misclassified pixels in image  $Y_i$  then
8:       User refinement through interaction;
9:       Generate refined output  $\hat{Y}_i$ ; /* toggles binary value for refinement */
10:       $Counter \leftarrow Counter + 1$ ; /* Increment counter */
11:       $Y'_i \leftarrow \hat{Y}_i$ ; /* Replace predicted output image by refined version */
12:    end if
13:  end for
14:  if  $Counter \geq c * k$  then
15:    Set  $Flag \leftarrow 1$ ;
16:     $[W] = FineTuning(X, W, Y')$  /*Weights of ISqEUNet are updated */
17:  end if
18: until ( $Flag == 0$ );
19: return (Refined Ground Truth  $Y$ );

```

### Algorithm 2 FineTuning ( $X, W, Y'$ )

**Input:** Data set  $X$ , ISqEUNet with weights  $W$ , Refined output  $Y'$ , No. of epochs  $Ep$

**Output:** Fine-tuned Weights  $W'$

```

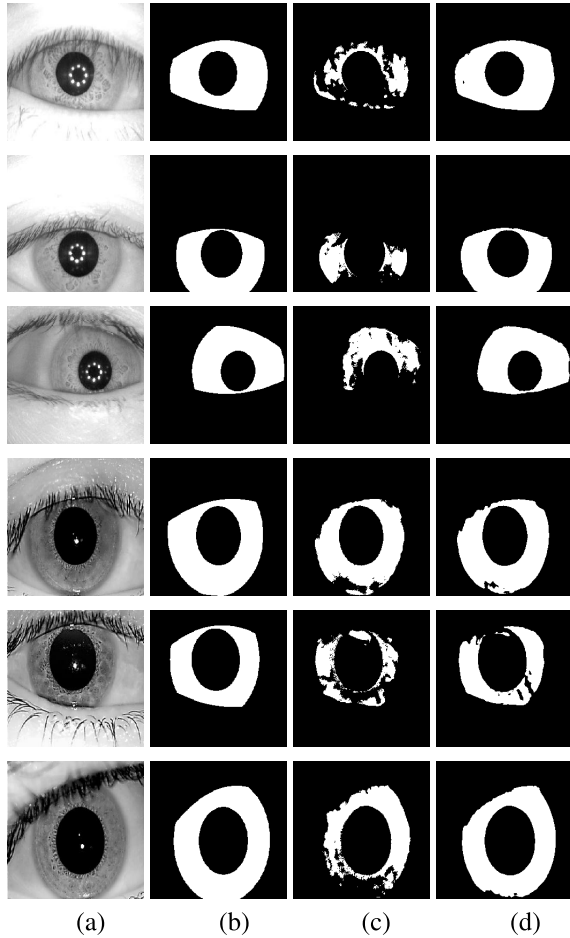
1: for  $i := 1$  to  $Ep$  do
2:   Train ISqEUNet to generate predicted output  $\hat{Y}_p$ 
3:   Update weights  $W'$  based on Loss Minimization,  $L_B(\hat{Y}_p, Y')$ ;
4: end for
5: return ( $W'$ );

```

ISqEUNet over 15 epochs (averaged over 5 runs), with batch size of 16 and learning rate 0.00001, was found to be sufficient (over 5 runs) for stopping based on cross validation. The optimum weights of the model and the hyperparameters were experimentally determined. The mean and standard deviation (Sd) of the quantitative metrics [eqns. (2)-(4)] were computed over 5 runs. The average values are presented in Table 2, with the value within parentheses indicating the Sd in each case, for the datasets CASIA-IrisV4-Interval and IITD. The number of parameters (as specified in Table 1)

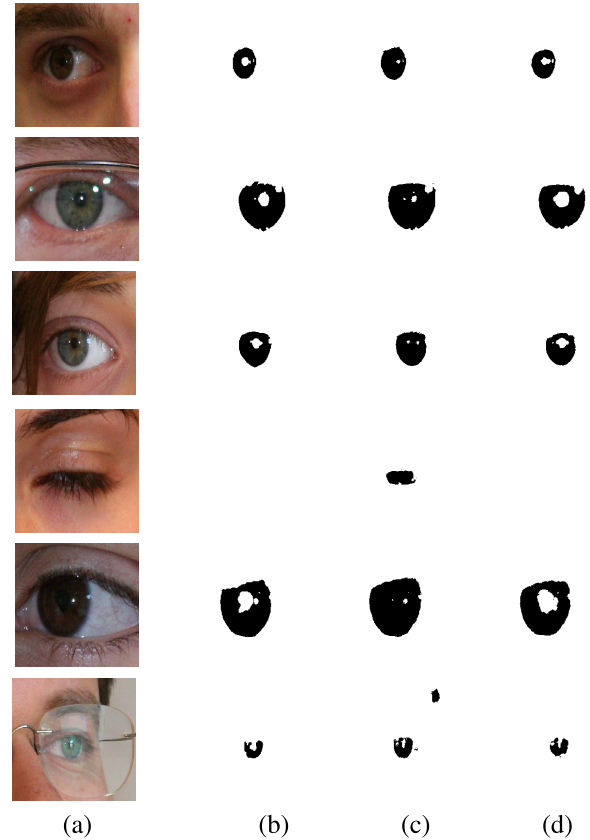
**TABLE 2.** Quantitative comparison of segmentation by *ISqEUNet* (without interactive learning).

Model	CASIA-IrisV4-Interval			IITD		
	<i>DSC</i>	<i>mTPR</i>	<i>MER</i>	<i>DSC</i>	<i>mTPR</i>	<i>MER</i>
<i>UNet</i>	0.913 (0.008)	0.908 (0.007)	0.996 (0.004)	0.934 (0.004)	0.922 (0.006)	0.954 (0.006)
<i>ISqEUNet</i>	0.987 (0.0013)	0.972 (0.0077)	0.176 (0.0027)	0.985 (0.0031)	0.980 (0.0048)	0.249 (0.0034)

**FIGURE 4.** Comparative segmentation (without interactive learning) of iris from sample images: Rows 1-3 from database CASIA-IrisV4-Interval, and rows 4-6 from IITD. (a) Original, and (b) Ground truth, with output obtained by (c) *UNet*, and (d) *ISqEUNet*.

were kept the same while evaluating the performance of the models over all three datasets. Comparative study was also made with *UNet* on the same data. It was observed that *ISqEUNet* performs better in all cases (as depicted in bold in the last row) for these unconstrained datasets.

A qualitative comparison of the segmentation output, with that of *UNet*, is presented in Fig. 4 for sample images. While the first three rows of the figure correspond to images from CASIA-IrisV4-Interval, the last three rows refer to images from IITD. We provide sample segmentation results on some noisy cases from both datasets in the figure. The CASIA-IrisV4-Interval and IITD datasets involve NIR images with high contrast, with many images suffering from the effects of (i) drooping eyelashes and low contrast between

**FIGURE 5.** Segmentation of iris from sample images of non-ideal NICE.I database. (a) Original, (b) Ground truth, (c) *UNet*, and (d) Prediction by *ISqEUNet* with interactive learning.

sclera and iris parts, (ii) having same pixel values in skin area and iris region, (iii) involving non-cooperative and blurring effects. The *ISqEUNet* learned to reduce segmentation error caused by noisy pixels, and was able to demarcate the visible iris part in a better manner as compared to conventional *UNet*.

#### D. INTERACTIVE LEARNING

Here we present the refinement in segmentation, evaluated quantitatively as well as qualitatively. The baseline model was *UNet* without the SE module. Interactive segmentation was incorporated into the *ISqEUNet* model, pre-trained on the CASIA-IrisV4-Interval dataset. Next the model was interactively fine-tuned on NICE.I database (without using available ground truth). Fine tuning was performed using Stochastic Gradient Descent (SGD) optimizer with small learning rate ( $10^{-8}$ ). Fig. 5 provides the segmentation output on NICE.I data from different sample images.

For an input image of size  $256 \times 256 \times 3$ , the run time of *ISqEUNet* was  $6.152e-06$  sec., on an Nvidia Quadro K6000 GPU with 12GB DDR5 RAM; whereas, in case of *UNet* it was  $9.106e-06$  sec. Table 3 lists a summary of comparative segmentation performance, with existing state-of-the-art methodologies, on dataset NICE.I. It is observed that the error *MER* is lower for our model *ISqEUNet*, as compared to *AttUNet* and *UNet*. Introducing interactive learning results in further improvement in performance.



**TABLE 3.** Comparative study on NICE.I dataset with related state-of-the-art approaches from literature.

Methods	NICE.I	
	<i>mTPR</i>	<i>MER</i>
Proença et al. [14]	-	1.87
Tan et al. [13]	-	1.72
Zhao et al. [9]	-	1.21
MFCN [27]	-	0.90
UNet [29]	0.948	0.898
AttUNet [4]	0.968	0.764
ISqEUNet (without interactive learning)	0.972	0.374
ISqEUNet (with interactive learning)	<b>0.983</b>	<b>0.261</b>

This is particularly true for cases where annotated data is scarce, mainly due to the expenses involved. Fine tuning or refinement, through user interaction, is found to perform very well in such scenario. It enables semi-supervised generation of effective annotations. This is evident from the results on NICE.I dataset. Given the small size of the data the value of *Ep* for fine tuning was experimentally chosen to be 5 with good results.

Some of the other related segmentation approaches [9], [13], [14], [27] were also compared to establish the effectiveness of our model. Qualitative comparison on some challenging image samples, involving hair occlusion (row 3), eye-glass occlusion (rows 2, 6), and non-cooperation (rows 1, 4, 5) are illustrated in Fig. 5. It is visually evident that our ISqEUNet model provides good segmentation results with interactive learning, for such non-ideal images. Interestingly, even in the absence of user interaction, the performance of ISqEUNet is found to be superior to that of UNet and AttUNet (as reported in [4]) using the architectural layout of table 1.

## VII. CONCLUSION

A novel deep ISqEUNet model was developed, with interactive learning, for efficient iris segmentation in a non-ideal environment; encompassing challenging situations like occlusion, reflection, and poor illumination. Incorporation of the Squeeze-Expand module enabled improved speed of network training by reducing the number of trainable parameters, along with enhanced accuracy of the segmentation. Fine tuning for interactive learning allowed robustness in iris segmentation, particularly with availability of insufficient annotations.

Segmentation results were compared with that of state-of-the-art iris segmentation methodologies. Results on CASIA-Irisv4-Interval, IITD, NICE.I databases established the superiority of our algorithm for non-ideal iris images, in terms of evaluation metrics *MER*, *DSC* and *mTPR*. Experimental results demonstrated that iris images captured in visible wavelength and NIR, including non-cooperative and non-ideal samples, could be appropriately segmented with ISqEUNet without involving any complex pre- or

post-processing. The mean error rate showed an improvement by at least 0.4% over the existing methods.

Semi-automated approaches for training data generation, with iterative refinement of deep learning models, is expected to become an important tool in a researcher's imaging toolset in the near future. Future study aims to improve upon the accuracy and optimization of the model pipeline, to enable it to perform in real time.

## REFERENCES

- [1] J. Daugman, "New methods in iris recognition," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 37, no. 5, pp. 1167–1175, Oct. 2007.
- [2] H. Proença and L. A. Alexandre, "Iris recognition: Analysis of the error rates regarding the accuracy of the segmentation stage," *Image Vis. Comput.*, vol. 28, no. 1, pp. 202–206, Jan. 2010.
- [3] K. W. Bowyer and M. J. Burge, Eds., *Handbook of Iris Recognition*. London, U.K.: Springer, 2016.
- [4] 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, pp. 296–304, Oct. 2018.
- [5] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1148–1161, Dec. 1993.
- [6] R. P. Wildes, J. C. Asmuth, G. L. Green, S. C. Hsu, R. J. Kolczynski, J. R. Matey, and S. E. McBride, "A machine-vision system for iris recognition," *Mach. Vis. Appl.*, vol. 9, no. 1, pp. 1–8, Jul. 1996.
- [7] S. Shah and A. Ross, "Iris segmentation using geodesic active contours," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 4, pp. 824–836, Dec. 2009.
- [8] A. Uhl and P. Wild, "Weighted adaptive Hough and ellipsoidal transforms for real-time iris segmentation," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar. 2012, pp. 283–290.
- [9] Z. Zhao and A. Kumar, "An accurate iris segmentation framework under relaxed imaging constraints using total variation model," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3828–3836.
- [10] Z. He, T. Tan, Z. Sun, and X. Qiu, "Toward accurate and fast iris segmentation for iris biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1670–1684, Sep. 2009.
- [11] S. J. Pundlik, D. L. Woodard, and S. T. Birchfield, "Non-ideal iris segmentation using graph cuts," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2008, pp. 1–6.
- [12] C.-W. Tan and A. Kumar, "Unified framework for automated iris segmentation using distantly acquired face images," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4068–4079, Sep. 2012.
- [13] C.-W. 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.
- [14] H. Proença, "Iris recognition: On the segmentation of degraded images acquired in the visible wavelength," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1502–1516, Aug. 2010.
- [15] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.
- [16] P. H. Pinheiro and R. Collobert, "Recurrent convolutional neural networks for scene labeling," in *Proc. 31st Int. Conf. Mach. Learn. (ICML)*, 2014, pp. 82–90.
- [17] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [19] N. Liu, M. Zhang, H. Li, Z. Sun, and T. Tan, "DeepIris: Learning pairwise filter bank for heterogeneous iris verification," *Pattern Recognit. Lett.*, vol. 82, pp. 154–161, Oct. 2016.
- [20] M. Trokielewicz, A. Czajka, and P. Maciejewicz, "Post-mortem iris recognition with deep-learning-based image segmentation," *Image Vis. Comput.*, vol. 94, Feb. 2020, Art. no. 103866.
- [21] T. Zhao, Y. Liu, G. Huo, and X. Zhu, "A deep learning iris recognition method based on capsule network architecture," *IEEE Access*, vol. 7, pp. 49691–49701, 2019.

- [22] H. Hofbauer, E. Jalilian, and A. Uhl, "Exploiting superior CNN-based iris segmentation for better recognition accuracy," *Pattern Recognit. Lett.*, vol. 120, pp. 17–23, Apr. 2019.
- [23] M. Arsalan, R. Naqvi, D. Kim, P. Nguyen, M. Owais, and K. 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. 5, p. 1501, May 2018.
- [24] 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, p. 263, Nov. 2017.
- [25] A. Gangwar and A. Joshi, "DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2301–2305.
- [26] 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.
- [27] 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. Int. Conf. Biometrics (ICB)*, Jun. 2016, pp. 1–8.
- [28] S. Bazrafkan, S. Thavalengal, and P. Corcoran, "An end to end deep neural network for iris segmentation in unconstrained scenarios," *Neural Netw.*, vol. 106, pp. 79–95, Oct. 2018.
- [29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015.
- [30] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: <http://arxiv.org/abs/1602.07360>
- [31] J. Egger, T. Kapur, A. Fedorov, S. Pieper, J. V. Miller, H. Veeraraghavan, B. Freisleben, A. J. Golby, C. Nimsky, and R. Kikinis, "GBM volumetry using the 3D slicer medical image computing platform," *Sci. Rep.*, vol. 3, no. 1, pp. 1073–1094, Dec. 2013.
- [32] V. Vezhnevets and V. Konouchine, "GrowCut: Interactive multi-label N-D image segmentation by cellular automata," in *Proc. GraphiCon*, 2005, pp. 150–156.
- [33] (2008). *CASIA Iris Image Database V3.0*. [Online]. Available: <http://biometrics.idealtest.org/>
- [34] (2007). *IIT Delhi Iris Image Database Version 1.0*. [Online]. Available: [http://www4.comp.polyu.edu.hk/~csajaykr/iitd/database\\_iris.htm](http://www4.comp.polyu.edu.hk/~csajaykr/iitd/database_iris.htm)
- [35] 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.
- [36] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018.
- [37] A. Kumar and A. Passi, "Comparison and combination of iris matchers for reliable personal authentication," *Pattern Recognit.*, vol. 43, no. 3, pp. 1016–1026, Mar. 2010.
- [38] H. Proenca and L. A. Alexandre, "The NICE.I: Noisy iris challenge evaluation—Part I," in *Proc. 1st IEEE Int. Conf. Biometrics, Theory, Appl., Syst.*, Sep. 2007, pp. 1–4.



**MOUSUMI SARDAR** received the B.Tech. degree from the College of Engineering and Management, Kolaghat, and the M.Tech. degree from the Maulana Abul Kalam Azad University of Technology (formerly known as the West Bengal University of Technology), in information technology. She is currently a RGNF Senior Research Fellow with the Machine Intelligence Unit, Indian Statistical Institute, Kolkata. Her current research interests include biometrics, image processing, and machine learning.



**SUBHASHIS BANERJEE** (Student Member, IEEE) is currently a Visvesvaraya Senior Research Fellow with the Machine Intelligence Unit, Indian Statistical Institute, Kolkata. He was selected as the Intel Student Ambassador in AI by Intel Corporation, USA, in 2018. He was a Visiting Researcher with the Division of Visual Information and Interaction, Uppsala University, Uppsala, Sweden, in 2019, with the Department of Computer Science, Bioengineering, Robotics, and



**SUSHMITA MITRA** (Fellow, IEEE) was with the RWTH, Aachen, Germany, as a DAAD Fellow, from 1992 to 1994. She was a Visiting Professor with the Computer Science Department, University of Alberta, Edmonton, AB, Canada, in 2004 and 2007, and with Meiji University, Japan, in 1999, 2004, 2005, and 2007, and with Aalborg University Esbjerg, Denmark, in 2002 and 2003. She is currently the INAE Chair Professor and a Fulbright-Nehru Senior Researcher with the Machine Intelligence Unit (MIU), Indian Statistical Institute, Kolkata. She has more than 150 research publications in referred international journals. She is the author of the books *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing and Data Mining: Multimedia, Soft Computing, and Bioinformatics* (John Wiley), and *Introduction to Machine Learning and Bioinformatics* (Chapman & Hall/CRC Press), beside a host of other edited books. Her current research interests include data mining, pattern recognition, soft computing, medical image processing, and bioinformatics.

Dr. Mitra is a Fellow of the Indian National Science Academy (INSA), International Association for Pattern Recognition (IAPR), and a Fellow of the Indian National Academy of Engineering (INAE) and The National Academy of Sciences, India (NASI). She received the National Talent Search Scholarship from 1978 to 1983 from NCERT, India, the University Gold Medal in 1988, the IEEE TNN Outstanding Paper Award in 1994, for her pioneering work in neuro-fuzzy computing, and the CIMPA-INRIA-UNESCO Fellowship in 1996. She has guest edited special issues of several journals, is an Associate Editor of IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, *Information Sciences*, *Fundamenta Informatica*, and is a Founding Associate Editor of *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery (WIRE DMKD)*. According to the Science Citation Index (SCI), two of her articles have been ranked Third and 15th in the list of Top-cited articles in Engineering Science from India, from 1992 to 2001. She has visited more than 30 countries as a Plenary/Invited Speaker or an Academic Visitor. She served as the capacity of Program Chair, the Tutorial Chair, and as a member of programme committees of many international conferences. She is an IEEE CIS Distinguished Lecturer and the current Chair Elect, IEEE Kolkata Section.

...