

# Robust Iris Segmentation Based on Fully Convolutional Networks and Generative Adversarial Networks

Cides S. Bezerra\*, Rayson Laroca\*, Diego R. Lucio\*, Evair Severo\*,

Lucas F. Oliveira\*, Alceu S. Britto Jr.<sup>†</sup> and David Menotti\*

\*Federal University of Paraná (UFPR), Curitiba, PR, Brazil

<sup>†</sup>Pontifical Catholic University of Paraná (PUCPR), Curitiba, PR, Brazil

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

**Abstract**—The iris can be considered as one of the most important biometric traits due to its high degree of uniqueness. Iris-based biometrics applications depend mainly on the iris segmentation whose suitability is not robust for different environments such as near-infrared (NIR) and visible (VIS) ones. In this paper, two approaches for robust iris segmentation based on Fully Convolutional Networks (FCNs) and Generative Adversarial Networks (GANs) are described. Similar to a common convolutional network, but without the fully connected layers (i.e., the classification layers), an FCN employs at its end a combination of pooling layers from different convolutional layers. Based on the game theory, a GAN is designed as two networks competing with each other to generate the best segmentation. The proposed segmentation networks achieved promising results in all evaluated datasets (i.e., BioSec, CasiaI3, CasiaT4, IITD-1) of NIR images and (NICE-I, CrEye-Iris and MICHE-I) of VIS images in both non-cooperative and cooperative domains, outperforming the baselines techniques which are the best ones found so far in the literature, i.e., a new state of the art for these datasets. Furthermore, we manually labeled 2,431 images from CasiaT4, CrEye-Iris and MICHE-I datasets, making the masks available for research purposes.

## I. INTRODUCTION

The identification of individuals based on their biological and behavioral characteristics has a higher degree of reliability compared to other means of identification, such as passwords or access cards. Several characteristics of the human body can be used for person recognition (e.g., face, signature, fingerprints, iris, sclera, retina, voice, etc.) [1]. The characteristics present in the iris make it one of the most representative and safe biometric modalities. This circular diaphragm forming the textured portion of the eye is capable of distinguishing individuals with a high degree of uniqueness [2], [3].

As described in [4], an automated biometric system for iris recognition is composed of four main steps: (i) image acquisition, (ii) iris segmentation, (iii) normalization and (iv) feature extraction and matching. The segmentation consists of locating and isolating the iris from other regions (e.g., the sclera, surrounding skin regions, etc.), therefore it is the most critical and challenging step of the system. Incorrect segmentation usually affects the subsequent steps, impairing the system performance [5].

Over the last decade, many approaches have been employed for iris segmentation, such as those based on edge detection [6], Hough transform [7], active contours [8], [9], integro-differential equation [10], Maximum Radial Suppression (MRS) [11], Markovian Texture Models (MTMs) [12], and Convolutional Neural Networks (CNNs) [13], [14] (see Section II for more details).

Leveraging the advent of CNNs we propose two approaches for iris segmentation task. The first is based on a Fully Convolutional Network (FCN) [15] and the second one is based on a Generative Adversarial Network (GAN) [16]. FCNs are used for segmentation in many different tasks since medical image analysis to aerospace image analysis [17], [18], while GAN is a young approach to semantic segmentation, which has outperformed the state of the art [19].

The proposed FCN and GAN iris segmentation approaches outperform three existing frameworks in the largest benchmark datasets found in the literature. There are two main contributions in this paper: (i) two CNN-based approaches that work well for near-infrared (NIR) and visible (VIS) images in both cooperative (highly controlled) and non-cooperative environments; and (ii) 2,431 new manually labeled masks from images of three existing iris datasets<sup>1</sup> (see Section IV-A).

The remainder of this paper is organized as follows: we briefly review related work in Section II. In Section III, the proposed approaches used for iris segmentation are described. Section IV presents the datasets, evaluation protocol and baselines used in the experiments. We report and discuss the results in Section V. Conclusions are given in Section VI.

## II. RELATED WORK

In this section, we briefly review relevant studies in the context of iris segmentation, which use from conventional image processing to deep learning techniques. For other studies on iris segmentation, please refer to [20], [21].

Jillela and Ross [22] presented an overview of classical approaches, evaluation methods and challenges related to

<sup>1</sup>The new masks are publicly available to the research community at <http://web.inf.ufpr.br/vri/databases/iris-segmentation-annotations/>.

iris segmentation in both NIR and VIS images. Daugman’s study [23] is considered the pioneer in iris segmentation. The integro-differential operator was used to approximate the boundary of the inner and outer iris, generating the central coordinates and both pupil and iris radius.

Liu et al. [6] first detected the inner boundary of the iris and then the outer boundary. In addition, noisy pixels were eliminated based on their high/low-intensity level. Proença and Alexandre [7] used the Fuzzy K-means algorithm to classify each pixel as belonging to a group, considering its coordinates and intensity distribution. Then, they applied the Canny edge detector in the image with the grouped pixels, creating an edge map. Finally, the inner and outer iris boundaries are detected by the circular Hough transform.

Shah and Ross [9] performed iris segmentation through Geodesic Active Contours, combining energy minimization with active contours based on curve evolution. The pupil is detected from a binarization and both inner and outer iris boundaries are approximated using the Fourier series coefficients.

The winning approach of the Noisy Iris Challenge Evaluation - Part I (NICE.I), proposed by Tan et al. [10], removes the reflection points using adaptive thresholding and bilinear interpolation. Region growing based on clustering and integro-differential constellation segments the iris. Podder et al. [11] applied an MRS technique to noise removal. Moreover, they applied the Canny edge detector and Hough transform to detect iris boundaries.

Haindl & Krupička [12] detected the iris using the Daugman’s operator [23] and removed the eyelids employing a third-order polynomial mean and standard deviation estimates. Adaptive thresholding and MTM were used to remove iris reflection. Ouabida et al. [8] applied the Optical Correlation based Active Contours (OCAC), that uses the Vander Lugt correlator algorithm, to detect the iris and pupil contours through spatial filtering.

Liu et al. [14] proposed two approaches called Hierarchical Convolutional Neural Networks (HCNNs) and Multi-scale Fully Convolutional Networks (MFCNs) to perform a dense prediction of the pixels using sliding windows, merging shallow and deep layers.

At present, CNNs are being employed to solve many computer vision problems with impressive results being obtained in several areas such as biometrics, medical imaging and security systems [24]–[26]. Teichmann et al. [27] proposed a CNN architecture, called MultiNet, to joint detection, classification and semantic segmentation. Inspired by the great results reported in their work, we apply the segmentation decoder of the MultiNet to the iris segmentation context, as detailed in Section III-B.

### III. PROPOSED APPROACH

This section describes the proposed approach and it is divided into two subsections, one for iris location and one for iris segmentation.

#### A. Iris Detection

The datasets used in this work have many different sizes, and just resizing the images would generate a distortion in the iris format. In order to avoid this distortion, we first performed the Periocular Region Detection (PRD).

YOLO [28] is a real-time object detection system, which regards detection as a regression problem. As great advances were recently attained through models inspired by YOLO [26], [29], we decided to fine-tune it for PRD. However, as we want to detect only one class (i.e., the iris), we chose to use a smaller model, called Fast-YOLO<sup>2</sup> [28], which uses fewer convolutional layers than YOLO and fewer filters in those layers. The Fast-YOLO’s architecture is shown in Table I.

TABLE I  
FAST-YOLO NETWORK USED FOR IRIS DETECTION.

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 was trained using the images, without any preprocessing, and the coordinates of the Region of Interest (ROI) as inputs. The annotations provided by Severo et al. [26] were used as ground truth. We applied a small padding in the detected patch to increase the chance that the iris is entirely within the ROI. Afterward, we enlarged the ROI to a square form with width and height that are power of 2.

By default, only objects detected with a confidence of 0.25 or higher are returned by Fast-YOLO [28]. We consider only the detection with the largest confidence in cases where more than one iris region is detected, since there is always only one region annotated in the evaluated datasets. If no region is detected, the next stage (iris segmentation) is performed on the image in its original size.

In our previous work on sclera segmentation [30], this same approach was used for iris detection.

#### B. Iris Segmentation

We chose FCN and GAN for iris segmentation since they presented good results in other segmentation applications [30]. These results can be explained by the fact that FCN has no fully connected layer which generally causes loss of spatial information, while the representations embodied by the pair

<sup>2</sup>For training Fast-YOLO we used the weights pre-trained on ImageNet, available at <https://pjreddie.com/darknet/yolo/>.

of networks in a GAN model (the generator and the discriminator) are able to capture the statistical distribution of training data, making possible less reliance on huge, well-balanced, and well-labelled datasets.

1) *Fully Convolutional Networks (FCNs)*: are deep neural networks in which an image is provided as input and a mask is generated at the output. This mask is a binary image (of the same size) where each pixel is classified as iris or not iris. Basically, we employed the MultiNet [27] segmentation decoder without the classification and detection decoders. The encoder consists of the first 13 layers of the VGG-16 network [31]. The features extracted from its fifth pooling layer were then used by the segmentation decoder, which follows the FCN architecture [32] (see Fig. 1).

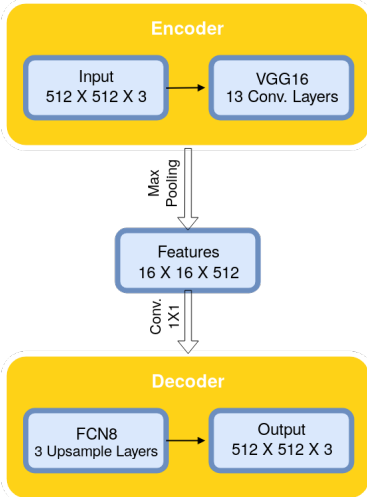


Fig. 1. FCN architecture for iris segmentation.

The fully-connected layers of the VGG-16 network were transformed into  $1 \times 1$  convolutional layers to produce a low-resolution segmentation. Then, three transposed convolution layers were used to perform up-sampling. Finally, high-resolution features were extracted through skip layers from lower layers to improve the up-sampled results.

The segmentation loss function was based on the cross-entropy. The pre-trained VGG-16 weights on ImageNet were used to initialize the encoder, the segmentation decoder, and the transposed convolutional layers. The training is based on the Adam optimizer algorithm [33], with the following parameters: learning rate of  $10^{-5}$ , dropout probability of 0.5, weight decay of  $5^{-4}$  and standard deviation of  $10^{-4}$  to initialize the skip layers.

2) *Generative Adversarial Networks (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 [34]. The GAN architecture for iris segmentation is shown in Fig. 2.

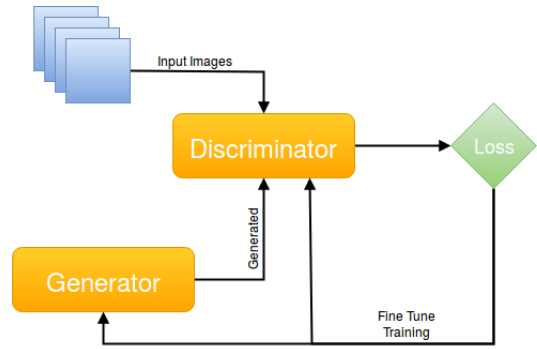


Fig. 2. GAN architecture for iris segmentation.

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

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

#### IV. EXPERIMENTS

In this section, we present the datasets, evaluation protocol and baselines used in our experiments for comparison of results and discussions.

##### A. Datasets

The experiments were carried out on well-known and challenging publicly available iris datasets with both NIR and VIS images having different sizes and characteristics. An overview of the number of images from each dataset is presented in Table II. The ground truths of the BioSec, CasiaI3 and IITD Iris Image Database 1.0 (IITD-1) datasets were provided by Hofbauer et al. [35]. In the following, details of the datasets are presented.

TABLE II  
OVERVIEW OF THE IRIS DATASETS USED IN THIS WORK, WHERE (\*) MEANS THAT ONLY PART OF THE DATASET WAS USED.

Dataset	Images	Subjects	Resolution	Wavelength
BioSec [36] (*)	400	25	$640 \times 480$	NIR
CasiaI3 [37]	2,639	249	$320 \times 280$	NIR
CasiaT4 [38] (*)	1,000	50	$640 \times 480$	NIR
IITD-1 [39]	2,240	224	$320 \times 240$	NIR
NICE.I [40]	945	n/a	$400 \times 300$	VIS
CrEye-Iris [41] (*)	1,000	120	$400 \times 300$	VIS
MICHE-I [42] (*)	1,000	75	Various	VIS

**BioSec**: a multimodal dataset [36] containing fingerprint, frontal face and iris images, as well as voice utterances. The entire dataset has 3,200 NIR iris images from 25 subjects with resolution of  $640 \times 480$  pixels, however, due to the available segmentation masks, we use only the first 400 images.

**CASIA-Iris-Interval-v3 (CasiaI3)**: a dataset [37] with 2,639 NIR iris images from 249 subjects with extremely clear iris texture details and resolution of  $320 \times 280$  pixels, acquired in an indoor environment.

**CASIA-Iris-Thousand-v4 (CasiaT4)**: a dataset [38] containing 20,000 NIR images from 1,000 subjects, collected in an indoor environment with different lightings setups. For our experiments, we manually labeled the first 1,000 images from 50 subjects.

**IITD-1**: a dataset [39] with 2,240 NIR images acquired from 224 subjects between 14-55 years comprising of 176 males and 48 females. All images have a resolution of  $320 \times 240$  pixels and were obtained in an indoor environment.

**Cross-Spectral Iris/Periocular (CrEye-Iris)**: a dataset composed of 3,840 images from 120 subjects [41]. The images were captured with a dual spectrum sensor (NIR and VIS) and divided into three subsets: iris, masked periocular and ocular images. We manually labeled the first 1,000 VIS images from the iris subset.

**Mobile Iris Challenge Evaluation I (MICHE-I)**: a dataset [42] with 3,191 VIS 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). The images have resolution of  $1536 \times 2048$ ,  $2320 \times 4128$  and  $640 \times 480$  pixels, respectively. We used the 569 ground truth masks made available by Hu et al. [43] and labeled another 431 to complete 1,000 images from 75 subjects.

**NICE.I**: a subset of the UBIRIS.v2 dataset [44]. The NICE.I [40] subset is composed of 500 images for training and 500 for testing. However, the test set provided by the organizers of the NICE.I contest has only 445 images. The subjects of the test set were not directly specified.

Fig. 3 shows two samples (NIR and VIS) of the masks created. We sought to eliminate all noise present in the iris, such as reflections and eyelashes.

### B. Evaluation protocol

A pixel-to-pixel comparison between the ground truth (manually labeled) and the algorithm prediction (i.e., the mask/segmentation) generate an average segmentation error  $E$  computed as a pixel divergence, given by the exclusive-or logical operator  $\otimes$  (i.e., XOR) [40], denoted by

$$E = \frac{1}{h \times w} \sum_i \sum_j M_k(i, j) \otimes GT_k(i, j), \quad (1)$$

where  $i$  and  $j$  are the coordinates in the mask  $M$  and ground truth  $GT$  images,  $h$  and  $w$  stand for the height and width of the image, respectively. Lower and higher  $E$  values represent better and worse results, respectively. We also reported the F-Measure (F1) measure which is a harmonic average of Precision and Recall [13].

In order to perform a fair evaluation and comparison of the proposed methodologies to the baselines in all datasets, we

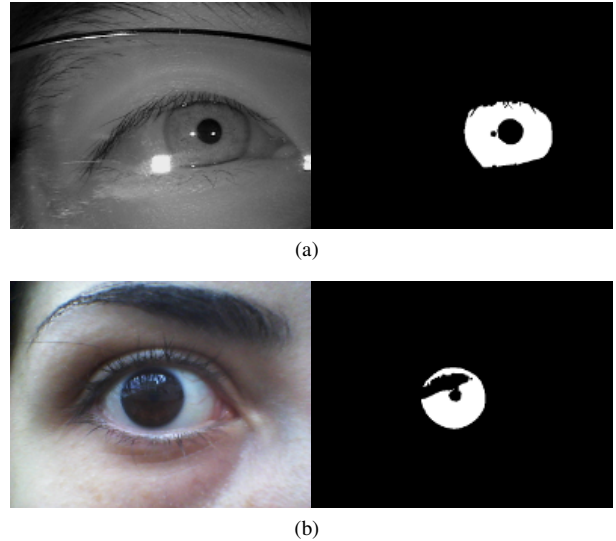


Fig. 3. Two examples of the masks created by us. (a) shows a NIR image (CasiaT4) and (b) a VIS image (MICHE-I).

randomly divided each dataset into two subsets, containing 80% of the images for training and the remainder for evaluation. The stopping learning criteria was 32,000 iterations.

As suggested in [27], we trained the FCN with 16,000 iterations. However, we noticed that the more iterations, the better was the model’s performance. Therefore, we doubled the number of iterations (i.e., 32,000) to ensure a good convergence of the model. According to our evaluations, 32,000 iterations were sufficient for all datasets.

### C. Benchmarks

We selected three baseline frameworks described (and available) in the literature to compare with our approaches with: Open Source Iris Recognition System Version 4.1 (OSIRISv4.1), Iris Segmentation Framework (IRIS-SEG) and Haindl & Krupička [12].

The **OSIRISv4.1** [45] framework is composed of four key modules: segmentation, normalization, feature extraction and matching. Nevertheless, we used only the segmentation module to compare it with our method. Although the performance of this framework was only reported in datasets with NIR images, we applied it on both NIR and VIS image datasets. This framework has input parameters such as minimum/maximum iris diameter. For a fair comparison, we tuned the parameters for each dataset in order to obtain the best results.

The **IRISSEG** [46] framework was designed specifically for non-ideal irises and is based on adaptive filtering, following a coarse-to-fine strategy. The authors emphasize that this approach does not require adjustment of parameters for different datasets. As in OSIRISv4.1, we report the performance of this framework on both NIR and VIS images.

The **Haindl & Krupička** [12] framework was used to evaluate the results achieved by the proposed approach on VIS datasets. This method was developed for colored eyes images obtained through mobile devices and used as the baseline in

the MICHE-II [47] contest. We did not report the Haindl & Krupička [12] performance on NIR images datasets since it was not possible to generate the segmentation masks using the executable provided by the authors.

## V. RESULTS AND DISCUSSIONS

The experiments were performed using two protocols: the protocol of the NICE.I contest and the one proposed in Section IV-B. Moreover, in order to analyze the robustness among sensors from the same environment (i.e., NIR or VIS) of the proposed FCN and GAN approaches, they were training using either all NIR or VIS image datasets and then evaluated on the same scenario. Finally, a visual and qualitative analysis showing some good and poor results is performed.

We report the mean F1 and  $E$  values by averaging the values obtained for each image. For all the experiments, we also carried out a statistical paired t-test with significance level of  $\alpha = 0.05$  between pairs of results for the same image, aiming to claim (statistical) significative difference between the results compared.

### A. The NICE.I Contest

The comparison of the results obtained by our approaches and those obtained by the baselines when using the NICE.I contest protocol is shown in Table III. As can be seen, the IRISSEG and OSIRISv4.1 frameworks presented the worst results. They achieved F1 values of 21.76% and 30.70% on the NICE.I test set, respectively. These results might be explained because these frameworks were developed for NIR images. Therefore, their performances are drastically compromised in VIS images. It is noteworthy that the distribution of F1 values for both frameworks presented high standard deviation (approximately  $\pm 32\%$ ). This occurs because, in some images, the False Positives (FPs) were high in both frameworks, including images that do not have iris, resulting in a very poor segmentation.

TABLE III  
IRIS SEGMENTATION RESULTS USING THE NICE.I CONTEST PROTOCOL.

Dataset	Method	F1 %	$E$ %
NICE.I (VIS)	OSIRISv4.1 [45]	30.70 $\pm$ 32.00	08.67 $\pm$ 06.29
	IRISSEG [46]	21.76 $\pm$ 32.13	14.03 $\pm$ 12.33
	Haindl & Krupička [12]	75.54 $\pm$ 22.93	03.27 $\pm$ 04.29
	<b>FCN Proposed</b>	<b>88.20 <math>\pm</math> 13.73</b>	<b>01.05 <math>\pm</math> 00.86</b>
	<b>GAN Proposed</b>	<b>91.42 <math>\pm</math> 03.81</b>	<b>03.09 <math>\pm</math> 01.76</b>

We expected to obtain good results using the Haindl & Krupička [12] framework, due to the fact that it was developed for VIS images and it was used for generating the reference masks (i.e., the ground truth) of the MICHE-I dataset in the recognition contest (MICHE-II). However, according to our experiments, its performance was not promising, although it obtained better results than IRISSEG and OSIRISv4.1.

The proposed FCN and GAN approaches achieved considerably better mean values for F1 and  $E$  metrics than the other approaches. We believe that these results were attained due

to the discriminating power of the deep learning approaches and also because our models were adjusted (i.e., trained) specifically for each dataset. We emphasize that OSIRISv4.1 was also adjusted for each dataset.

Although higher standard deviation of F1 was presented for the FCN approach, the paired t-test has shown that the GAN approach presented a statistically better F1 value, however, the FCN approach has presented a statistically smaller  $E$  value.

### B. Our protocol

We trained and tested the FCN and GAN approaches on each dataset to compare them with the benchmarks. Table IV shows the results obtained when using the proposed evaluation protocol (see Section IV-B).

TABLE IV  
IRIS SEGMENTATION RESULTS USING THE PROPOSED PROTOCOL.

Dataset	Method	F1 %	$E$ %
BioSec (NIR)	OSIRISv4.1 [45]	92.62 $\pm$ 03.19	01.21 $\pm$ 00.47
	IRISSEG [46]	93.94 $\pm$ 05.88	01.06 $\pm$ 01.20
	<b>FCN Proposed</b>	<b>97.46 <math>\pm</math> 00.74</b>	<b>00.44 <math>\pm</math> 00.12</b>
	<b>GAN Proposed</b>	<b>96.82 <math>\pm</math> 02.83</b>	<b>00.74 <math>\pm</math> 01.40</b>
CasiaI3 (NIR)	OSIRISv4.1 [45]	89.49 $\pm$ 05.78	05.35 $\pm$ 02.40
	IRISSEG [46]	94.61 $\pm$ 03.28	02.85 $\pm$ 01.62
	<b>FCN Proposed</b>	<b>97.90 <math>\pm</math> 00.68</b>	<b>01.15 <math>\pm</math> 00.37</b>
	<b>GAN Proposed</b>	<b>96.13 <math>\pm</math> 05.35</b>	<b>01.45 <math>\pm</math> 03.71</b>
CasiaT4 (NIR)	OSIRISv4.1 [45]	87.76 $\pm$ 08.01	01.34 $\pm$ 00.64
	IRISSEG [46]	91.39 $\pm$ 08.13	00.95 $\pm$ 00.54
	<b>FCN Proposed</b>	<b>94.42 <math>\pm</math> 07.54</b>	<b>00.61 <math>\pm</math> 00.58</b>
	<b>GAN Proposed</b>	<b>95.38 <math>\pm</math> 03.72</b>	<b>01.40 <math>\pm</math> 00.93</b>
IITD-1 (NIR)	OSIRISv4.1 [45]	92.20 $\pm$ 06.07	04.37 $\pm$ 02.69
	IRISSEG [46]	94.25 $\pm$ 03.89	03.39 $\pm$ 02.16
	<b>FCN Proposed</b>	<b>97.44 <math>\pm</math> 01.78</b>	<b>01.48 <math>\pm</math> 01.01</b>
	<b>GAN Proposed</b>	<b>95.84 <math>\pm</math> 04.13</b>	<b>01.33 <math>\pm</math> 02.65</b>
NICE.I (VIS)	OSIRISv4.1 [45]	38.15 $\pm$ 33.61	07.92 $\pm$ 06.20
	IRISSEG [46]	28.64 $\pm$ 35.14	13.48 $\pm$ 12.36
	Haindl & Krupička [12]	70.59 $\pm$ 26.11	04.72 $\pm$ 05.87
	<b>FCN Proposed</b>	<b>89.54 <math>\pm</math> 13.79</b>	<b>01.00 <math>\pm</math> 00.70</b>
	<b>GAN Proposed</b>	<b>91.12 <math>\pm</math> 05.08</b>	<b>03.34 <math>\pm</math> 02.31</b>
CrEye-Iris (VIS)	OSIRISv4.1 [45]	46.53 $\pm$ 29.25	13.22 $\pm$ 06.33
	IRISSEG [46]	61.72 $\pm$ 33.55	10.58 $\pm$ 10.38
	Haindl & Krupička [12]	76.81 $\pm$ 23.73	05.69 $\pm$ 04.58
	<b>FCN Proposed</b>	<b>97.04 <math>\pm</math> 01.21</b>	<b>00.96 <math>\pm</math> 00.36</b>
<b>GAN Proposed</b>	<b>92.61 <math>\pm</math> 05.86</b>	<b>03.02 <math>\pm</math> 03.22</b>	
MICHE-I (VIS)	OSIRISv4.1 [45]	33.85 $\pm$ 35.86	01.99 $\pm$ 02.90
	IRISSEG [46]	19.34 $\pm$ 33.03	01.90 $\pm$ 03.37
	Haindl & Krupička [12]	63.12 $\pm$ 33.30	01.32 $\pm$ 02.10
	<b>FCN Proposed</b>	<b>83.01 <math>\pm</math> 19.47</b>	<b>00.37 <math>\pm</math> 00.43</b>
<b>GAN Proposed</b>	<b>87.42 <math>\pm</math> 13.08</b>	<b>03.27 <math>\pm</math> 03.13</b>	

Remark that both IRISSEG and OSIRISv4.1 frameworks presented good results in NIR datasets, always reaching F1 values over 90%. Nonetheless, our proposed approaches presented statistically better F1 values for all datasets even in the NIR datasets, which are the IRISSEG and OSIRISv4.1 specific image domain. Observe that there are no results for the approach by Haindl & Krupička [12] since it was not developed for NIR images.

Looking at VIS datasets, the results obtained were slightly worse than in the NIR datasets. This is because VIS images usually have more noise, e.g., reflections. The best F1 and

TABLE V  
SUITABILITY (BOLD LINES) FOR NIR AND VIS ENVIRONMENTS.

Dataset	Method	F1 %	E %
BioSec	FCN	97.24 ± 00.81	00.58 ± 00.30
	GAN	90.19 ± 05.52	02.22 ± 01.39
CasiaI3	FCN	97.43 ± 00.74	00.55 ± 00.29
	GAN	97.10 ± 01.83	00.75 ± 01.10
CasiaT4	FCN	95.87 ± 02.66	01.25 ± 00.67
	GAN	82.65 ± 13.98	05.52 ± 04.15
IITD-1	FCN	96.47 ± 01.56	00.72 ± 00.59
	GAN	96.18 ± 02.52	01.09 ± 01.80
NIR	<b>FCN</b>	<b>96.69 ± 01.43</b>	<b>00.78 ± 00.63</b>
	<b>GAN</b>	<b>94.04 ± 07.93</b>	<b>01.72 ± 02.69</b>
NICE.I	FCN	90.68 ± 14.01	02.67 ± 02.04
	GAN	91.40 ± 05.18	01.22 ± 00.71
CrEye-Iris	FCN	96.71 ± 01.11	01.12 ± 00.80
	GAN	93.21 ± 02.30	01.88 ± 00.53
MICHE-I	FCN	88.36 ± 11.88	01.90 ± 02.20
	GAN	89.49 ± 06.76	03.11 ± 02.24
VIS	<b>FCN</b>	<b>89.56 ± 12.36</b>	<b>02.40 ± 02.21</b>
	<b>GAN</b>	<b>92.58 ± 04.89</b>	<b>02.80 ± 02.05</b>

$E$  values achieved for the VIS datasets were achieved by the FCN approach with 97.04%(±01.21) and 00.37%(±00.43), respectively, in the CrEye-Iris and MICHE-I datasets.

It is worth noting that the FCN approach is the one with the smallest  $E$  values in almost all scenarios. This result can be explained by the fact that the FCN approach took advantage of transfer learning, while the GAN approach was trained from scratch.

### C. Suitability and Robustness

Here, experiments for evaluating the suitability and robustness of the proposed approaches are presented. By suitability, we expect that models trained with a specific kind of images, i.e. NIR or VIS images, work as well as when training on a specific dataset. By robustness, we expect that models trained with all kind of images (NIR and VIS) perform as well as when training on a specific dataset.

In summary, the suitability is evaluated by training the models using only NIR or VIS images (i.e., FCN and GAN trained on the NIR merged and VIS also merged datasets). The robustness is evaluated by training the models using all images available (NIR and VIS merged). The results are presented in Tables V and VI, respectively. Note that we report the results of the separate test subsets as well, to facilitate visual comparison between the tables.

By comparing the values presented in Table V with those reported in Table IV, we can observe that the values vary slightly, and thus we can state that the proposed approaches are stable in the suitability scenario.

When comparing the results presented in Table V and Table VI, we noticed that the obtained values of  $F1$  and

TABLE VI  
ROBUSTNESS (BOLD LINES) OF THE IRIS SEGMENTATION APPROACHES.

Dataset	Method	F1 %	E %
BioSec	FCN	96.57 ± 01.14	00.70 ± 00.24
	GAN	85.48 ± 07.63	03.45 ± 01.97
CasiaI3	FCN	97.69 ± 00.82	00.50 ± 00.33
	GAN	93.33 ± 01.98	00.87 ± 00.92
CasiaT4	FCN	95.39 ± 03.20	01.46 ± 01.12
	GAN	85.68 ± 12.92	03.98 ± 02.80
IITD-1	FCN	97.11 ± 01.70	00.61 ± 00.67
	GAN	94.99 ± 03.88	01.28 ± 01.73
NIR	FCN	96.89 ± 06.60	00.82 ± 00.59
	GAN	89.87 ± 07.93	02.39 ± 01.78
NICE.I	FCN	89.25 ± 14.06	03.31 ± 02.77
	GAN	65.56 ± 23.32	11.53 ± 05.87
CrEye-Iris	FCN	96.15 ± 01.90	01.38 ± 01.16
	GAN	88.96 ± 08.98	04.57 ± 04.63
MICHE-I	FCN	80.49 ± 20.65	02.73 ± 02.76
	GAN	61.93 ± 24.97	10.95 ± 06.22
VIS	FCN	88.63 ± 09.15	02.47 ± 02.23
	GAN	72.15 ± 19.03	09.01 ± 05.54
All	<b>FCN</b>	<b>94.36 ± 09.90</b>	<b>01.26 ± 01.73</b>
	<b>GAN</b>	<b>86.62 ± 17.71</b>	<b>04.03 ± 05.28</b>

$E$  were similar in NIR datasets. On the other hand, the performance was considerably lower in VIS datasets. Therefore, the proposed approaches are robust for both NIR and VIS images. However, the GAN approach presented a decrease in the results, while the FCN obtained little variation.

### D. Visual & Qualitative Analysis

Here we perform a visual and qualitative analysis. First, in Fig. 4, we show poor and well-performed iris segmentation results obtained in each dataset by the FCN and GAN approaches. Some images were poorly segmented, thus explaining the high standard deviations obtained.

Then, in Fig. 5, we show iris segmentation performed by both the FCN and GAN approaches, as well as the baselines. We only show one image from each the CasiaI3 and CrEye-Iris datasets due to lack of space.

We particularly chose images where all methods perform fairly well and also where our methods performed better, which is the case in most situations. One can observe that our approach performed better in both NIR and VIS images.

## VI. CONCLUSION

This work presented two approaches (FCN and GAN) for robust iris segmentation in NIR and VIS images in both cooperative and non-cooperative environments. The proposed approaches were compared with three baselines methods and reported better results in all test cases. The transfer learning for each domain (or dataset) was essential to achieve outstanding results since the number of images for training the FCN is relatively small. Therefore, the use of pre-trained models



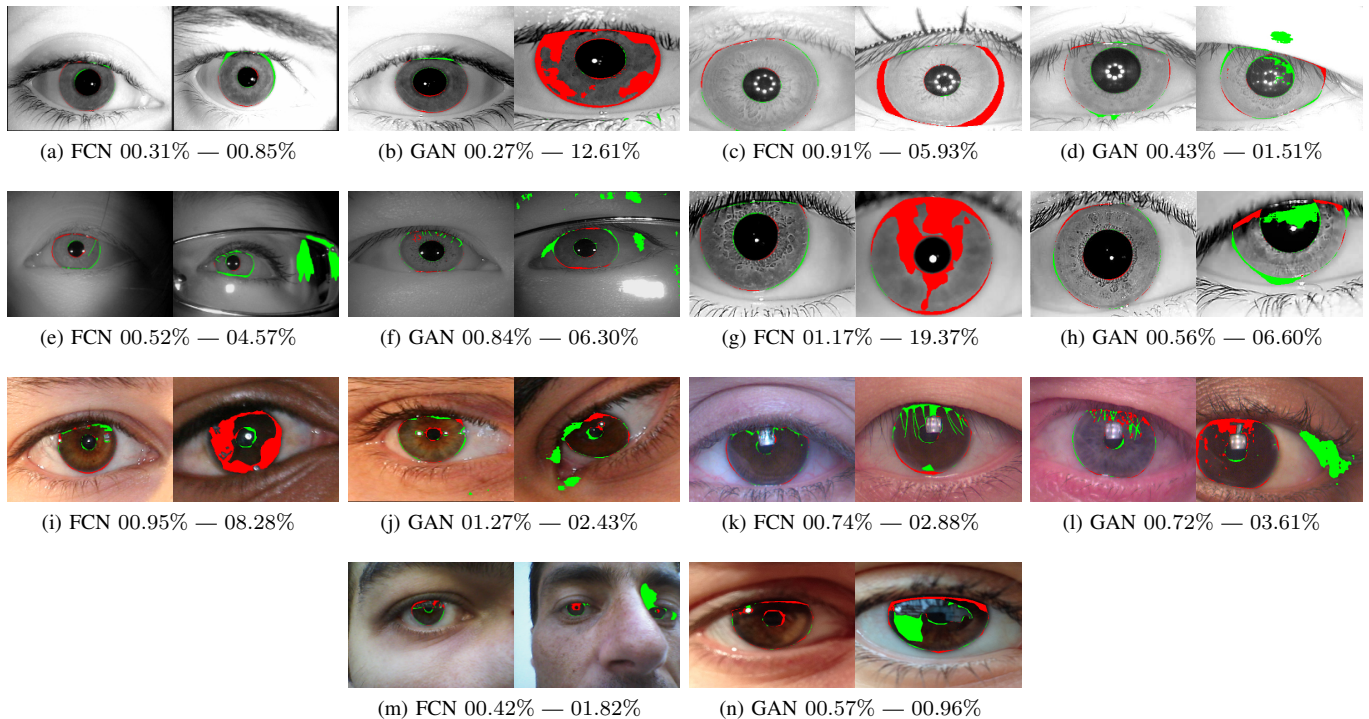


Fig. 4. FCN and GAN qualitative results: good (left) and bad (right) results based on the error  $E$ . Green and red pixels represent the False Positives (FPs) and False Negatives (FNs), respectively. (a)-(b) BioSec; (c)-(d) Casia3; (e)-(f) CasiaT4; (g)-(h) IITD-1; (i)-(j) NICE.I; (k)-(l) CrEye-Iris; (m)-(n) MICHE-I.

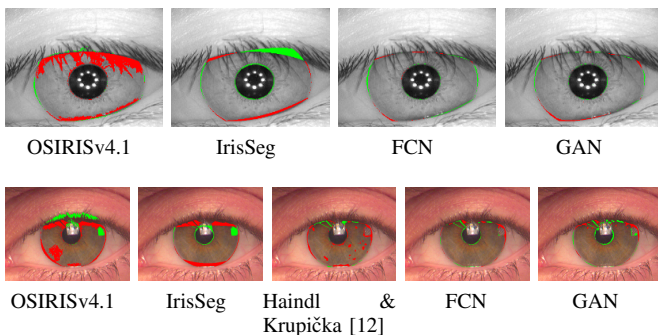


Fig. 5. Qualitative results achieved by the FCN, GAN and baselines. Green and red pixels represent the FP and FN, respectively. The first and second rows correspond, respectively, to images from the Casia3 and CrEye-Iris datasets.

from other datasets brings excellent benefits in learning deep networks. Moreover, specific data augmentation techniques can be applied for improving the performance of the GAN approach.

We also labeled more than 2,000 images for iris segmentation. These masks (manually labeled) are publicly available to the research community, assisting the development and evaluation of new iris segmentation approaches.

Despite the outstanding results, our approach presented high standard deviation rates in some datasets. Therefore, as future work we intend to (i) evaluate the impact of performing the segmentation in two steps, that is, first perform iris detection

and then segment the iris in the detected patch; (ii) create a post-processing stage to refine the prediction, since many images have minor errors (especially at the limbus); (iii) first classify the sensor or image type and then segment each image with a specific and tailored convolutional network model, in order to design a general approach.

#### ACKNOWLEDGMENTS

This work was supported by grants from the National Council for Scientific and Technological Development (CNPq) (# 428333/2016-8, # 313423/2017-2 and # 307277/2014-3) and the Coordination for the Improvement of Higher Education Personnel (CAPES). The Titan Xp GPU used for this research was donated by the NVIDIA Corporation.

#### REFERENCES

- [1] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognition Letters*, vol. 79, pp. 80–105, 2016.
- [2] R. P. Wildes, "Iris recognition: an emerging biometric technology," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348–1363, 1997.
- [3] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, Jan 2004.
- [4] R. Jillela and A. A. Ross, "Segmenting iris images in the visible spectrum with applications in mobile biometrics," *Pattern Recognition Letters*, vol. 57, pp. 4–16, 2015.
- [5] A. Rattani and R. Derakhshani, "Ocular biometrics in the visible spectrum: A survey," *Image and Vision Computing*, vol. 59, pp. 1–16, 2017.
- [6] X. Liu, K. W. Bowyer, and P. J. Flynn, "Experiments with an improved iris segmentation algorithm," in *IEEE AutoID'05*, 2005, pp. 118–123.

- [7] H. Proença and L. A. Alexandre, "Iris segmentation methodology for non-cooperative recognition," *IEEE Proceedings - Vision, Image and Signal Processing*, vol. 153, no. 2, pp. 199–205, 2006.
- [8] E. Ouabida, A. Essadique, and A. Bouzid, "Vander lugt correlator based active contours for iris segmentation and tracking," *Expert Systems with Applications*, vol. 71, pp. 383–395, 2017.
- [9] S. Shah and A. Ross, "Iris segmentation using geodesic active contours," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 824–836, Dec 2009.
- [10] T. Tan, Z. He, and Z. Sun, "Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition," *Image and Vision Computing*, vol. 28, no. 2, pp. 223–230, 2010.
- [11] P. Podder, T. Z. Khan, M. H. Khan, M. M. Rahman, R. Ahmed, and M. S. Rahman, "An efficient iris segmentation model based on eyelids and eyelashes detection in iris recognition system," in *Int. Conf. on Computer Communication and Informatics*, 2015, pp. 1–7.
- [12] M. Haindl and M. Krupička, "Unsupervised detection of non-iris occlusions," *Pattern Recognition Letters*, vol. 57, pp. 60–65, 2015.
- [13] E. Jalilian, A. Uhl, and R. Kwitt, "Domain adaptation for cnn based iris segmentation," in *Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*, Sept 2017, pp. 1–6.
- [14] 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 *Int. Conf. on Biometrics*, 2016, pp. 1–8.
- [15] M. Teichmann, M. Weber, J. M. Zöllner, R. Cipolla, and R. Urtasun, "Multinet: Real-time joint semantic reasoning for autonomous driving," *CoRR*, vol. abs/1612.07695, 2016.
- [16] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *CoRR*, vol. abs/1611.07004, 2016. [Online]. Available: <http://arxiv.org/abs/1611.07004>
- [17] C. Henry, S. M. Azimi, and N. Merkle, "Road segmentation in SAR satellite images with deep fully-convolutional neural networks," *CoRR*, vol. abs/1802.01445, 2018.
- [18] 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*, vol. 66, pp. 90–99, jun 2018.
- [19] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," *CoRR*, vol. abs/1611.08408, 2016.
- [20] F. Jan, "Segmentation and localization schemes for non-ideal iris biometric systems," *Signal Processing*, vol. 133, pp. 192–212, 2017.
- [21] M. D. Marsico, M. Nappi, F. Narducci, and H. Proença, "Insights into the results of MICHE I - Mobile Iris CHallenge Evaluation," *Pattern Recognition*, vol. 74, pp. 286–304, 2018.
- [22] R. Jillela and A. A. Ross, *Methods for Iris Segmentation*. Springer London, 2016, pp. 137–184.
- [23] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, 1993.
- [24] K. Ahuja, R. Islam, F. A. Barbhuiya, and K. Dey, "Convolutional neural networks for ocular smartphone-based biometrics," *Pattern Recognition Letters*, vol. 91, pp. 17–26, 2017.
- [25] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *arXiv preprint arXiv:1603.07285*, 2016.
- [26] 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*, vol. abs/1803.01250, 2018. [Online]. Available: <http://arxiv.org/abs/1803.01250>
- [27] M. Teichmann, M. Weber, M. Zoellner *et al.*, "Multinet: Real-time joint semantic reasoning for autonomous driving," *arXiv preprint arXiv:1612.07695*, 2016.
- [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)*, 2016, pp. 779–788.
- [29] 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*, vol. abs/1802.09567, 2018. [Online]. Available: <http://arxiv.org/abs/1802.09567>
- [30] D. R. Lucio, R. Laroca, E. Severo, A. S. Britto Jr., and D. Menotti, "Fully convolutional networks and generative adversarial networks applied to sclera segmentation," *CoRR*, vol. abs/1806.08722, 2018. [Online]. Available: <http://arxiv.org/abs/1806.08722>
- [31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.
- [32] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, April 2015.
- [33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.
- [34] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*. Curran Associates, Inc., 2014, pp. 2672–2680.
- [35] H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, and A. Uhl, "A ground truth for iris segmentation," in *2014 22nd Int. Conf. on Pattern Recognition*, Aug 2014, pp. 527–532.
- [36] J. Fierrez, J. Ortega-Garcia, D. T. Toledano, and J. Gonzalez-Rodriguez, "Biocase baseline corpus: A multimodal biometric database," *Pattern Recognition*, vol. 40, no. 4, pp. 1389–1392, 2007.
- [37] T. Tan and Z. Sun, "CASIA-IrisV3," *Chinese Academy of Sciences Institute of Automation*, <http://www.cbsr.ia.ac.cn/IrisDatabase.htm>, Tech. Rep, 2005.
- [38] —, "CASIA-IrisV4," *Chinese Academy of Sciences Institute of Automation*, <http://biometrics.idealtest.org/dbDetailForUser.do?id=4.htm>, Tech. Rep, 2005.
- [39] A. Kumar and A. Passi, "Comparison and combination of iris matchers for reliable personal authentication," *Pattern Recognition*, vol. 43, no. 3, pp. 1016–1026, 2010.
- [40] H. Proença and L. A. Alexandre, "Toward covert iris biometric recognition: Experimental results from the NICE contests," *IEEE Trans. on Information Forensics and Security*, vol. 7, no. 2, pp. 798–808, 2012.
- [41] A. Sequeira, L. Chen, P. Wild, J. Ferryman, F. Alonso-Fernandez, K. B. Raja, R. Raghavendra, C. Busch, and J. Bigun, "Cross-Eyed - cross-spectral iris/periorcular recognition database and competition," in *Int. Conf. of the Biometrics Special Interest Group*, Sept 2016, pp. 1–5.
- [42] M. Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols," *Pattern Recognition Letters*, vol. 57, pp. 17–23, 2015.
- [43] Y. Hu, K. Sirlantzis, and G. Howells, "Improving colour iris segmentation using a model selection technique," *Pattern Recognition Letters*, vol. 57, pp. 24–32, 2015.
- [44] H. Proença, S. Filipe, R. Santos, J. Oliveira, and L. Alexandre, "The UBIRIS.v2: A database of visible wavelength images captured on-the-move and at-a-distance," *IEEE TPAMI*, vol. 32, no. 8, pp. 1529–1535, 2010.
- [45] N. Othman, B. Dorizzi, and S. Garcia-Salicetti, "OSIRIS: An open source iris recognition software," *Pat. Rec. Letters*, vol. 82, pp. 124–131, 2016.
- [46] A. Gangwar, A. Joshi, A. Singh, F. Alonso-Fernandez, and J. Bigun, "IrisSeg: A fast and robust iris segmentation framework for non-ideal iris images," in *Int. Conf. on Biometrics (ICB)*, June 2016, pp. 1–8.
- [47] M. Marsico, M. Nappi, and H. Proença, "Results from MICHE II - Mobile Iris CHallenge Evaluation II," *Pattern Recognition Letters*, vol. 91, pp. 3–10, 2017.