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


*Federal University of Paraná (UFPR), Curitiba, PR, Brazil
†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-I) 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† (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

†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 correlation 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 reflection. Ouabida et al. [8] applied the Optical Correlation based Active Contours (OCAC), that uses the Vander Lught 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-YOLO² [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>
<thead>
<tr>
<th>Layer</th>
<th>Filters</th>
<th>Size</th>
<th>Input</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>conv</td>
<td>16</td>
<td>3 × 3/1</td>
<td>416 × 416 × 1/3</td>
</tr>
<tr>
<td>1</td>
<td>max</td>
<td>416</td>
<td>2 × 2/2</td>
<td>416 × 416 × 16</td>
</tr>
<tr>
<td>2</td>
<td>conv</td>
<td>32</td>
<td>3 × 3/1</td>
<td>208 × 208 × 16</td>
</tr>
<tr>
<td>3</td>
<td>max</td>
<td>2 × 2/2</td>
<td>208 × 208 × 32</td>
<td>104 × 104 × 32</td>
</tr>
<tr>
<td>4</td>
<td>conv</td>
<td>64</td>
<td>3 × 3/1</td>
<td>104 × 104 × 32</td>
</tr>
<tr>
<td>5</td>
<td>max</td>
<td>2 × 2/2</td>
<td>104 × 104 × 64</td>
<td>52 × 52 × 64</td>
</tr>
<tr>
<td>6</td>
<td>conv</td>
<td>128</td>
<td>3 × 3/1</td>
<td>52 × 52 × 64</td>
</tr>
<tr>
<td>7</td>
<td>max</td>
<td>2 × 2/2</td>
<td>52 × 52 × 128</td>
<td>26 × 26 × 128</td>
</tr>
<tr>
<td>8</td>
<td>conv</td>
<td>256</td>
<td>3 × 3/1</td>
<td>26 × 26 × 128</td>
</tr>
<tr>
<td>9</td>
<td>max</td>
<td>2 × 2/2</td>
<td>26 × 26 × 256</td>
<td>13 × 13 × 256</td>
</tr>
<tr>
<td>10</td>
<td>conv</td>
<td>512</td>
<td>3 × 3/1</td>
<td>13 × 13 × 256</td>
</tr>
<tr>
<td>11</td>
<td>max</td>
<td>2 × 2/1</td>
<td>13 × 13 × 512</td>
<td>13 × 13 × 512</td>
</tr>
<tr>
<td>12</td>
<td>conv</td>
<td>1024</td>
<td>3 × 3/1</td>
<td>13 × 13 × 512</td>
</tr>
<tr>
<td>13</td>
<td>conv</td>
<td>1024</td>
<td>3 × 3/1</td>
<td>13 × 13 × 1024</td>
</tr>
<tr>
<td>14</td>
<td>conv</td>
<td>30</td>
<td>1 × 1/1</td>
<td>13 × 13 × 1024</td>
</tr>
<tr>
<td>15</td>
<td>detection</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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.](image1)

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.](image2)

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

**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 we 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

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

(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 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 (IRISSEG) 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 & Krupiˇcka [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 trained 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 ±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>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>F1 %</th>
<th>E %</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioSec (NIR)</td>
<td>OSIRISv4.1 [45]</td>
<td>92.62 ± 0.19</td>
<td>01.21 ± 0.47</td>
</tr>
<tr>
<td>FCN Proposed</td>
<td>97.46 ± 0.74</td>
<td>00.44 ± 0.12</td>
<td></td>
</tr>
<tr>
<td>GAN Proposed</td>
<td>96.82 ± 0.28</td>
<td>00.74 ± 0.40</td>
<td></td>
</tr>
<tr>
<td>Casia3 (NIR)</td>
<td>OSIRISv4.1 [45]</td>
<td>89.49 ± 0.57</td>
<td>05.35 ± 0.24</td>
</tr>
<tr>
<td>FCN Proposed</td>
<td>97.90 ± 0.68</td>
<td>01.15 ± 0.37</td>
<td></td>
</tr>
<tr>
<td>GAN Proposed</td>
<td>96.13 ± 0.35</td>
<td>01.45 ± 0.31</td>
<td></td>
</tr>
<tr>
<td>Casia4 (NIR)</td>
<td>OSIRISv4.1 [45]</td>
<td>87.76 ± 0.08</td>
<td>01.34 ± 0.64</td>
</tr>
<tr>
<td>FCN Proposed</td>
<td>94.42 ± 0.74</td>
<td>00.61 ± 0.08</td>
<td></td>
</tr>
<tr>
<td>GAN Proposed</td>
<td>95.38 ± 0.72</td>
<td>01.40 ± 0.93</td>
<td></td>
</tr>
<tr>
<td>IIITD-1 (NIR)</td>
<td>OSIRISv4.1 [45]</td>
<td>92.20 ± 0.07</td>
<td>04.37 ± 0.29</td>
</tr>
<tr>
<td>FCN Proposed</td>
<td>97.44 ± 0.17</td>
<td>01.48 ± 0.01</td>
<td></td>
</tr>
<tr>
<td>GAN Proposed</td>
<td>95.84 ± 0.14</td>
<td>01.33 ± 0.26</td>
<td></td>
</tr>
<tr>
<td>NICE.I (VIS)</td>
<td>IRISSEG [46]</td>
<td>28.64 ± 35.14</td>
<td>13.48 ± 12.36</td>
</tr>
<tr>
<td>FCN Proposed</td>
<td>89.54 ± 13.79</td>
<td>01.00 ± 0.70</td>
<td></td>
</tr>
<tr>
<td>GAN Proposed</td>
<td>91.12 ± 05.08</td>
<td>03.34 ± 02.31</td>
<td></td>
</tr>
<tr>
<td>CrEye-Iris (VIS)</td>
<td>OSIRISv4.1 [45]</td>
<td>46.53 ± 29.25</td>
<td>13.22 ± 06.33</td>
</tr>
<tr>
<td>Haindl &amp; Krupiˇcka [12]</td>
<td>70.59 ± 26.11</td>
<td>04.72 ± 05.87</td>
<td></td>
</tr>
<tr>
<td>FCN Proposed</td>
<td>97.04 ± 01.21</td>
<td>00.96 ± 00.36</td>
<td></td>
</tr>
<tr>
<td>GAN Proposed</td>
<td>92.61 ± 05.86</td>
<td>03.02 ± 03.22</td>
<td></td>
</tr>
<tr>
<td>MICHE-I (VIS)</td>
<td>OSIRISv4.1 [45]</td>
<td>33.85 ± 35.86</td>
<td>01.99 ± 02.90</td>
</tr>
<tr>
<td>Haindl &amp; Krupiˇcka [12]</td>
<td>63.12 ± 33.30</td>
<td>01.32 ± 02.10</td>
<td></td>
</tr>
<tr>
<td>FCN Proposed</td>
<td>83.01 ± 19.47</td>
<td>00.37 ± 00.43</td>
<td></td>
</tr>
<tr>
<td>GAN Proposed</td>
<td>87.42 ± 13.08</td>
<td>03.27 ± 03.13</td>
<td></td>
</tr>
</tbody>
</table>

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 & Krupiˇcka [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
smallest, respectively, in the CrEye-Iris and MICHE-I datasets.

When comparing the results presented in Table V and Table VI, 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 $E$ 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.

**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 kinds 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 $E$ 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) CasiaI3; (e)-(f) CasiaT4; (g)-(h) IITD-1; (i)-(j) NICE.I; (k)-(l) CrEye-Iris; (m)-(n) MICHE-I.

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


