

Sclera Segmentation and Joint Recognition Benchmarking Competition: SSRBC 2023

Abhijit Das^{1,*,\dagger}, Saurabh Atreya^{1,\dagger}, Aritra Mukherjee¹, Matej Vitek², Haiqing Li³, Caiyong Wang³, Guangzhe Zhao³, Fadi Boutros⁴, Patrick Siebke⁴, Jan Niklas Kolf⁴, Naser Damer^{4,5}, Sun Ye⁶, Lu Hexin⁶, Fan Aobo⁶, You Sheng⁶, Sabari Nathan⁷, R. Suganya⁸, R. S. Rampriya⁸, Geetanjali Sharma⁹, Priyanka P⁹, Aditya Nigam⁹, Peter Peer², Umapada Pal¹⁰, Vitomir Štruc²

¹Birla Institute of Technology and Science, Pilani (BITS Pilani), India, ²University of Ljubljana (UL, Slovenia), ³Beijing University of Civil Engineering and Architecture (BUCEA), China, ⁴Fraunhofer Institute for Computer Graphics Research IGD (IGD), Germany, ⁵TU Darmstadt, Germany, ⁶Jilin University (JLU, China), ⁷Cougar Inc (Japan), ⁸Vellore Institute of Technology, Vellore (VIT, India), ⁹Indian Institute of Technology, Mandi (IIT-MD, India), ¹⁰Indian Statistical Institute, Kolkata (ISI, India)

Abstract

This paper presents the summary of the Sclera Segmentation and Joint Recognition Benchmarking Competition (SSRBC 2023) held in conjunction with IEEE International Joint Conference on Biometrics (IJCB 2023). Different from the previous editions of the competition, SSRBC 2023 not only explored the performance of the latest and most advanced sclera segmentation models, but also studied the impact of segmentation quality on recognition performance. Five groups took part in SSRBC 2023 and submitted a total of six segmentation models and one recognition technique for scoring. The submitted solutions included a wide variety of conceptually diverse deep-learning models and were rigorously tested on three publicly available datasets, i.e., MASD, SBVPI and MOBIUS. Most of the segmentation models achieved encouraging segmentation and recognition performance. Most importantly, we observed that better segmentation results always translate into better verification performance.

1. Introduction

The human sclera (i.e., the white portion of the eye) represents an emerging biometric modality that contains a rich vascular structure that can be exploited for automatic identity inference. While ocular biometric traits, such as the iris, are already utilized on mobile devices in various authentication schemes, recent literature is increasingly looking at alternative ocular traits that can complement these schemes and ensure higher levels of security, spoof resistance and better overall recognition results. Among such traits, the

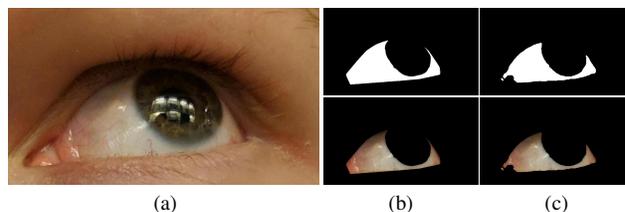


Figure 1. The (a) captured image of the eye, along with the (b) manually segmented mask, are used to train segmentation models that can generate masks like (c). The mask is then superimposed on the image to finally attain the sclera images, which are used to train and test recognition models.

sclera is particularly appealing due to its desirable characteristics [1–3] such as (i) the uniqueness of the vasculature and its stability over time, (ii) the robustness with respect to certain eye conditions and presence of accessories (astigmatism, cataracts, lenses), which negatively affect retina- and iris-based recognition systems, (iii) the difficulty of forging the vascular patterns, which makes sclera biometrics (partially) resilient to presentation attacks and spoofing, and (iv) the possibility of using cost-effective off-the-shelf cameras, for data acquisition in the visible spectrum unlike the iris or retina, which typically require specialized capturing hardware. These characteristics have fueled recent research on exploring sclera as a biometric, including techniques for recognition [3–9], segmentation models [2, 10, 11], presentation attacks detection approaches [2, 3, 6], adaptability of the trait [12, 13], information fusion with the iris [14, 15] and synthetic sclera generation models [16].

Despite this plethora of work, comprehensive studies investigating the characteristics of sclera segmentation and recognition models jointly and their impact on the overall recognition performance in biometric systems relying on the vasculature of the sclera are still underrepresented in the literature. The aim of the sclera segmentation task is to locate the sclera region in the eye image as precisely as pos-

*Corresponding author: Abhijit Das: abhijitdas@hyderabad.bits-pilani.ac.in

^{\dagger}Abhijit Das and Saurabh Atreya are the first authors and have equal contributions

sible and, consequently, to facilitate accurate recognition. Thus, the segmentation step is expected to considerably impact all downstream tasks, including the feature extraction and similarity computation steps. As illustrated in Figure 1, the segmentation process needs to effectively identify all pixels in the image that belong to the sclera, exclude background areas and ocular components that could interfere with the recognition stage, and do so reliably even if the images exhibit challenging characteristics, such as blur, extreme gazes or are captured in unconstrained (mobile) settings and under varying illumination conditions. Given the challenges faced by the segmentation models, it is critical to understand what kind of performance degradation can be expected for the recognition tasks due to issues with the segmentation results.

To get detailed insight into these issues, the Sclera Segmentation and Recognition Benchmarking Competition (SSRBC 2023) was organized as part of the International Joint Conference on Biometrics (IJCB 2023). The competition focused on a joint evaluation of sclera segmentation and recognition models, both of which are key components of sclera-based recognition systems. It is quite obvious that improper segmentation will adversely impact the feature extraction and recognition performance, but the extent of this impact is not clear based on the insights currently available in the open literature. Hence, SSRBC 2023 was organized around three main datasets. The Multi-angle sclera dataset (MASD) was used for training, and the Mobile Ocular Biometrics in Unconstrained Settings (MOBIUS) and the Sclera Blood Vessels, Periocular, and Iris (SBVPI) datasets for testing. The experimental setup used for the benchmarking competition allowed us to explore a number of research questions, such as: How do the segmentation models perform in cross-dataset (and cross-sensor) settings? What differences can be expected in performance due to different model architectures? How are segmentation errors propagated to the recognition stage? To answer these questions **6 segmentation models were contributed** to the competition from 5 different research groups and analyzed for their performance. The joint effort of all participants resulted in the following contributions that are presented in this paper:

- A rigorous analysis of several contemporary sclera segmentation and recognition models using images captured in challenging mobile scenarios.
- A detailed analysis of the sensitivity of sclera recognition models on errors in the segmentation process.

2. Related Work

SSRBC 2023 is the eighth edition in the series of competitions focused on problems in sclera biometrics, initiated in 2015 as part of BTAS 2015. This series of competitions had a considerable impact on the state of technology in sclera

biometrics over the last decade attracted a good amount of interest from the community.

The first and third iterations of SSRBC aimed at benchmarking segmentation models and introduced novel datasets for this task together with manually annotated ground truth masks [17]. The second iteration (SSRBC 2016) investigated the performance of segmentation and for the first time, recognition techniques in a group evaluation setting [18] was also accomplished. In the fourth iteration, i.e., SSERBC 2017, ocular recognition models for the iris, sclera and peri-ocular region were benchmarked with images of varying gaze directions in addition to the standard segmentation task [19]. SSBC 2018, the fifth edition of this series, explored the impact of cross sensor image capture on sclera segmentation [20], and the sixth iteration, SSBC 2019, investigated the performance of segmentation models in cross-resolution settings [21]. In the last edition, SSBC 2020, sclera-segmentation techniques applied to images, captured in the mobile environments, were benchmarked [22]. A follow-up group benchmarking effort following SSBC 2020 also looked at bias in sclera segmentation models due to the capturing sensors, ethnicity, and eye color [23].

The 2023 edition of SSRBC aims to continue the series with a benchmarking competition focused on a *joint segmentation and recognition problem*, which is highly relevant for the deployment of sclera biometrics in real-life applications. To make this task challenging and as close to real-world operating conditions, we utilize diverse ocular data captured with different sensors, with varying gaze directions and acquisition conditions, which is expected to be an additional important contribution to the community.

3. Benchmarking methodology

This section presents the benchmarking methodology used for SSRBC 2023. It first describes the datasets and logistics of the competition and then presents the performance metrics utilized for the final comparative evaluation.

3.1. SSRBC 2023 datasets

Three datasets were used for SSRBC 2023: (i) the Multi-Angle Sclera Dataset (MASD) [24], (ii) the Mobile Ocular Biometrics in Unconstrained Settings (MOBIUS) dataset [23], and (iii) the Sclera Blood Vessels, Periocular, and Iris (SBVPI) [2]. The high-level characteristics of these datasets are given in Table 1, a few visual examples are presented in Figure 2 and a more detailed description is provided below:

MASD The MASD dataset consists of high-resolution ocular images captured using a DSLR camera (Nikon D800 28-300mm lens). There are a total of 2624 RGB images obtained from 82 distinct participants. Each eye is treated as a separate identity, so there are a total 164 different IDs in the

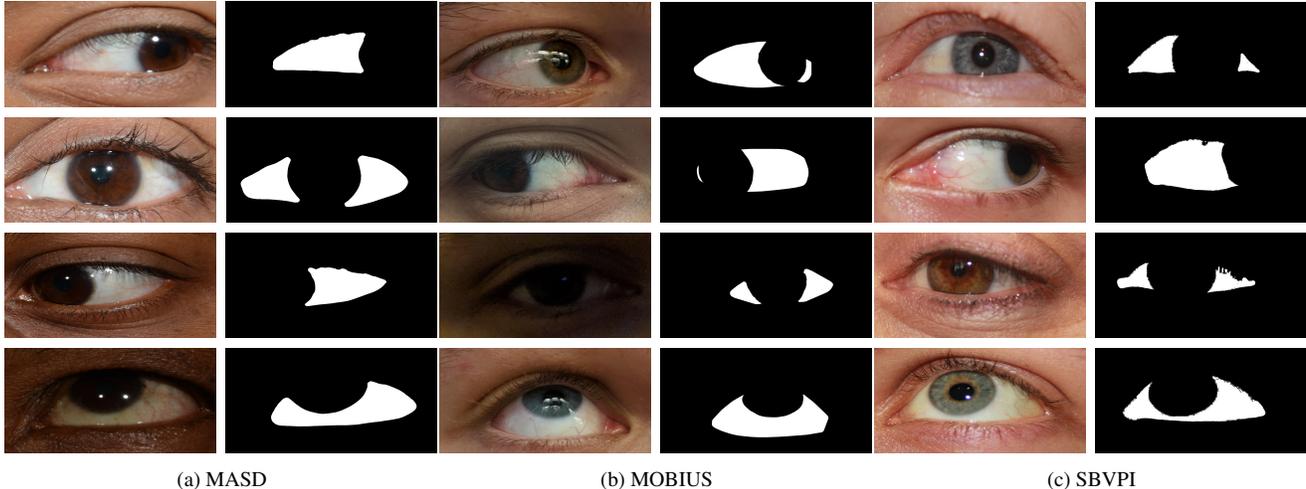


Figure 2. Sample images and ground truth masks for the datasets used in the competition. The figure displays images and sclera ground truth masks from (a) MASD, (b) MOBIUS, and (c) SBVPI. Note the variability across image qualities and gaze directions present in the datasets.

Table 1. Overview of the datasets used for SSRBC 2023. Information on the number of images, number of subjects, the image resolution (in pixels), main sources of variability and purpose in the competition is provided.

Dataset	#Images	#IDs	Resolution	Variability	Purpose
MASD	2624	82	7500 × 5000	GZ, BL	Training
MOBIUS	3542	35	3000 × 1700	MD, CN, GZ, BL	Testing
SBVPI	1858	55	3000 × 1700	CN, GZ	Testing

[†]GZ - gaze, BL - blur, CN - acquisition condition, MD - mobile device.

dataset available for the experiments, and each ID has images that correspond to four gaze directions (straight, left, right and up). A total of 4 images are taken for each direction. The participants consist of a mix of male and female subjects with some of them wearing contact lenses. The ambient lighting and overall illumination factors are varied by capturing images at different times of the day. There are also some images in which the participants blinked or partially closed the eyes. The images are of a resolution 7500×5000 and come with manually annotated segmentation masks for the sclera region.

MOBIUS. The second dataset used for SSRBC 2023 is a recent one, intended for use in mobile ocular biometrics. MOBIUS has 100 subjects and over 16,000 RGB ocular images. For SSRBC 2023, the segmentation split of the entire dataset is used that contains 3542 manually annotated RGB images captured from 35 participants (i.e. 70 eyes). Different smartphones, such as Apple’s iPhone 6s, Sony’s Xperia Z5 Compact, and Xiaomi’s Pocophone F1, in various lighting conditions were used for the acquisition process. The lighting conditions can be further grouped into *Good*: controlled indoor lighting, *Neutral*: uncontrolled outdoor lighting, and *Poor*: uncontrolled indoor lighting. The subjects of

the dataset are a mix of Caucasian male and female persons. The variability in the captured images is due to the different types of mobile photo-sensors utilized and changes in the acquisition conditions. All images are manually annotated with masks for the sclera, iris and pupil and the imperfections were fixed by a semi-automated post-processing technique. Because SSRBC 2023 is only focusing on sclera segmentation, the masks of the other ocular parts are not considered in the competition.

SBVPI. The third dataset used [2, 3] for SSRBC 2023 consists of 55 participants (110 eyes), and comprises a total of 1858 semi-high resolution RGB images, captured by a DSLR camera. The resolution of each image is scaled to 3000×1700 and the image quality used during capturing was set to the highest possible setting under controlled lighting. Similar to MASD the dataset, SBVPI contains images of four gaze directions (i.e. straight, up, right and left). The data variability in the dataset is due to identity, gender, age and eye color. Each image is manually annotated with masks for the sclera and periocular regions.

Since the datasets are publicly available, abstraction was used on the annotations of the images and masks. For each task, pseudonyms were used in place of the original file-names so as to not reveal the labels for identity, gaze direction, lighting condition and capturing device. Participants were instructed to make their submissions in the same format.

3.2. Competition tasks and protocol

SSRBC 2023 defined three main competition tasks targeting, i.e., (i) sclera segmentation, (ii) sclera recognition and (iii) joint sclera segmentation and recognition. The tasks are described in detail below:

- **Segmentation task (ST).** For the segmentation task, participants had to learn segmentation models on the MASD datasets and then test them on the MOBIUS datasets. Here, the manually generated (ground truth) segmentation masks were provided along with the eye images of MASD for the training procedure. For testing purposes, only the ocular images from the MOBIUS dataset were provided to the participants without any ground truth annotations.
- **Recognition task (RT).** For the recognition task, the participants were asked to develop recognition models on the MASD datasets. The performance evaluation was then conducted on the sequestered SBVPI dataset. In this setting, the manually generated (ground truth) segmentation mask were used to get the region-of-interest (ROI) before subjecting the images to the recognition/feature extraction models. Thus, Oracle-type of experiments were performed in this task.
- **Joint Segmentation and Recognition task (SRT).** For the joint segmentation-recognition task, the participants were asked to develop segmentation as well as recognition models on the MASD datasets and then submit the trained models for scoring to the organizers. The performance evaluation was conducted on the sequestered MOBIUS dataset. The segmentation masks generated by the participants' models were used to extract ROIs.

For the segmentation task, the participants were asked to submit two types of results to the organizers for scoring, i.e., (i) *binary masks*, where a thresholding procedure was already applied to the segmentation output by the participants, and (ii) *probabilistic predictions*, where each pixel in the segmentation map is represented on a scale between 0 and 1 and a value of 1 implies absolute certainty that the pixel belongs to the sclera region. A visual example of the requested submission material is shown in Figure 3. For the two recognition-oriented tasks, the participants had to submit complete recognition models, which were then evaluated by the organizers.

3.3. Evaluation protocol

SSRBC 2023 was implemented in two separate stages. In the first stage, the MASD dataset was provided to the participants along with the ground truth masks and labels for training their models for segmentation and recognition, respectively. In the second stage, the MOBIUS and SBVPI datasets were provided (barring the ground truth segmentation masks) to the participants. The timeframe for delivering the final result was four days from release of the two datasets to ensure that the participants could not manually inspect and label the data.



Figure 3. Illustration of results to be submitted (from left to right): original image, generated binary segmentation mask, probabilistic (grey-scale) segmentation prediction.

3.4. Performance metrics

The following segmentation (Sg) and recognition (Rc) measures and graphs were used for SSRBC 2013 to judge the performance of the submitted models:

- **Precision (Sg)**, defined as the ratio of pixels that are true positives or correctly classified for the sclera, to the total number of pixels classified as belonging to sclera, including the false positives, i.e., $(\frac{TP}{TP+FP})$ [25–28].
- **Recall (Sg)**, defined as the ratio of correctly classified pixels by the segmentation mask, to the overall number of ground truth sclera pixel, i.e. $\frac{TP}{TP+FN}$ [25–28].
- **F_1 score (Sg)**, defined as the harmonic mean of precision and recall, calculated as: $2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$. Some of the results may be biased towards better precision or recall. The F_1 score mitigates this bias by considering both and is, therefore, used as the primary performance indicator for the segmentation task in SSRBC 2023.
- **Intersection over union (IoU) (Sg)**, also termed as the Jaccard index, is the ratio between the intersection of the predicted and ground truth sclera pixels over their union. IoU is computed as $\frac{TP}{TP+FP+FN}$.
- **ROC curves (Rc)**, Receiver Operating Characteristics curves, that plot the verification rate against the false acceptance rate (FAR) for various values of the decision threshold. Here, the verification rate is defined as $1 - \text{FRR}$, where FRR stands for the false rejection rate. The Area Under the ROC curve is also reported as a performance measure for the verification experiments.
- **Equal Error Rate (EER) (Rc)**, defined as the operating point on the ROC curve, where the values of FAR and FRR are equal.

In the equations above, TP denotes *true positives*, i.e., the number of pixels correctly classified as sclera pixels, FP denotes *false positives*, i.e., the number of background pixels misclassified as sclera pixels, and FN denotes *false negatives*, i.e., the number of correct sclera pixels misclassified as background pixels. The results based on binary mask only show a partial picture of the overall performance for a certain decision threshold and may lead to biased conclusions. That is why complete *precision-recall curves* were generated by moving the threshold over the probabilistic segmentation masks [29, 30]. An optimal F_1 score

Table 2. Summary of participants and list of submitted approaches to SSRBC 2023. The table lists the abbreviations of the models, as used in the experimental section.

Segmentation Algorithms			
No.	Group [†]	Model Acronym	DL/Other
1.	School of Computer Science, Jilin University, China	Unet-VGG	DL
2.	Beijing University of Civil Engineering and Architecture (BUCEA)	Sclera-TransFuseCNN	DL
3.	Indian Institute of Technology, Mandi (IIT Mandi)	SegDeep+	DL
4.	Couger Inc., Japan and Vellore Institute of Technology (VIT Vellore)	Attention-Sclera-net	DL
5.	Fraunhofer Institute for Computer Graphic Research IGD	IGD-EyeMMS	DL
6.	Fraunhofer Institute for Computer Graphic Research IGD	IGD-U-Net	DL
Recognition Algorithms			
No.	Group	Model Acronym	DL/Other
1.	Beijing University of Civil Engineering and Architecture (BUCEA)	Res2Net	DL

[†]For details on the participants from the institutions see the author list.

(F_1^{opt}) was also calculated based on these curves, and ultimately the Area Under Curve *Area Under the precision-recall Curve* was included in the evaluation, as another performance metric [31].

4. Summary of submitted approaches

Five groups entered SSRBC 2023 and submitted 6 segmentation models and one recognition approach for scoring. Table 2 presents a summary of the participating groups, while a brief description of the submitted models is provided below.

4.1. Segmentation Algorithms

Unet-VGG uses the VGG model pretrained on ImageNet as the backbone feature extraction part of its Unet. The enhanced features extraction is done by using the five initial feature layers obtained from the pre-trained model for up-sampling and feature fusion using both, the cross entropy loss and the dice loss.

Sclera-TransFuseCNN uses a novel two-stream hybrid model to integrate ResNet-34 and Swin Transformer encoders [32] into a dual-encoder architecture for segmentation. The model employs an encoder-decoder structure. The dual encoders first separately extract (in a hierarchical manner) coarse- and fine-grained feature representations. Then a novel Cross-Domain Fusion module is introduced to efficiently fuse the multi-scale features extracted from the dual encoders. Finally, the fused features are progressively up-sampled and aggregated to predict the sclera masks in the decoding process. Additionally, deep supervision strategies are employed to learn intermediate feature representations better and faster.

SegDeep+ is based on the DeepLabV3plus [33] architec-

ture, which is widely used for semantic segmentation tasks. It consists of a feature extraction backbone network followed by an encoder-decoder structure. In this case, a pre-trained ResNet50 model is used as the backbone network, which has been trained on a large dataset to extract high-level features from images. The input images for the segmentation task are resized. After passing the input image through the ResNet50 backbone, the feature maps are extracted from the "conv4 block6 2 relu" layer. These feature maps contain rich spatial information. To capture multi-scale contextual information, the Dilated Spatial Pyramid Pooling layer is introduced. This layer performs average pooling on the feature maps to reduce their spatial dimensions. Then, different convolutional blocks with different dilation rates (1, 6, 12, 18) are applied to capture contextual information at multiple scales. These blocks have convolutional layers with appropriate dilation rates to expand the receptive field and capture both local and global contextual information. The output feature maps from the Dilated Spatial Pyramid Pooling layer and the convolutional blocks are concatenated and passed through another convolutional layer to refine the segmentation predictions.

Attention-Sclera-net draws inspiration from the AB Sclera Net architecture. It consists of four encoder and four decoder blocks. In the encoder part, DenseNet layers with down-sampling are employed, while the decoder utilizes up-sampling layers followed by a residual block with concurrent spatial and channel attention mechanisms. The encoder skip connection is passed through the CBAM block [34], and the output of the CBAM block is connected to the decoder block. The fourth encoder block is connected to the residual block and linked to decoder block 1. Furthermore, all decoder blocks undergo additional upsampling and are fused together at the end. The incorporation of supervision

Table 3. Comparative assessment of the segmentation algorithms on the MOBIUS dataset. The results are ordered according to the achieved F_1 scores. The F_1 , Precision, Recall and IoU scores were computed from the submitted binary masks. The optimal F_1 score on the precision-recall curve (F_1^{opt}) and AUC values were calculated from the probabilistic segmentation predictions.

Segmentation Model	Number of Parameters (in Millions)	From binary masks				From probabilistic predictions	
		F_1	Precision	Recall	IoU	F_1^{opt}	AUC
Sclera-TransFuseCNN	192.4	0.863213	0.854184	0.883867	0.768490	0.863649	0.818853
IGD-EyeMMS	22.7	0.820630	0.922743	0.763570	0.714149	0.822482	0.718979
IGD-U-Net	31.0	0.762797	0.933832	0.710583	0.664876	0.766242	0.705874
Unet-VGG	24.9	0.607015	0.888362	0.544279	0.504948	0.675423	0.687316
Attention-Sclera-net	7.1	0.518656	0.973090	0.430581	0.423650	0.722804	0.730540
SegDeep+	11.9	0.470149	0.688844	0.405713	0.324566	0.485030	0.409351

at different levels facilitates efficient learning by enabling gradient propagation and weight updates in these layers. Finally, the output of each level is concatenated along the channel axis, followed by the application of the CBAM attention block before the final supervision stage.

IGD-EyeMMS extends over Multi-scale segmentation solutions (Eye-MMS [35, 36]) by having an additional input to the first module of Eye-MMS from an encoder network (ResNet34 [37]). IGD-EyeMMS is a convolutional neural network (CNN) that consists of inter-connected refinement modules. Each module consists of two convolutional layers, each followed by layer normalization and a Leaky ReLU with non-linearity. The submitted IGD-EyeMMS is trained using the Intersection over Union (IoU) loss and the SGD optimizer with an initial learning rate of 1e-1 and batch size of 16 for 200 epochs. The learning rate is exponentially reduced by $\gamma = 0.99$ every epoch. The predicted segmentation is rounded to the nearest integer values to represent the discrete labels.

IGD-U-Net is based on the U-Net for brain segmentation [38]. The U-Net architecture consists of two parts, an encoder and a decoder. Each of these two parts is built using four levels of blocks, each containing two convolutional layers with batch normalization and ReLU activation function and one max pooling layer in the encoder and up-convolutional layers in the decoder. Details of the utilized architecture are described in [38]. The submitted IGD-U-Net model is trained using the Intersection over Union (IoU) loss and SGD optimizer with an initial learning rate of 1e-1 and batch size of 16 for 200 epochs. The learning rate is exponentially reduced by $\gamma = 0.99$ every epoch. The predicted segmentation is rounded to the nearest integer values to represent the discrete labels.

4.2. Recognition Algorithms

Res2Net adopts a Res2Net model [39] pre-trained on ImageNet as the backbone, and the pixel-by-pixel product of the sclera mask and the original ocular image is regarded as

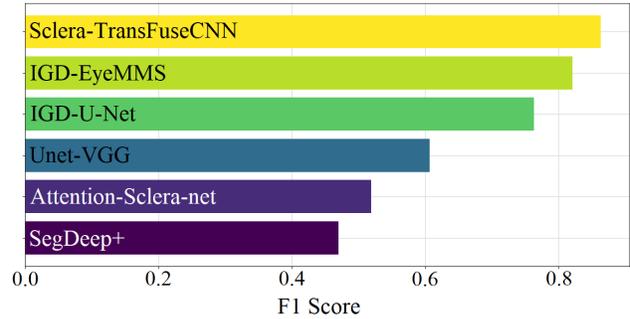


Figure 4. Performance comparison of the submitted models in terms of the F_1 score achieved over all test images from the MOBIUS dataset. Best viewed in colour.

the input for feature extraction. During training, the model is fine-tuned for sclera classification by using a regular cross entropy loss, where the provided (limited) training dataset is enriched with various augmentation strategies, such as horizontal flipping, rotation, and shifting. Next, the segmented sclera image is resized to 224×224 and the CLAHE algorithm is performed twice on the red, green and blue layers of the resized image to enhance the vessel structure. Finally, the image is fed into the model for training. After training, a feature vector is extracted from the penultimate layer of the model and is utilized as the identity representation of the input test image. The normalized Euclidean distance is applied to calculate the similarity of any two images. In order to obtain the score that an image belongs to a subject, the score between the image and all gallery images from the subject are calculated. The highest value of the scores is treated as the final comparison score. Therefore, the identity label of the subject with the highest score is assigned as the identity of the input image.

5. Benchmarking results and discussion

In this section, the results of SSRBC 2023 are presented. A comprehensive analysis is conducted to analyze the performance of the submitted segmentation models and study

their impact on the recognition accuracy.

5.1. Comparative assessment

We first evaluate the submitted segmentation models and analyze their performance in cross-sensor environments.

Results on binary segmentation masks. Binary segmentation masks represent the most common output of modern segmentation models. Typically, a thresholding procedure is utilized within the model that determines the trade-off between precision and recall, and, consequently, the final F_1 score. Because this is the default output of the majority of modern segmentation models, we first analyze the submitted binary mask generated by the SSBRC entries. The results of the comparison are presented in Table 3 and Figure 4.

It can be seen that the Sclera-TransFuseCNN is the top performer of the competition w.r.t. to recall, F1 and IoU scores. The second-best algorithm, IGD-EyeMMS, performed very closely to Sclera-TransFuseCNN, with recall, F1 and IoU scores slightly below the results of the best performing approach, however, with better precision. The other four models achieved somewhat lower F_1 , recall, and IoU scores. Nonetheless, the IGD-U-Net and Attention-Sclera-net models produced higher precision scores than the top performers of SSRBC 2023.

It is also interesting to observe that the performance of the models is not always correlated with their size, i.e., the number of parameters of the models. IGD-EyeMMS, for example, ensures better performance than IGD-U-Net with a lower number of parameters, whereas Sclera-TransFuseCNN with around 192.4 million parameters achieved comparable results as IGD-EyeMMS. Overall, we can conclude that most models produced solid results on the segmentation task, but also that there are considerable differences in the results between the top-performers and the rest, both in terms of performance characteristics as well as model size.

Results on probabilistic segmentation predictions. For better insight into the performance of the submitted segmentation models, we proceed to analyze them w.r.t. the probabilistic segmentation predictions. From the right part of Table 3 and the precision-recall curves in Figure 7, it can be observed that the two top-performing submission of SSBRC 2023, i.e., Sclera-TransFuseCNN and IGD-EyeMMS, also lead to the best results in terms of optimal F_1 score, i.e., F_1^{opt} . Moreover, given that the optimal F_1 scores are close to the ones observed with the binary masks, this also suggests that the two techniques are well calibrated.

Qualitative comparison. A qualitative comparison of the submitted models is shown in Figure 6 in terms of the binary masks. Three challenging samples from the MOBIUS dataset, where the acquisition environment is varying, are

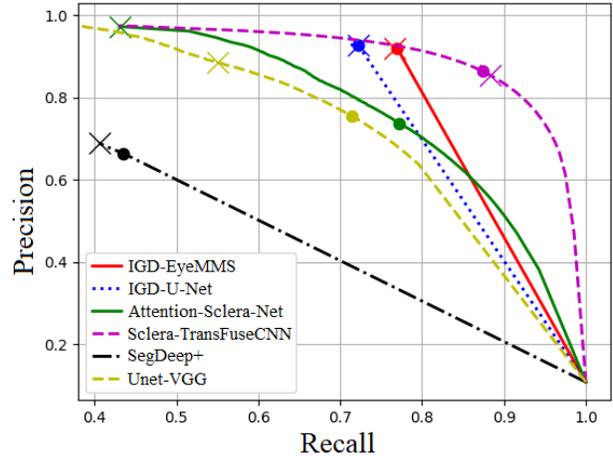


Figure 5. Precision-recall curves of the submitted models. The operating points denoted with a full circle represent the best possible F_1 score (F_1^{opt}), whereas the cross denotes the precision-recall point produced by the binary masks. The dotted lines denote the standard deviation. The figure is best viewed in colour and zoomed in.

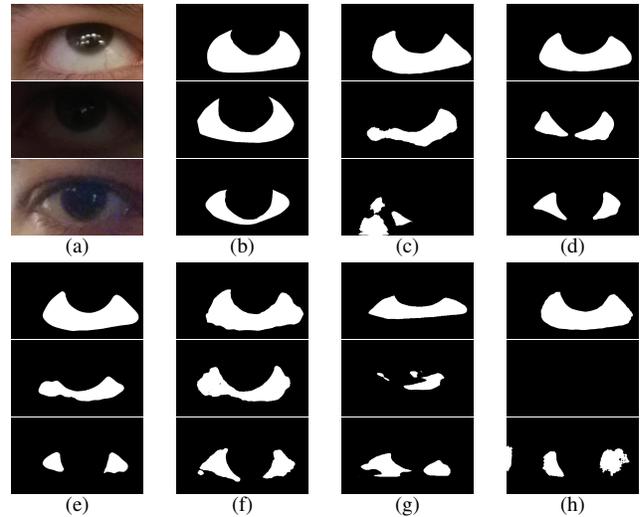


Figure 6. Qualitative comparison of the submitted models (in terms of binary masks) on selected MOBIUS images. Observe the difference in the segmentation quality across the evaluated models. The figure shows (a) the original image; (b) the ground truth mask; and the submitted binary masks from: (c) Attention-Sclera-Net, (d) IGD-EyeMMS, (e) IGD-U-Net, (f) Sclera-TransFuseCNN, (g) SegDeep+, (h) Unet-VGG.

presented to illustrate the capabilities of the submitted segmentation models. As can be seen, there are significant differences between the best and worst performing models in terms of overlap with the manually annotated ground truth masks, which is also reflected in quantitative scores in Table 3.

Table 4. Comparative assessment of two recognition algorithms applied on the segmented images for MOBIUS, obtained from each of the segmentation algorithms.

Segmentation Model	Task	Res2Net		SqueezeNet	
		AUC	EER	AUC	EER
Ground Truth	RT	0.948729	0.123418	0.920501	0.106319
Sclera-TransFuseCNN	SRT	0.944574	0.130192	0.910725	0.116070
IGD-EyeMMS	SRT	0.926692	0.151305	0.913165	0.117795
IGD-U-Net	SRT	0.921143	0.160622	0.902389	0.146742
Unet-VGG	SRT	0.863086	0.213110	0.838594	0.204368
Attention-Sclera-net	SRT	0.906672	0.164379	0.865137	0.165836
SegDeep+	SRT	0.853497	0.236792	0.820378	0.210503

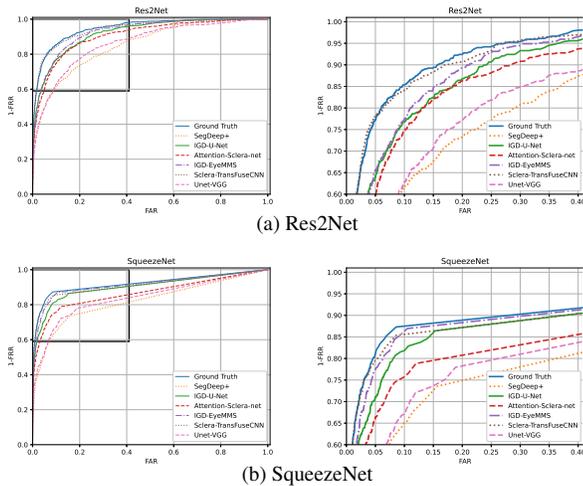


Figure 7. ROC curves generated for the (a) submitted Res2Net model and the (b) baseline SqueezeNet model. The figures illustrate the sensitivity of the recognition models on differences in segmentation performance in a joint evaluation setting.

5.2. Impact of segmentation on recognition performance

In this section, we analyze the impact of segmentation performance on the overall sclera-recognition pipeline. Because only the BUCEA group submitted a recognition approach for the RT and SRT tasks, we implement and train another baseline model, SqueezeNet [40], for the analysis and utilize it to put the results of the BUCEA Res2Net model into perspective. As can be seen from the results in Table 4 and Figure 7, the Res2Net model performs somewhat better than the lightweight SqueezeNet approach and yields higher AUC scores in all experiments, suggesting that larger models with better capacity are needed for competitive results - note that SqueezeNet has 0.76 and Res2Net 2.4 M parameters. Nonetheless, the baseline SqueezeNet model produced competitive EER scores and consistently outperforms Res2Net in this operating point. In the recognition tasks (RT), where the ground truth segmentation masks

were used on the SBVPI dataset, for example, the Res2Net model yielded an AUC score of a little above 0.94 and an equal error rate (EER) of 0.123, whereas SqueezeNet resulted in an AUC score of 0.92, but a slightly lower EER of 0.106.

To further analyze the sensitivity of the recognition models to changes in the quality of the segmentation masks, we ran verification experiments on a subset of the MOBIUS dataset and investigated the performance of the Res2Net and SqueezeNet models with all 6 submitted sets of segmentation masks. From the reported results, it can be clearly seen that in most cases better segmentation also leads to better recognition performance. Additionally, it can be observed that Sclera-TransFuseCNN i.e. the best submission for the segmentation task, ensured very similar performance to the one observed with the manually generated ground truth masks, both in terms of AUC and EER. It is also worth noting that recognition performance ensured by the Attention-Sclera-net segmentation is better than that of Unet-VGG, despite the fact that the latter exhibited stronger segmentation results. The reason behind this is assumed to be the higher precision attained by Attention-Sclera-net during segmentation on the MOBIUS dataset. However, it is questionable whether such performance can also generalize to other datasets given the lower recall rate. In other words, it can be assumed that other information in addition to the sclera pattern is also retrieved and considered during the recognition stage when the segmentation masks generated by Attention-Sclera-net are used.

6. Conclusion

The eight edition of the Sclera Segmentation Benchmarking Competition (SSRBC 2023) was organized in conjunction with IJCB 2023 to benchmark and record the recent developments and performance of sclera segmentation and recognition models with varying sensors, gaze angles, and acquisition conditions. A total of 5 groups from 6 institutions participated in the competition and contributed 6 segmentation techniques and 1 recognition model for the group evaluation. The submitted models ensured solid segmentation and recognition results. It can be concluded that better segmentation can lead to better recognition results. In future we will aim at investigating the impact of covariates such as lighting condition and gaze angle on the SRT.

Acknowledgements

The authors are grateful to CodaLab Competitions [41] for providing a platform to conduct the competition.

The authors also gratefully acknowledge the computing time provided on the high performance computing facility, Sharanga, at the Birla Institute of Technology and Science - Pilani, Hyderabad Campus.

References

- [1] Abhijit Das, Umapada Pal, Michael Blumenstein, and Miguel Angel Ferrer Ballester. Sclera recognition—a survey. In *2013 2nd IAPR Asian Conference on Pattern Recognition*, pages 917–921. IEEE, 2013. [1](#)
- [2] Peter Rot, Matej Vitek, Klemen Grm, Žiga Emeršič, Peter Peer, and Vitomir Štruc. Deep sclera segmentation and recognition. In *Handbook of vascular biometrics*, pages 395–432. Springer, Cham, 2020. [1](#), [2](#), [3](#)
- [3] Matej Vitek, Peter Rot, Vitomir Štruc, and Peter Peer. A comprehensive investigation into sclera biometrics: a novel dataset and performance study. *Neural Computing and Applications*, pages 1–15, 2020. [1](#), [3](#)
- [4] Zhi Zhou, Eliza Yingzi Du, N. Luke Thomas, and Edward J. Delp. A new human identification method: Sclera recognition. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 42(3):571–583, 2011. [1](#)
- [5] Daniel Riccio, Nadia Brancati, Maria Frucci, and Diego Gragnaniello. An unsupervised approach for eye sclera segmentation. In *Iberoamerican Congress on Pattern Recognition*, pages 550–557. Springer, 2017. [1](#)
- [6] Abhijit Das, Umapada Pal, Miguel Angel Ferrer, and Michael Blumenstein. A framework for liveness detection for direct attacks in the visible spectrum for multimodal ocular biometrics. *Pattern Recognition Letters*, 82:232–241, 2016. [1](#)
- [7] Abhijit Das, Umapada Pal, Miguel A. Ferrer Ballester, and Michael Blumenstein. Sclera recognition using dense-sift. In *2013 13th International Conference on Intelligent Systems Design and Applications*, pages 74–79. IEEE, 2013. [1](#)
- [8] Sinan Alkassar, Wai Lok Woo, Satnam Singh Dlay, and Jonathon A Chambers. Robust sclera recognition system with novel sclera segmentation and validation techniques. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(3):474–486, 2015. [1](#)
- [9] Abhijit Das, Prabir Mondal, Umapada Pal, Miguel Angel Ferrer, and Michael Blumenstein. Fast and efficient multimodal eye biometrics using projective dictionary pair learning. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages 1402–1408. IEEE, 2016. [1](#)
- [10] Petru Radu, James Ferryman, and Peter Wild. A robust sclera segmentation algorithm. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6. IEEE, 2015. [1](#)
- [11] Sinan Alkassar, Wai-Lok Woo, Satnam Dlay, and Jonathon Chambers. Sclera recognition: on the quality measure and segmentation of degraded images captured under relaxed imaging conditions. *IET Biometrics*, 6(4):266–275, 2016. [1](#)
- [12] Abhijit Das, Rituraj Kunwar, Umapada Pal, Miguel A. Ferrer, and Michael Blumenstein. An online learning-based adaptive biometric system. In *Adaptive Biometric Systems*, pages 73–96. Springer, 2015. [1](#)
- [13] Abhijit Das, Umapada Pal, Miguel A. Ferrer Ballester, and Michael Blumenstein. A new efficient and adaptive sclera recognition system. In *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, pages 1–8. IEEE, 2014. [1](#)
- [14] Abhijit Das, Umapada Pal, Miguel A. Ferrer, and Michael Blumenstein. A decision-level fusion strategy for multimodal ocular biometric in visible spectrum based on posterior probability. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 794–798. IEEE, 2017. [1](#)
- [15] Vikas Gottemukkula, Sashi Saripalle, Sriram P. Tankasala, and Reza Derakhshani. Method for using visible ocular vasculature for mobile biometrics. *IET Biometrics*, 5(1):3–12, 2016. [1](#)
- [16] Abhijit Das, Prabir Mondal, Umapada Pal, Michael Blumenstein, and Miguel A. Ferrer. Sclera vessel pattern synthesis based on a non-parametric texture synthesis technique. In *Proceedings of international conference on computer vision and image processing*, pages 241–250. Springer, 2017. [1](#)
- [17] Abhijit Das, Umapada Pal, Miguel A. Ferrer, and Michael Blumenstein. SSBC 2015: Sclera Segmentation Benchmarking Competition. In *Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pages 742–747, 2015. [2](#)
- [18] Abhijit Das, Umapada Pal, Miguel A. Ferrer, and Michael Blumenstein. SSRBC 2016: sclera segmentation and recognition benchmarking competition. In *International Conference on Biometrics (ICB)*, pages 1–6, 2016. [2](#)
- [19] Abhijit Das, Umapada Pal, Miguel A. Ferrer, Michael Blumenstein, Dejan Štepec, Peter Rot, Žiga Emeršič, Peter Peer, Vitomir Štruc, and S.A. Kumar. SSERBC 2017: Sclera segmentation and eye recognition benchmarking competition. In *International Joint Conference on Biometrics (IJCB)*, pages 742–747, 2017. [2](#)
- [20] Abhijit Das, Umapada Pal, Miguel A. Ferrer, Michael Blumenstein, Dejan Štepec, Peter Rot, Peter Peer, and Vitomir Štruc. SSBC 2018: Sclera Segmentation Benchmarking Competition. In *International Conference on Biometrics (ICB)*, pages 303–308, 2018. [2](#)
- [21] Abhijit Das, Umapada Pal, Michael Blumenstein, Caiyong Wang, Yong He, Yuhao Zhu, and Zhenan Sun. Sclera segmentation benchmarking competition in cross-resolution environment. In *IAPR International Conference on Biometrics. IEEE*, 2019. [2](#)
- [22] Matej Vitek, Abhijit Das, Yann Pourcenoux, Alexandre Missler, Calvin Paumier, Sumanta Das, Ishita De Ghosh, Diego R. Lucio, Luiz A. Zanlorensi Jr., David Menotti, Fadi Boutros, Naser Damer, Jonas Henry Grebe, Arjan Kuijper, Junxing Hu, Yong He, Caiyong Wang, Hongda Liu, Yunlong Wang, Zhenan Sun, Daile Osorio-Roig, Christian Rathgeb, Christoph Busch, Juan Tapia Farias, Andres Valenzuela, Georgios Zampoukis, Lazaros Tsochatzidis, Ioannis Pratikakis, Sabari Nathan, R Suganya, Vineet Mehta, Abhinav Dhall, Kiran Raja, Gourav Gupta, Jalil Nourmohammadi Khirak, Mohsen Akbari-Shahper, Farhang Jaryani, Meysam Asgari-Chenaghlu, Ritesh Vyas, Sristi Dakshit, Sagnik Dakshit, Peter Peer, Umapada Pal, and Vitomir Štruc. SSBC

- 2020: Sclera segmentation benchmarking competition in the mobile environment. In *IEEE International Joint Conference on Biometrics (IJCB)*, 2020. 2
- [23] Matej Vitek, Abhijit Das, Diego Rafael Lucio, Luiz Antonio Zanolensi, David Menotti, Jalil Nourmohammadi Khiarak, Mohsen Akbari Shahpar, Meysam Asgari-Chenaghlu, Farhang Jaryani, Juan E Tapia, et al. Exploring bias in sclera segmentation models: A group evaluation approach. *IEEE Transactions on Information Forensics and Security*, 18:190–205, 2022. 2
- [24] Abhijit Das, Umapada Pal, Miguel A. Ferrer Ballester, and Michael Blumenstein. Multi-angle based lively sclera biometrics at a distance. In *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, pages 22–29. IEEE, 2014. 2
- [25] Žiga Emeršič, Luka Lan Gabriel, Vitomir Štruc, and Peter Peer. Pixel-wise ear detection with convolutional encoder-decoder networks. *IET Biometrics*, 2017. 4
- [26] Peter Rot, Žiga Emeršič, Vitomir Štruc, and Peter Peer. Deep multi-class eye segmentation for ocular biometrics. In *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOB)*, pages 1–8. IEEE, 2018. 4
- [27] Juš Lozej, Blaž Meden, Vitomir Štruc, and Peter Peer. End-to-end iris segmentation using U-Net. In *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOB)*, pages 1–6. IEEE, 2018. 4
- [28] Juš Lozej, Dejan Štepec, Vitomir Štruc, and Peter Peer. Influence of segmentation on deep iris recognition performance. In *2019 IEEE International Work Conference on Bioinspired Intelligence (IWOB)*, pages 1–6, 2019. 4
- [29] David Martin Powers. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2:37–63, 2011. 4
- [30] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLoS one*, 10(3):e0118432, 2015. 4
- [31] Kendrick Boyd, Kevin H. Eng, and C. David Page. Area under the precision-recall curve: point estimates and confidence intervals. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 451–466. Springer, 2013. 5
- [32] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 5
- [33] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, 2018. 5
- [34] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018. 5
- [35] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Eye-mms: Miniature multi-scale segmentation network of key eye-regions in embedded applications. In *2019 IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019, Seoul, Korea (South), October 27-28, 2019*, pages 3665–3670. IEEE, 2019. 6
- [36] Fadi Boutros, Naser Damer, Kiran B. Raja, Raghavendra Ramachandra, Florian Kirchbuchner, and Arjan Kuijper. Iris and periocular biometrics for head mounted displays: Segmentation, recognition, and synthetic data generation. *Image Vis. Comput.*, 104:104007, 2020. 6
- [37] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016. 6
- [38] Mateusz Buda, Ashirbani Saha, and Maciej A Mazurowski. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in Biology and Medicine*, 109, 2019. 6
- [39] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):652–662, 2019. 6
- [40] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016. 8
- [41] Adrien Pavao, Isabelle Guyon, Anne-Catherine Letournel, Xavier Baró, Hugo Escalante, Sergio Escalera, Tyler Thomas, and Zhen Xu. Codalab competitions: An open source platform to organize scientific challenges. *Technical report*, 2022. 8