Iris + Ocular: Generalized Iris Presentation Attack Detection Using Multiple Convolutional Neural Networks

Steven Hoffman, Renu Sharma, Arun Ross
Michigan State University, USA
steven.mw.hoffman@gmail.com, sharma90@msu.edu, rossarun@msu.edu

Abstract

An iris recognition system is vulnerable to presentation attacks, or PAs, where an adversary presents artifacts such as printed eyes, plastic eyes or cosmetic contact lenses to defeat the system. Existing PA detection schemes do not have good generalization capability and often fail in cross-dataset scenarios, where training and testing are performed on vastly different datasets. In this work, we address this problem by fusing the outputs of three Convolutional Neural Network (CNN) based PA detectors, each of which examines different portions of the input image. The first CNN (I-CNN) focuses on the iris region only, the second CNN (F-CNN) uses the entire ocular region and the third CNN (S-CNN) uses a subset of patches sampled from the ocular region. Experiments conducted on two publicly available datasets (LivDetW15 and BERC-IF) and on a proprietary dataset (IrisID) confirm that the use of a bag of CNNs is effective in improving the generalizability of PA detectors.

1. Introduction

An iris recognition system uses the rich texture of the iris as a biometric cue for recognizing individuals [6]. While iris systems have demonstrated impressive recognition performance in practical applications, they are vulnerable to a number of adversarial attacks including presentation attacks. A presentation attack (PA) occurs when an adversarial user presents a fake or altered biometric sample to the sensor in order to spoof another user’s identity or to obfuscate their own identity or to create a virtual identity.

A number of iris PAs have been noted in the literature including the presentation of a printed iris image, a plastic/glass/doll eye, a cosmetic contact lens, a video of another individual’s eye, a cadaver eye, a robotic eye, a holographic eye image, or even a real iris of an individual held against their will. Examples of iris PAs are provided in Figure 1. Addressing the problem of PAs has led to development of sophisticated presentation attack detection (PAD) schemes. The aim of the PA detector is to detect all these PAs besides new or unknown PAs that could be developed in the future.

The methods used to detect PAs can be either hardware-based or software-based. Hardware-based methods utilize additional sensor or devices, besides the iris sensor itself, whereas software-based methods use images captured by existing iris sensors. Examples of hardware-based PAD solutions include using an external light stimulus to analyze pupillary responses [5], employing structured light to distinguish between the 3D shape of bonafide irides and cosmetic contacts [4], analyzing the reflectance properties of the iris and sclera by using sophisticated multi-spectral illumination [11] and using eye tracking devices to observe eye movement [17].

In software-based methods, researchers have investigated the use of handcrafted features – such as LBP, LPQ, and BSIF – in conjunction with a Support Vector Machine or other classifiers [18, 15, 7]. Galbally et al. [8] utilized iris quality metrics for detecting the iris liveness, while Pacut et al. [13] analyzed image frequency. In recent years, Convolutional Neural Networks (CNNs) have achieved state-of-the-art performance on many computer vision tasks, including biometrics [14]. As a result, Menotti et al. [12], He et al. [9], and Ramachandra et al. [16] proposed iris PAD solutions based on CNNs where patches from normalized iris images are fed into CNN. Chen and Ross [2, 3] and Hoffman et al. [10] have also developed CNN-based PAD models that utilize the pre-normalized iris rather than the unwrapped iris. While most of these methods resulted in very high PA detection rates, they were primarily evaluated in intra-dataset scenarios where training and testing were based on the same types of PAs. Evaluation using sensors and PAs that were not present in the training set showed a dramatic decrease in detection accuracy [18, 7, 16]. Furthermore, these solutions examined either the entire ocular image provided, with no distinction between the iris and ocular region, or they examined the iris region alone.
Figure 1: Examples of artifacts used to launch iris presentation attacks (PAs): (a) printed images, (b) plastic eyes, and (c) cosmetic contacts (http://hoovervisioncenter.com/2015/10/21/halloween-hazard-the-dangers-of-cosmetic-contact-lenses/).

Because the iris and the surrounding ocular region may include distinct cues for discriminating bonafide samples from PAs, not making use of this information may hinder these software-based solutions. The relative importance of the iris and the ocular region is also studied in [1], though for a different application (sex prediction). In this paper, we develop a software-based iris PAD method that performs well in both intra-dataset and cross-dataset scenarios. This paper also investigates the benefits of examining both the iris region and the surrounding ocular region distinctly when discriminating between bonafide and PA iris samples.

Description of three CNN based solutions discussed in Section 2. Datasets used for the evaluation are described in Section 3. Experimental setup and results are provided in Section 4. Finally, concluding remarks and future work are given in Section 5.

2. Proposed Method

2.1. Iris Region vs Ocular Region

Some authors of PAD solutions have utilized only the iris region of an image, as in [9] and [16]. This is likely due to the fact that the iris region is the only portion of the image guaranteed to contain PA artifacts. For example, when wearing a cosmetic contact, it is likely that the only artifacts that can be used to distinguish between an acceptable clear contact and the adversarial cosmetic contact can be found within the iris region.

Nonetheless, in many cases, the region surrounding the iris, which we call the ocular region, may provide useful cues when a PA is present. Figure 2 shows several such examples. Often times, fingers may be seen in the ocular region holding up a PA. For both print and replay PAs, a white border is sometimes seen surrounding the PAs. Also, for plastic PAs, an unnaturally clear boundary between the iris and sclera is often seen, unobscured by eyelids or eye lashes. It is possible that these features in the ocular region could be used by a CNN to help it increase its PAD accuracy. Some PAD solutions in the literature have analyzed the full iris image, including the ocular region [12]. However, these solutions have not accounted for the inherent differences in the features which can be extracted from the iris region versus those that can be extracted from the ocular region.

We propose three different solutions based on the region that is input to the CNN. The first solution, which we call the Iris CNN (I-CNN), looks primarily at the iris region. The second solution, called the Full Image CNN (F-CNN), looks at the full ocular image, whereas the third solution, called Sampled Ocular CNN (S-CNN), looks at a subset of patches sampled from the ocular region. In all three solutions, we account for the differences between the iris region and the ocular region through the use of segmentation masks, as described in Section 2.3. By analyzing and fusing these different methods, we are able to examine the importance of both the iris region and the ocular region for iris PAD in a novel way.

2.2. Data Preprocessing

As mentioned in Section 2.1, the iris region is the only area of an iris image guaranteed to contain discriminatory artifacts for PA detection (in the context of an adversary targeting an iris system). Therefore, for our first solution (I-CNN), we segment out the iris region to reduce confounding ocular information. Since the iris size in an ocular image varies significantly within and across datasets, we resize all cropped iris images to 300×300 pixels. The size of 300×300 was chosen so that the vast majority of images in the datasets used would be upsampled during resizing, as
Iris and Ocular Presentation Attack Detection Using Multiple Convolutional Neural Networks.

For iris-based PAD detection, the authors present three CNN models: I-CNN, F-CNN, and S-CNN. The I-CNN focuses on iris region patches, while the F-CNN processes the entire iris image into overlapping patches. The S-CNN combines the iris region with ocular region patches.

The I-CNN architecture consists of eight convolutional layers, four max pooling layers, and a fully connected layer. It uses a ReLU non-linearity function and outputs a PA score.

The F-CNN approach tessellates the iris image into overlapping patches of size 96x96, allowing the CNN to identify specific regions. This method is particularly effective in detecting PA artifacts along the iris-sclera boundary.

The S-CNN takes a subset of patches from the ocular region, combined with iris region patches, to detect PA artifacts. This approach demonstrates high performance in PAD detection with a novel approach in this work.
only in the iris and ocular regions but also in the pupil region. For instance, printed iris attacks are often presented by cutting out the pupil region in order to confound sensors that look for specular reflection [17]. Similarly, attacks based on cosmetic contacts often do not obscure the pupil.

To accommodate these findings, we added a second channel to our CNN’s input which we refer as a segmentation mask (segmask). It is a 2-dimensional ternary matrix having the same size as an input image patch, where a value of +1, 0 and −1 corresponds to the iris region, ocular region, and pupil region, respectively. Adding segmask as a second input channel, helps the CNN to learn the importance of each region of the iris image automatically without introducing any additional human bias. An example segmask is shown in Figure 6.

Each one of our three solutions makes use of this same CNN architecture. I-CNN takes as input only those patches extracted from the iris region (Figure 3). F-CNN takes as input patches coming from the entire iris image (Figure 4) whereas S-CNN taking as input a pre-defined subset of patches from the ocular region of the image (Figure 5). Other than the input to the CNNs, the three solutions are identical so that we can robustly analyze the significance of each of the three data input approaches.

In summary, the key aspects of our proposed approach are: (a) The CNNs take patches as input rather than the full iris or ocular image to provide a natural form of data augmentation; (b) The input iris patches are taken from the unnormalized iris image rather than the normalized iris to avoid the downsampling that occurs during iris normalization; (c) A single CNN is trained on patches originating from multiple locations within an image, and, as such, the CNN does not attempt to learn location artifacts but focuses on PA artifacts; (d) Domain knowledge is incorporated by accounting for the number of iris, pupil, and ocular pixels in an input patch through the inclusion of segmentation masks in the input and through defining patch-level fusion functions, as seen in Section 2.4; and (e) Three different CNNs trained respectively on the iris region, the iris and ocular region, and sampled patches from the ocular region are available, allowing us to analyze the role of each of these regions in PA detection.

2.4. Fusion Techniques

As described in Section 2.2, each ocular image presented to our system is tessellated into a number of different patches, and each one of these patches produces its own PA score after passing through a CNN model. A fusion method is needed in order to consolidate the patch PA scores and render a decision. One possible fusion method is to take the average score. But as mentioned in Section 2.3, the iris pixels provide more information for discriminating PAs from bonafide samples compared to pupil pixels. This led to the idea that patch scores should be weighted such that they are rewarded for having more iris pixels and penalized for having more pupil pixels. We designed two score fusion techniques based on this intuition, the iris-only ratio ($s_{io}$) score and the iris-pupil ratio ($s_{ip}$) score:

$$s_{io} = \left( \frac{1}{\sum_{i=1}^{K} a_i} \right) \sum_{i=1}^{K} \left[ 8_i [ i ] -1 \right] a_i \right] \cdot \left[ 0,1 \right]$$

$$s_{ip} = \left( \frac{1}{\sum_{i=1}^{K} \frac{a_i}{1+b_i}} \right) \sum_{i=1}^{K} \left[ 8_i [ i ] -1 \right] \frac{a_i}{1+b_i} \right] \cdot \left[ 0,1 \right]$$

where, $s_i$ is the score of the $i$th iris patch, $a_i$ and $b_i$ are the proportion of iris pixels and the proportion of pupil pixels in the $i$th patch, respectively, $K$ is the total number of patches per image, $\left[ \cdot \right]_{[-1,+1]}$ and $\left[ \cdot \right]_{[0,1]}$ are the functions that convert scores to a given range. During this computation, we first convert the scores to a $[-1,+1]$ range so that the proportion of iris and pupil pixels is not affected by a 0 score value. We divide by $1+b_i$ rather than $b_i$ to avoid a divide-by-zero situation. This produces the iris-only score, which is a weighted average giving higher priority to patches having a larger proportion of iris pixels, and the iris-pupil score, which gives higher priority to a patch having more iris pixels and less pupil pixels. These fusion techniques do not directly take into account the spatial location of the patch, only considering the proportion of iris or pupil contained in a patch.
3. Datasets

To evaluate the proposed solutions, we collected our own dataset called the IrisID Dataset. The aim of the dataset is to introduce a large amount of variations among PAs. This dataset contains 1343 bonafide images, 1651 print PAs, 352 plastic/glass PAs, and 130 Kindle replay PAs. All of these images were collected using multiple IrisID 7000 NIR camera across several locations. For the print PAs, six different printers, two different paper types (glossy and matte), two different print sizes, and two different types of prints (ones with pupil cut out and ones without any cutouts) were used to increase the variations in the prints. In addition, five different bonafide iris datasets were used to source the images to print for the dataset. For the plastic PAs, three different plastic/glass eye brands were used with 10 distinct colors. For the replay PAs, we used the Kindle 8 device as modern computer or mobile screens do not reflect NIR light and we observed that the E-Ink display of this version of Kindle does reflect NIR illumination. The large degree of variability included in this dataset should help algorithms trained on it to better generalize to unseen datasets or PAs. Example images from this dataset can be seen in Figure 2. In addition to this, we evaluated our methods on two publicly available datasets: the LivDet-Iris 2015 Warsaw dataset [19] and the BERC-Iris-Fake dataset [11]. We used the Verieye segmentation software to locate the irides and pupils in the images whenever such information was not provided with the datasets. Details about the training and testing partitions on the three datasets are summarized in Table 1.

4. Results

We trained CNNs on each dataset separately for our three solutions, experimenting with different batch sizes and learning rates to optimize performance. When a dataset contained multiple PAs (i.e., BERC-IF), we trained a single CNN on all PA types simultaneously to increase the model’s generalizability across PA types. During testing, however, we evaluated the performance of each PA type separately to obtain a more insightful picture of how our CNN models perform against each type of PA. We reported TDR (True Detection Rate - percentage of PA samples that were correctly detected) values at a FDR (False Detection Rate - percentage of bonafide samples that were misclassified as PA) value of 1%. The iris-pupil ratio fusion technique (Section 2.4) is used for each CNN trained on the BERC-IF dataset and for the I-CNN trained on LivDetW15, as it performed the best for them; the average fusion technique performed the best for all the remaining CNNs. When a CNN is tested on the same dataset that it was trained on, we call it an intra-dataset evaluation; when the same CNN is evaluated on a dataset that is different from what it was trained on, it is called cross-dataset evaluation.

4.1. Intra-Dataset Results

Our intra-dataset results for the I-CNN, F-CNN, and S-CNN are shown in Table 2. The current state-of-the-art techniques [19] [11] give 100% accuracy on BERC-H and LivDetW15 datasets in intra-dataset scenario. All of our solutions did very well when tested against our IrisID dataset. We achieved perfect detection on both the plastic and Kindle PAs while achieving 98.90% TDR on printed PAs. We believe that the slightly lower TDR on prints is due to the large degree of variability introduced in IrisID print set. Upon further analysis, it was found that all of the print images on which the IrisID F-CNN failed were of very high quality coming from the same printer and paper type. Adding more prints from this printer/paper combination into our training set could help us achieve perfect PAD results in this category. Nonetheless, all of our CNNs are able to achieve comparable results with the state-of-the-art, with two of our CNNs matching Federico et al. [19] results, thereby confirming the efficacy of our solutions. Furthermore, BERC-H method [11] used a hardware-based method to perfectly classify the BERC-IF dataset, relying on a sensor to capture images at both 750nm and 850nm. We were able to obtain similar accuracy using all three of our solutions without taking into account the differences in spectra, using a software-based solution.

4.2. Cross-Dataset Results

In the literature, very little has been reported on the generalizability of iris PAD algorithms across PAs, sensors, or datasets not seen in the training data. Some authors [7] [18] [16] evaluated their models on a cross-sensor scenario and it is found that the TDR was often much lower than those seen in intra-dataset conditions, while the FDR was much higher. This is likely due to the large variations in image features obtained across sensors, and it highlights the difficulties in cross-sensor testing scenarios. A much more difficult scenario occurs when testing is performed across different datasets (cross-dataset). During cross-sensor testing, one must only account for variations in cameras, but during cross-dataset testing, one must account for variations in the sensors, data acquisition environment, subject population, and PA generation procedures. This makes cross-dataset testing a very difficult problem [10]. Nonetheless, for real-world applications, an iris PAD algorithm should not only work in a cross-dataset scenario, but should also be able to operate in a cross-attack scenario, for it is impossible to know in advance exactly what attacks an adversary may deploy.

We evaluated each of our three solutions under both cross-dataset and cross-attack scenarios. The results can be seen in Table 2. It can be noted that in many cases, especially for the I-CNN, there is a decrease in accuracy when transitioning from intra-dataset to cross-dataset tests. How-
Table 1: Summary of the datasets used in the paper along with the number of images in each dataset, the number of images on which automatic segmentation failed, and the number of images remaining after those with failures have been removed.

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th></th>
<th>Testing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bonafide</td>
<td>Print</td>
<td>Plastic</td>
<td>CC</td>
</tr>
<tr>
<td>IrisID</td>
<td>Original</td>
<td>769</td>
<td>1254</td>
<td>282</td>
</tr>
<tr>
<td></td>
<td>Seg Fails</td>
<td>19</td>
<td>78</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Remaining</td>
<td>750</td>
<td>1176</td>
<td>270</td>
</tr>
<tr>
<td>LivDetW15</td>
<td>Original</td>
<td>603</td>
<td>582</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Seg Fails</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Remaining</td>
<td>603</td>
<td>582</td>
<td>-</td>
</tr>
<tr>
<td>BERC-IF</td>
<td>Original</td>
<td>2258</td>
<td>1280</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Seg Fails</td>
<td>609</td>
<td>683</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Remaining</td>
<td>1649</td>
<td>597</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 2: The results of the proposed approaches under intra-dataset, cross-dataset and cross-attack conditions. Numbers in italics indicate intra-dataset results, numbers in bold indicate cross-attack results and rest indicate cross-dataset results. Note that these results are without fusing information from the three CNNs. For fusion results, see Table 3.

<table>
<thead>
<tr>
<th>CNN Type</th>
<th>Training Set</th>
<th>IrisID* (TDR@FDR = 1%)</th>
<th>LivDetW15* (TDR@FDR = 1%)</th>
<th>BERC* (TDR@FDR = 1%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Print</td>
<td>Plastic</td>
<td>Kindle</td>
</tr>
<tr>
<td>I-CNN</td>
<td>IrisID</td>
<td>98.90%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>LivDetW15</td>
<td>52.19%</td>
<td>31.82%</td>
<td>20.83%</td>
</tr>
<tr>
<td></td>
<td>BERC-IF</td>
<td>76.37%</td>
<td>48.48%</td>
<td>75.00%</td>
</tr>
<tr>
<td>F-CNN</td>
<td>IrisID</td>
<td>99.17%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>LivDetW15</td>
<td>84.62%</td>
<td>51.52%</td>
<td>87.50%</td>
</tr>
<tr>
<td></td>
<td>BERC-IF</td>
<td>65.66%</td>
<td>19.70%</td>
<td>50.00%</td>
</tr>
<tr>
<td>S-CNN</td>
<td>IrisID</td>
<td>98.07%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>LivDetW15</td>
<td>85.71%</td>
<td>81.81%</td>
<td>79.17%</td>
</tr>
</tbody>
</table>

ever, especially for our F-CNNs and S-CNNs trained on the IrisID and LivDetW15 datasets, there is some very good cross-dataset and cross-attack results. Our best performing CNNs were the F-CNNs and S-CNNs trained on the IrisID dataset, achieving 100% or high 90% TDR values in all but one scenario. The one experiment for these CNNs that did not result in greater than 90% TDR was when testing the S-CNN on the cross-attack BERC-IF cosmetic contact PA. The CNN likely did worse on this experiment because cosmetic contacts do not provide any PA artifacts outside of the iris region and the S-CNN acquires most of its patches from outside of or at the edge of the iris region. Overall, the F-CNNs and S-CNNs performed much better at cross-dataset experiments than the I-CNNs. Comparatively, the F-CNN seems to perform slightly better than the S-CNN. As a consequence, this CNN takes 0.0131 secs which is four times as long as the S-CNN to process an image (0.032 secs). Thus, it would be best to choose the CNN based on the operational scenario in which it should be used.

To further improve our performance, we tried fusing the PA scores produced by our nine trained CNNs. To fuse the scores, we used the max and average fusion rule. Altogether, we tested over 1000 different CNN fusion configurations. The best five CNN fusion results are shown in Table 3. These fusion schemes performed extremely well, producing nearly perfect results when combining only two or three CNNs. It is important to note that all of the reported fusion schemes make use of our two best performing models: the F-CNN and the S-CNN trained on the IrisID dataset. This shows how useful having a sufficiently diverse training set can be towards producing generalizable results. It also shows the importance in considering both the ocular and iris region as distinct features when performing iris presentation attack detection.

4.3. Feature Map and Heat Map Analysis

To gain a better understanding of what our CNN models learned, we analyzed the intermediate representations (feature maps) of image patches as they passed through the CNN. We looked at our first solution, the I-CNN trained...
Table 3: The results obtained by fusing the scores of several of our CNNs. Here numbers indicate CNN models. 1, 2 and 3 correspond to the F-CNN, S-CNN and I-CNN trained on IrisID dataset, receptively, whereas 4 and 5 correspond to S-CNN and F-CNN trained on LivDetW15. Numbers that are in italics indicate intra-dataset results and numbers that are bold indicate cross-attack results.

<table>
<thead>
<tr>
<th>Fusion Rule</th>
<th>Models Fused</th>
<th>IrisID*</th>
<th>LivDetW15*</th>
<th>BERC-IF*</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAX</td>
<td>1 + 2</td>
<td>99.18%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>AVG</td>
<td>1 + 2 + 3</td>
<td>99.73%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>MAX</td>
<td>1 + 2 + 4</td>
<td>99.18%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>MAX</td>
<td>1 + 2 + 5</td>
<td>99.18%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>MAX</td>
<td>1 + 2 + 4 + 5</td>
<td>99.18%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Figure 7: An example image patch from a BERC cosmetic contact PA sample.

Figure 8: An example image patch from a BERC bonafide sample.

Figure 9: Cross-dataset images misclassified by the I-CNN trained on BERC-IF. Also shown are their corresponding heatmaps, ground truth, and PA scores. The threshold is 0.64. In the heatmap, red color represents a score of 1 (PA) whereas green represents a score of 0 (Bonafide).

on the BERC-IF dataset, and viewed the outputs of the 1st, 3rd, 5th, and 7th convolutional layers for a PA input sample (Figure 7) and for a bonafide input sample (Figure 8). An analysis of these images reveals that the first few layers of the CNN work as edge detectors, emphasizing edges formed by both the iris texture and by the low resolution of the PA. In particular, the outputs of the 1st and 3rd convolutional layers for the PA sample reveal a fine-grained grid pattern formed from the pixelation in the cosmetic contact. These fine-grained pixelation edges get smoothed out later in the CNN such that by the 7th convolutional layer, the output feature maps contain small degrees of intensity variation. For the bonafide sample, the irregular iris pattern is such that a strong grid pattern is not seen in the first few layers. This leads the CNN to have a large amount of variation in the 7th layer, allowing the bonafide samples to be discriminated from PAs. This analysis suggests that our model should do well at identifying PAs when their iris textual details are of lower resolution. Conversely, when a PA sample has a higher resolution and the inherent pixelation is not evident in the resulting image, our model is not as likely to do well at identifying it.

For further analysis, we generated “PA Score Heatmaps” of each image by coloring each pixel of the iris image a color corresponding to that pixel’s average PA score pertaining to the patches containing that pixel. These heatmaps allowed us to quickly understand which parts of an image our CNN was able to correctly classify and which areas it had more difficulty with. A few observations can be made. First, we noticed that a large number of errors in both bonafide and PA images occurred when there was a significant amount of glare. Glare obscures the underlying object’s texture. We also noticed that some of the misclassified bonafide images, such as in Figure 9a, the texture of the iris was rather smooth, not containing a lot of discriminative texture. The same can be seen for some misclassified spoof images, like in Figure 9b, where the PA sample has high contrast, making it difficult to discern any texture. This corresponds with our feature map analysis, suggesting that...
when neither the iris texture nor the PA pixelations can be identified by the CNN’s edge-detecting filters, the CNN is unable to reliably classify the image.

5. Discussion

Upon analyzing the results, we concluded that the proposed CNN solutions were able to perform exceptionally well on intra-dataset, cross-dataset, and cross-attack results. Analyzing both the ocular and iris region, and using segmentation masks which help the CNNs to distinguish between them, proved to be an effective and novel way to produce state-of-the-art cross-dataset and cross-attack results. Furthermore, this work also introduces the IrisID dataset, a new dataset exhibiting a large degree of variations amongst PAs to help algorithms learn generalizable features and that introduces a new PA type: Kindle replay attacks. For the S-CNN, we are currently pre-selecting a set of patches sampled from the ocular region. This process can be automated by detecting different parts of the ocular region where the probability of finding PA artifacts is high. For further generalization, we can also obtain cues from the scene when the user is interacting with the biometric sensor.

Acknowledgment

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2017 - 17020000004. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References