Fully Convolutional Networks and Generative Adversarial Networks Applied to Sclera Segmentation

Diego R. Lucio[1], Rayson Laroca[1], Evair Severo[1], Alceu S. Britto Jr.[2], David Menotti[1]

[1]Federal University of Paraná (UFPR), Curitiba, PR, Brazil

[2]Pontifical Catholic University of Paraná (PUCPR), Curitiba, PR, Brazil

{drlucio,rblsantos,ebsevero,menotti}@inf.ufpr.br alceu@ppgia.pucpr.br

Abstract

Due to the world's demand for security systems, biometrics can be seen as an important topic of research in computer vision. One of the biometric forms that has been gaining attention is the recognition based on sclera. The initial and paramount step for performing this type of recognition is the segmentation of the region of interest, i.e. the sclera. In this context, two approaches for such task based on the Fully Convolutional Network (FCN) and on Generative Adversarial Network (GAN) are introduced in this work. FCN is similar to a common convolution neural network, however the fully connected layers (i.e., the classification layers) are removed from the end of the network and the output is generated by combining the output of pooling layers from different convolutional ones. The GAN is based on the game theory, where we have two networks competing with each other to generate the best segmentation. In order to perform fair comparison with baselines and quantitative and objective evaluations of the proposed approaches, we provide to the scientific community new 1,300 manually segmented images from two databases[1]. The experiments are performed on the UBIRIS.v2 and MICHE databases and the best performing configurations of our propositions achieved F-score's measures of 87.48% and 88.32%, respectively.

1. Introduction

In recent years, the interest in using biometrics to automatically identify and/or verify a person's identity has greatly increased. Many characteristics can be used to identify a person, such as physical, biological and behavioral traits [7, 32]. Biometrics are especially important as they can not be changed, forgotten, lost or stolen, providing an unquestionable connection between the individual and the application that makes use of it [4].

Several characteristics of the human body can be used as biometrics, such as fingerprint, face, iris, retina and voice, each one with its advantages and disadvantages. The iris and retina are among the most accurate biometrics [7]. However, biometric systems based on iris and retina having high degree of reliability require, respectively, user collaboration and an intrusive image acquisition scheme [8]. In addition to the iris and retina biometric traits, the eye has also a white region around the eyeball known as sclera that contains a pattern of blood vessels that can be used for personal identification [6, 9, 11].

Typically, segmentation is the first step in which efforts should be applied in a reliable sclera-based recognition system. Incorrect segmentation can either reduce the region of blood vessels or introduce new patterns such as eyelashes and eyelids, impairing the system effectiveness.

In order to avoid the above mentioned problems and to encourage the creation of new sclera segmentation techniques, some competitions were performed [6, 8, 9]. The state of the art method in these competitions was obtained using a neural network technique named autoencoder in the Multi-Angle Sclera Dataset.v1 (MASD.v1) database. When images of a single sensor were used, the best attained recall and precision rates were 96.65% and 95.64%, respectively.

In this work, we proposed two new approaches to sclera segmentation based on Convolutional Neural Networks (CNNs), one based on Fully Convolutional Network (FCN) [31] and another one based on Generative Adversarial Network (GAN) [16]. To the best of our knowledge, both kinds of networks have never been studied in the sclera segmentation scenario. FCN is used for segmentation in a large range of applications, from medical to satellite image analysis [14, 29], while GAN is a new approach to semantic segmentation, which has outperformed the state of the art [21]. The results yielded by the proposed approaches outperform the ones of previous works on a subject of UBIRIS.v2 database [25]. This subject contains 201 images kindly provided by the authors of [12] and

---

more 300 ones manually labeled by us. We also present promising results for sclera segmentation on a subset (1000 images) of the Mobile Iris Challenge Evaluation I (MICHE-I) database [22], which was never been used in this context and initially proposed for iris segmentation and recognition from mobile images. For producing a fair and quantitative comparison among proposed approaches and a baseline one (SegNet [3]), we manually labeled 1,300 images from those databases, making the masks publicly available for research purposes. Regarding the MASD.v1 database which was the focus of the previous mentioned competitions for sclera segmentation, we do not have yet implementations on Matlab as required by the organizers of those competitions.

The main contributions of this paper can be summarized as follows: 1) Two new approaches for sclera segmentation; 2) A comparative evaluation of the proposed approaches with a baseline one; 3) Two datasets composed of 1,300 sclera images manually labeled, being 1,000 from the MICHE-I database and 300 from the UBIRIS.v2 database.

The remainder of this paper is organized as follows: we briefly review the related work in Section 2. In Section 3, the proposed segmentation approach is described. Section 4 and 5 present the experiments and the results obtained, respectively. Finally, conclusions and future work are discussed in Section 6.

2. Related Work

In this section, we present a review of the most relevant studies in the sclera segmentation context. We start by presenting some relevant sclera recognition methods in which different sclera segmentation techniques were proposed. Finally, we describe some important works specially dedicated to sclera segmentation techniques.

2.1. Sclera-based Recognition Methods

The works presented in this subsection employed sclera segmentation as preprocessing for sclera-based recognition. Note that, in such cases, the authors did not report the [precision, recall, F-score] achieved in the segmentation stage.

Zhou et al. [32] proposed a new concept for human identification based on the pattern of sclera vessels. In their work, they presented a fully automated sclera segmentation approach for both color and grayscale images. In color images, the sclera region is estimated using the best representation between two color-based techniques. On the other hand, the Otsu's thresholding method is applied to find the sclera region in grayscale images. The UBIRIS.v1 [24] and IUPUI [32] databases were used in the experiments.

Das et al. [7] presented a methodology where the right and left sclera are segmented separately. The time adaptive active contour-based region growing segmentation technique proposed by Chan & Vese [5] was employed. The authors applied the Daugman's integro-differential operator

to find the seed point required in the region growing-based segmentation [10]. The UBIRIS.v1 database was used in the experiments.

Delna et al. [11] presented a sclera identification based on a single-board computer. The sclera region is segmented as a rectangle from the Hough circular transform applied for iris location. All images used in the experiments were obtained from a webcam connected to a Raspberry Pi.

2.2. Sclera Segmentation Techniques

Das et al. [8] presented a benchmark for sclera segmentation where four research groups proposed their solutions for this task. All approaches were evaluated in the MASD.v1 database, proposed in the competition. The best results were obtained by Team 4 [27], where the authors presented a novel sclera segmentation algorithm for color images which operates at pixel-level. Exploring various color spaces, the proposed approach is robust to image noise and different gaze directions. The algorithm's robustness is enhanced by a two-stage classifier. At the first stage, a set of simple classifiers is employed, while at the second stage, a neural network classifier operates on the probabilities space generated by the classifier at the first stage. The reported precision and recall rates were 95.05% and 94.56%, respectively, on the MASD.v1 database.

Das et al. [9] proposed a new benchmark, which addresses sclera segmentation and recognition. The best segmentation results reached 85.21% precision and 80.21% recall. This result is achieved using a method based on Fuzzy C Means, which considers spatial information and uses Gaussian kernel function to calculate the distance between the center of the cluster and the data points.

Alkassar et al. [1] proposed a segmentation algorithm which fuses multiple color space skin classifiers to overcome noise introduced by sclera acquisition, such as motion, blur, gaze and rotation. This approach was evaluated using the UBIRIS.v1, UBIRIS.v2 [25] and UTIRIS [15] databases.

Das et al. [6] presented a new sclera segmentation benchmark where seven research groups proposed their solutions for this task. The best results attained precision and recall rates of 95.34% and 96.65%, respectively. These results were obtained by using a neural network architecture based on the encoder-decoder approach called SegNet [3].

2.3. Final Remarks

As one may see, the use of CNNs was not much explored in the challenging sclera segmentation task. Thus, one of the contributions of this paper is to explore this aspect in the attempt of obtaining improvements in terms of sclera segmentation accuracy.

3. Proposed Approach

Some images might present specular highlights in regions of the subject's face. In our preliminary tests, many of these regions were erroneously classified as sclera. Therefore, we propose to first locate the periocular region and then perform the sclera segmentation in the detected patch.

This section describes the proposed approach and it is divided into two subsections, one for Periocular Region Detection (PRD) and one for sclera segmentation.

3.1. Periocular Region Detection

YOLO [28] is a object detection framework based on CNN, which regards detection as a regression problem. As great advances were recently achieved through YOLO-inspired models [18, 23], we decided to fine-tune it for PRD. However, as we want to detect only one class and the computational cost is one of our main concerns, we chose to use a smaller model, called Fast-YOLO[2] [28], which uses fewer convolutional layers than YOLO and fewer filters in those layers. The Fast-YOLO's architecture is shown in Table 1.

Table 1. Fast-YOLO network used to detect the periocular region. We reduced the number of filters in the last convolutional layer from 125 to 30 in order to output 1 class instead of 20.

	Layer	Filters	Size	Input	Output
0	conv	16	$3 \times 3/1$	$416 \times 416 \times 1/3$	$416 \times 416 \times 16$
1	max		$2 \times 2/2$	$416 \times 416 \times 16$	$208 \times 208 \times 16$
2	conv	32	$3 \times 3/1$	$208 \times 208 \times 16$	$208 \times 208 \times 32$
3	max		$2 \times 2/2$	$208 \times 208 \times 32$	$104 \times 104 \times 32$
4	conv	64	$3 \times 3/1$	$104 \times 104 \times 32$	$104 \times 104 \times 64$
5	max		$2 \times 2/2$	$104 \times 104 \times 64$	$52 \times 52 \times 64$
6	conv	128	$3 \times 3/1$	$52 \times 52 \times 64$	$52 \times 52 \times 128$
7	max		$2 \times 2/2$	$52 \times 52 \times 128$	$26 \times 26 \times 128$
8	conv	256	$3 \times 3/1$	$26 \times 26 \times 128$	$26 \times 26 \times 256$
9	max		$2 \times 2/2$	$26 \times 26 \times 256$	$13 \times 13 \times 256$
10	conv	512	$3 \times 3/1$	$13 \times 13 \times 256$	$13 \times 13 \times 512$
11	max		$2 \times 2/1$	$13 \times 13 \times 512$	$13 \times 13 \times 512$
12	conv	1024	$3 \times 3/1$	$13 \times 13 \times 512$	$13 \times 13 \times 1024$
13	conv	1024	$3 \times 3/1$	$13 \times 13 \times 1024$	$13 \times 13 \times 1024$
14	conv	30	$1 \times 1/1$	$13 \times 13 \times 1024$	$13 \times 13 \times 30$
15	detection				

The PRD network is trained using the images, without any preprocessing, and the coordinates of the Region of Interest (ROI) as inputs. As ground truth, we used the annotations provided by Severo et al. [30]. As these annotations were made for iris location, we applied a padding (chosen based on the validation set) in the detected patch (i.e., iris), so that the sclera is entirely within the ROI.

By default, Fast-YOLO only returns objects detected with a confidence of 0.25 or higher. We consider only the detection with the largest confidence in cases where more than one periocular region is detected, since there is always only one region annotated in the evaluated databases. If no

region is detected, the next stage (sclera segmentation) is performed on the image in its original size.

3.2. Sclera Segmentation

We employ three approaches for sclera segmentation, since they presented good results in other segmentation applications. These approaches are: FCN, Encoder-Decoder (ED) and GAN. Its noteworthy that ED was employed for sclera segmentation in [6], obtaining state-of-the-art results. Therefore, we made use of ED in the databases used in this paper, considering it as the baseline for comparison with the proposed approach.

3.2.1 Fully Convolutional Network

This segmentation approach was proposed by Long et al. [20]. The network has only convolutional layers and the segmentation process can take input images of arbitrary sizes, producing correspondingly-sized output with efficient inference and learning.

In this work, we employ the FCN approach presented by Teichmann et al. [31]. As shown in Figure 1, features are extracted using a CNN without the fully connected layers (i.e., VGG-16 without the last 3 layers).
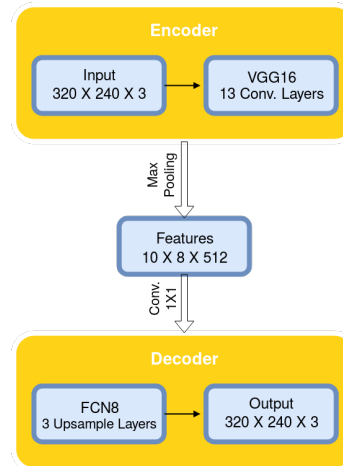

Figure 1. FCN architecture for sclera segmentation.

Next, the extracted features pass through two 1×1 convolutional layers, generating an output of dimension $10 \times 8 \times 6$. The output of these convolutional layers is processed by the FCN8 architecture proposed in [20], which performs the up-sampling combining the last three layers from the VGG-16.

3.2.2 Encoder-Decoder

A convolutional ED, also called autoencoder, is a neural network trained in order to copy its input to the output. The

---

Table 2. SegNet architecture.

Layer		Filters	Size	Input	Output	Layer		Filters	Size	Input	Output
1	enc	64	3×3	$320 \times 240 \times 3/1$	$320 \times 240 \times 64$	19	up		2×2	$10 \times 8 \times 512$	$20 \times 15 \times 512$
2	enc	64	3×3	$320 \times 240 \times 64$	$320 \times 240 \times 64$	20	dec	512	3×3	$20 \times 15 \times 512$	$20 \times 15 \times 512$
3	max		2×2	$320 \times 240 \times 64$	$160 \times 120 \times 64$	21	dec	512	3×3	$20 \times 15 \times 512$	$20 \times 15 \times 512$
4	enc	128	3×3	$160 \times 120 \times 64$	$160 \times 120 \times 128$	22	dec	512	3×3	$20 \times 15 \times 512$	$20 \times 15 \times 512$
5	enc	128	3×3	$160 \times 120 \times 128$	$160 \times 120 \times 128$	23	up		2×2	$20 \times 15 \times 512$	$40 \times 30 \times 512$
6	max		2×2	$160 \times 120 \times 128$	$80 \times 60 \times 128$	24	dec	512	3×3	$40 \times 30 \times 512$	$40 \times 30 \times 512$
7	enc	256	3×3	$80 \times 60 \times 128$	$80 \times 60 \times 256$	25	dec	512	3×3	$40 \times 30 \times 512$	$40 \times 30 \times 512$
8	enc	256	3×3	$80 \times 60 \times 256$	$80 \times 60 \times 256$	26	dec	256	3×3	$40 \times 30 \times 512$	$40 \times 30 \times 256$
9	enc	256	3×3	$80 \times 60 \times 256$	$80 \times 60 \times 256$	27	up		2×2	$40 \times 30 \times 256$	$80 \times 60 \times 256$
10	max		2×2	$80 \times 60 \times 256$	$40 \times 30 \times 256$	28	dec	256	3×3	$80 \times 60 \times 256$	$80 \times 60 \times 256$
11	enc	512	3×3	$40 \times 30 \times 256$	$40 \times 30 \times 512$	29	dec	256	3×3	$80 \times 60 \times 256$	$80 \times 60 \times 256$
12	enc	512	3×3	$40 \times 30 \times 512$	$40 \times 30 \times 512$	30	dec	128	3×3	$80 \times 60 \times 256$	$80 \times 60 \times 128$
13	enc	512	3×3	$40 \times 30 \times 512$	$40 \times 30 \times 512$	31	up		2×2	$80 \times 60 \times 512$	$160 \times 120 \times 128$
14	max		2×2	$40 \times 30 \times 512$	$20 \times 15 \times 512$	32	dec	128	3×3	$160 \times 120 \times 128$	$160 \times 120 \times 128$
15	enc	512	3×3	$20 \times 15 \times 512$	$20 \times 15 \times 512$	33	dec	64	3×3	$160 \times 120 \times 128$	$160 \times 120 \times 64$
16	enc	512	3×3	$20 \times 15 \times 512$	$20 \times 15 \times 512$	34	up		2×2	$160 \times 120 \times 64$	$320 \times 240 \times 64$
17	enc	512	3×3	$20 \times 15 \times 512$	$20 \times 15 \times 512$	35	dec	64	3×3	$320 \times 240 \times 64$	$320 \times 240 \times 64$
18	max		2×2	$20 \times 15 \times 512$	$10 \times 8 \times 512$	36	dec	2	3×3	$320 \times 240 \times 64$	$320 \times 240 \times 2$

purpose is to learn data encoding which can be used for dimensionality reduction or even for file compression [2].

The ED (SegNet) used in this work was presented in [3]. SegNet consists of a stack of encoders followed by a corresponding stack of decoders which feed a soft-max classification layer. Decoders map low-resolution features extracted by encoders to an image with the same dimension as the input. The architecture used is presented in Table 2.

3.2.3 Generative Adversarial Network

GANs are deep neural networks composed by both generator and discriminator networks, pitting one against the other. First, the generator network receives noise as input and generates samples. Then, the discriminator network receives samples of training data and those of the generator network, being able to distinguish between the two sources [13]. A generic GAN architecture is shown in Figure 2.
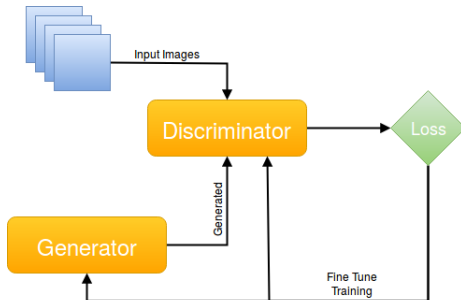

Figure 2. GAN architecture for sclera segmentation.

Basically, the generator network learns to produce more realistic samples throughout each iteration, and the discriminator network learns to better distinguish real and synthetic

datas.

Isola et al. [16] presented the GAN approach used in this work, which is a conditional GAN able to learn the relation between a image and its label file, and, from that, generate a variety of image types, which can be employed in various tasks such as photo generation and semantic segmentation.

4. Experiments

In this section, we present the databases and the evaluation protocol used in our experiments.

4.1. Databases

The experiments were carried out in two subsets of well-known iris databases: UBIRIS.v2 and MICHE-I. An overview of both subsets can be seen in Table 3. Remark that we do not use the SSBC [8] and SSRBC [9] databases in our experiments as only the test sets were made available by the authors.

Table 3. Overview of the databases used in this work. All of these are a subset of the original database.

Database	Images	Subjects	Resolution
UBIRIS.v2	500	261	400×300
MICHE-I	1,000	92	Various
MICHE-GS4	333	92	Various
MICHE-IP5	323	92	Various
MICHE-GT2	344	92	640×480

UBIRIS.v2: this database is composed of 11,102 images collected from both eyes from 261 subjects and have a resolution of 400×300 pixels.

MICHE-I: this database consists of 3,191 images captured from 92 subjects under uncontrolled settings using three

mobile devices: iPhone 5, Galaxy Samsung IV and Galaxy Tablet II (1,262, 1,297 and 632 images, respectively), with many different resolutions [22].

4.2. Preprocessing

As discussed in Section 3.1, it was necessary to first detect the periocular region as some images present specular highlights, impairing the performance of the system. After detecting the periocular region, only the ROI was maintained in each image, providing a great improvement over the results obtained at first. Figure 3 shows an example of each subset of the original image (without PRD) and the segmentation mask created by us.

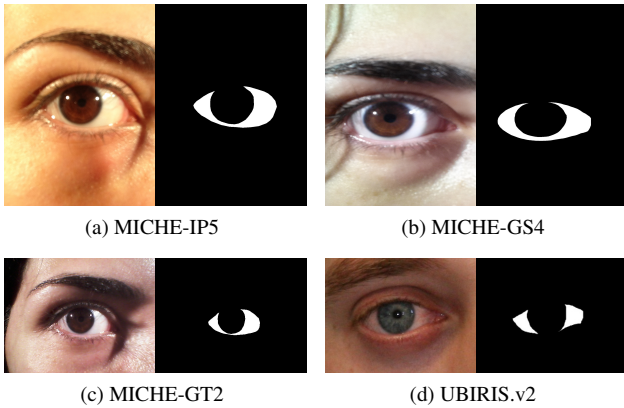(a) MICHE-IP5 (b) MICHE-GS4

(c) MICHE-GT2 (d) UBIRIS.v2

Figure 3. Four examples of the masks created by us.

Figure 4 shows, instead, four cropped images after PRD and their respective masks. It is noteworthy that most specular highlights are removed after PRD.

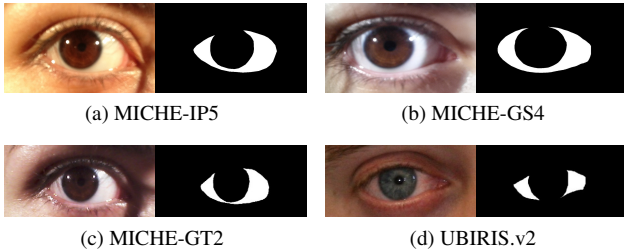(a) MICHE-IP5 (b) MICHE-GS4

(c) MICHE-GT2 (d) UBIRIS.v2

Figure 4. Periocular regions detected and replicated to the masks.

At last, the ROI is resized according to each approach proposed in Section 3. The input sizes were chosen based on the original architectures of the approaches (see Table 4).

Table 4. Image dimensions used in each approach.

Approach	Image - Dimension	Mask - Dimension
FCN	$320 \times 240 \times 3$	$320 \times 240 \times 1$
GAN	$256 \times 256 \times 3$	$256 \times 256 \times 3$
SegNet	$320 \times 240 \times 3$	$320 \times 240 \times 1$

4.3. Evaluation Protocol

The performance evaluation of an automatic segmented mask is performed in a pixel-to-pixel comparison between the ground truth and the predicted image. Therefore, we use the following metrics: Precision, Recall and F-score.

To perform a fair evaluation and comparison of the proposed approaches in all databases, we divided each into three subsets, being 40% of the images for training, 40% for testing and 20% for validation.

5. Results and Discussions

The experiments were carried out using the protocol presented in Section 4.3. We also performed some additional experiments using the cross-sensor methodology.

5.1. Proposed Protocol

The results obtained by both the baseline (SegNet) and the proposed approaches are shown in Table 5. The baseline presented considerably worse results. We believe this is due to the size of the training set, since SegNet was originally employed in a large dataset [27]. Radu et al. [27] generated a dataset with 54,000 images using data augmentation. However, we did not have access to the database used by them, and thus a more direct comparison with their methodology was not possible to be done.

Table 5. Results achieved using the proposed protocol.

Database	Approach	Recall %	Precision %	F-score %
UBIRIS.v2	GAN	87.48 ± 08.19	87.10 ± 08.16	86.82 ± 05.88
	SegNet	72.48 ± 17.15	87.52 ± 08.53	77.82 ± 13.08
	FCN	$\mathbf{87.31 \pm 06.68}$	$\mathbf{88.45 \pm 06.98}$	$\mathbf{87.48 \pm 03.90}$
MICHE-I	GAN	87.07 ± 10.81	86.39 ± 12.02	86.27 ± 09.97
	SegNet	64.59 ± 24.73	83.39 ± 18.53	69.87 ± 22.34
	FCN	$\mathbf{87.59 \pm 11.28}$	$\mathbf{89.90 \pm 09.82}$	$\mathbf{88.32 \pm 09.80}$
MICHE-GS4	GAN	85.72 ± 12.53	86.12 ± 13.01	85.20 ± 11.31
	SegNet	66.50 ± 26.34	76.09 ± 23.80	67.92 ± 23.87
	FCN	$\mathbf{88.24 \pm 12.03}$	$\mathbf{88.65 \pm 10.62}$	$\mathbf{88.12 \pm 10.56}$
MICHE-IP5	GAN	88.11 ± 07.40	87.71 ± 07.71	87.42 ± 05.43
	SegNet	31.90 ± 26.05	79.40 ± 32.93	40.95 ± 29.19
	FCN	$\mathbf{87.51 \pm 11.61}$	$\mathbf{89.32 \pm 05.22}$	$\mathbf{87.80 \pm 08.24}$
MICHE-GT2	GAN	86.20 ± 15.02	83.81 ± 15.73	84.50 ± 14.28
	SegNet	73.77 ± 21.20	76.46 ± 18.29	72.33 ± 18.26
	FCN	$\mathbf{87.86 \pm 12.23}$	$\mathbf{88.50 \pm 12.68}$	$\mathbf{87.94 \pm 11.59}$

Better results were obtained using the proposed approaches. In the UBIRIS.v2 subset, the GAN-based sclera segmentation attained a F-score value of 86.82% (±5.88), while the approach based on FCN achieved 87.48% (±3.90). Although there is little difference between the F-score values obtained by both methods, the standard deviation presented when using FCN was slightly lower than when GAN was employed for the segmentation.

The same happened in all subsets used in our experiments, fact that makes us believe that the FCN approach

is best suited for sclera segmentation. However, the results obtained with the GAN-based segmentation should not be diminished, since they were very close to the best results.

Here we perform a visual analysis. For this task, we randomly chose an image from the UBIRIS.v2 subset. Figures 5a, 5b and 5c demonstrate a very poor outcome in the segmentation of the sclera. As can be seen, the FCN approach presented a considerably better segmentation result when compared to the baseline and GAN. The same occurs in many other images, but the results are not always so discrepant (see Figures 5d, 5e and 5f). It is noteworthy the consistency presented with FCN-based segmentation technique, observed in all sclera images generated in this work.
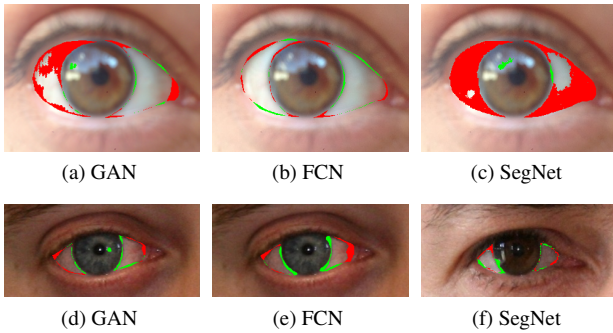

(a) GAN   (b) FCN   (c) SegNet

(d) GAN   (e) FCN   (f) SegNet

Figure 5. Samples of segmented scleras using the ground truth for highlighting errors: green and red pixels represent the FPs and FNs respectively.

5.2. Additional Experiments Using Cross-Sensor

In this section, we present the results obtained using a cross-sensor methodology, where two experiments were performed. In the first one, we used MICHE-I (see Table 3) as training set and UBIRIS.v2 as test set. In the second experiment, we inverted the order and used UBIRIS.v2 as training set and MICHE-I as test set.

Table 6 presents the results obtained using MICHE-I as training set and UBIRIS.v2 as test set. As we can see, the obtained F-score was very close to that obtained when the training and test sets were from the same database, reaching a F-score value 1% higher. However, in this case the best segmentation was achieved with the GAN-based approach.

Table 6. Results obtained using MICHE-I as training set and UBIRIS.v2 as test set.

Approach	Recall %	Precision %	F-score %
GAN	**90.02 ± 05.46**	**85.96 ± 07.90**	**87.52 ± 03.74**
SegNet	83.98 ± 08.85	82.64 ± 08.40	82.58 ± 05.35
FCN	87.92 ± 05.28	87.04 ± 07.11	87.14 ± 03.53

The same did not happen when we used UBIRIS.v2 as training set and MICHE-I as test set, as shown in Table 7. The attained F-score values were much lower than those obtained when MICHE-I was used as training.

Table 7. Results obtained using UBIRIS.v2 as training set and MICHE-I as test set.

Approach	Recall %	Precision %	F-score %
GAN	68.45 ± 19.13	71.05 ± 21.57	67.98 ± 18.41
SegNet	27.77 ± 22.43	60.58 ± 37.42	30.36 ± 22.79
FCN	**74.99 ± 20.14**	**77.41 ± 16.15**	**73.40 ± 17.07**

This might have occurred because the subset of the MICHE-I database used in this work has more diversity and it is larger than the UBIRIS.v2 subset, which allows the generated model to better discriminate the pixels belonging to sclera.

6. Conclusions and Future Work

This work introduced two new approaches for sclera segmentation and compared them with a baseline (SegNet) method chosen in the literature. Both proposed approaches (FCN and GAN) attained higher precision and recall values in all evaluated scenarios. Furthermore, these approaches presented promising results when evaluated in cross-sensor scenarios.

We also labeled $1,300$ images for sclera segmentation. These masks (manually labeled) are publicly available to the research community, assisting in the fair comparison among published works.

There is still room for improvements in sclera segmentation, so we intend to: 1) design new and better network architectures; 2) create a unique architecture that integrates the detection stage of the periocular region; 3) employ a post-processing stage to refine the segmentation given by the proposed approaches; 4) design a general and independent sensor approach, where firstly the image sensor is classified and then the sclera is segmented with a specific approach; 5) compare the proposed approaches with methods applied in other domains such as iris segmentation [17, 19] and periocular-based recognition [26].

References

[1] S. Alkassar, W. L. Woo, S. S. Dlay, and J. A. Chambers. Enhanced segmentation and complex-sclera features for human recognition with unconstrained visible-wavelength imaging. In *2016 International Conference on Biometrics (ICB)*, pages 1–8, June 2016.

[2] N. Audebert, B. Le Saux, and S. Lefèvre. Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. In *Computer Vision – ACCV 2016*, pages 180–196, 2017.

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, Dec 2017.

[4] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior. *Guide to Biometrics*. Springer Publishing Company, Incorporated, 2004.

[5] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, Feb 2001.

[6] A. Das et al. SSERBC 2017: Sclera segmentation and eye recognition benchmarking competition. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 742–747, Oct 2017.

[7] A. Das, U. Pal, M. A. F. Ballester, and M. Blumenstein. Sclera recognition using dense-SIFT. In *2013 13th International Conference on Intellient Systems Design and Applications*, pages 74–79, Dec 2013.

[8] A. Das, U. Pal, M. A. Ferrer, and M. Blumenstein. SSBC 2015: Sclera segmentation benchmarking competition. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6, Sept 2015.

[9] A. Das, U. Pal, M. A. Ferrer, and M. Blumenstein. SSRBC 2016: Sclera segmentation and recognition benchmarking competition. In *2016 International Conference on Biometrics (ICB)*, pages 1–6, June 2016.

[10] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, Nov 1993.

[11] K. V. Delna, K. A. Sneha, and R. P. Aneesh. Sclera vein identification in real time using single board computer. In *2016 International Conference on Next Generation Intelligent Systems (ICNGIS)*, pages 1–5, Sept 2016.

[12] J. Fialho Pinheiro, J. D. Sousa de Almeida, G. Braz Junior, A. Cardoso de Paiva, and A. Corrêa Silva. Sclera segmentation in face images using image foresting transform. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 229–236, 2018.

[13] I. Goodfellow et al. Generative adversarial networks. In *27th Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[14] C. Henry, S. M. Azimi, and N. Merkle. Road segmentation in SAR satellite images with deep fully-convolutional neural networks. *CoRR*, abs/1802.01445, 2018.

[15] M. S. Hosseini, B. N. Araabi, and H. Soltanian-Zadeh. Pigment melanin: Pattern for iris recognition. *IEEE Trans. on Instrumentation and Measurement*, 59(4):792–804, 2010.

[16] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016.

[17] A. Lakra, P. Tripathi, R. Keshari, M. Vatsa, and R. Singh. SegDenseNet: Iris segmentation for pre and post cataract surgery. *CoRR*, abs/1801.10100, 2018.

[18] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Gonçalves, W. R. Schwartz, and D. Menotti. A robust real-time automatic license plate recognition based on the YOLO detector. *CoRR*, abs/1802.09567, 2018.

[19] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan. Accurate iris segmentation in non-cooperative environments using fully convolutional networks. In *2016 International Conference on Biometrics (ICB)*, pages 1–8, June 2016.

[20] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, June 2015.

[21] P. Luc, C. Couprie, S. Chintala, and J. Verbeek. Semantic segmentation using adversarial networks. *CoRR*, abs/1611.08408, 2016.

[22] M. Marsico, M. Nappi, D. Riccio, and H. Wechsler. Mobile iris challenge evaluation (MICHE)-i, biometric iris dataset and protocols. *Pat. Rec. Letters*, 57:17–23, 2015.

[23] G. Ning, Z. Zhang, C. Huang, X. Ren, H. Wang, C. Cai, and Z. He. Spatially supervised recurrent convolutional neural networks for visual object tracking. In *IEEE International Symposium on Circuits and Systems*, pages 1–4, 2017.

[24] H. Proença and L. A. Alexandre. UBIRIS: A noisy iris image database. In *Image Analysis and Processing – ICIAP*, pages 970–977, 2005.

[25] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre. The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1529–1535, Aug 2010.

[26] H. Proença and J. C. Neves. Deep-prwis: Periocular recognition without the iris and sclera using deep learning frameworks. *IEEE Transactions on Information Forensics and Security*, 13(4):888–896, April 2018.

[27] P. Radu, J. Ferryman, and P. Wild. A robust sclera segmentation algorithm. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6, Sept 2015.

[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, June 2016.

[29] H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, and K. Mori. An application of cascaded 3d fully convolutional networks for medical image segmentation. *Computerized Medical Imaging and Graphics*, 66:90–99, jun 2018.

[30] E. Severo, R. Laroca, C. S. Bezerra, L. A. Zanlorensi, D. Weingaertner, G. Moreira, and D. Menotti. A benchmark for iris location and a deep learning detector evaluation. *CoRR*, abs/1803.01250, 2018.

[31] M. Teichmann, M. Weber, J. M. Zöllner, R. Cipolla, and R. Urtasun. Multinet: Real-time joint semantic reasoning for autonomous driving. *CoRR*, abs/1612.07695, 2016.

[32] Z. Zhou, E. Y. Du, N. L. Thomas, and E. 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, May 2012.