Counteracting Presentation Attacks in Face, Fingerprint and Iris Recognition

Allan Pinto[1], Helio Pedrini[1], Michael Krumdick[2], Benedict Becker[2], Adam Czajka[2,3,4], Kevin W. Bowyer[2], and Anderson Rocha[1]

[1]Institute of Computing, University of Campinas, Brazil
[2]Computer Science and Engineering, University of Notre Dame, U.S.A.
[3]Research and Academic Computer Network (NASK), Poland
[4]Warsaw University of Technology, Poland

April 1, 2017

Abstract

This chapter explores data-driven approaches to presentation attack detection for three biometric modalities: face, iris and fingerprint. The primary aim of this chapter is to show how pre-trained deep neural networks can be used to build classifiers that can distinguish between authentic images of faces, irises and fingerprints and their static imitations. The most important, publicly available benchmarks representing various attack types were used in a unified presentation attack detection framework in both same-dataset and cross-dataset experiments. The pre-trained VGG neural networks, being the core of this solution, tuned independently for each modality and each dataset present almost perfect accuracy for all three biometric techniques. In turn, low classification accuracies achieved in cross-dataset evaluations show that models based on deep neural networks are sensitive not only to features specific to biometric imitations, but also to dataset-specific properties of samples. Thus, such models can provide a rapid solution in scenarios in which properties of imitations can be predicted but appropriate feature engineering is difficult. However, these models will perform worse if the properties of imitations being detected are unknown. This chapter includes also a current literature review summarizing up-to-date data-driven solutions to face, iris and finger liveness detection.

1 Introduction

Biometric authentication is a technology designed to recognize humans automatically based on their behavior, physical and chemical traits. Recently, this technology emerged as an important mechanism for access control in many modern applications, in which the

traditional methods including the ones based on knowledge (*e.g.*, keywords) or based on tokens (*e.g.*, smart cards) might be ineffective since they are easily shared, lost, stolen or manipulated [36]. Biometric technologies are increasingly used as the main authenticating factor for access control and also jointly with traditional authentication mechanisms, as a "step-up authentication" factor in two- or three-factor authentication systems.

In this context, face, iris and fingerprint are the most commonly-used biometric traits. In fact, the choice of the trait to be used takes into account some issues such as universality, easiness to measure the biometric characteristics, performance, or difficulty to circumvent the system [36]. However, a common disadvantage of these traits is that an impostor might produce a synthetic replica that can be presented to the biometric sensor to circumvent the authentication process. In the literature, the mechanisms to protect the biometric system against this type of attack are referred to as *spoofing detection*, *liveness detection* or *presentation attack detection* (PAD). Hereinafter, we will use the most generic term, presentation attack detection (PAD), which was initially proposed by SC37 experts in ISO/IEC 30107 – Presentation Attack Detection – Framework (Part 1), Data Formats (Part 2), and Testing and Reporting (Part 3).

The idea of spoofing biometric recognition is surprisingly older than biometrics itself. A careful reader of the Old Testament can find an impersonation attempt described in the Book of Genesis, based on presentation of a goat's fur put on Jacob's hand to imitate properties of Esau's skin, so that Jacob would be blessed by Isaac. A fictitious example that is surprisingly realistic is the description of how to copy someone's fingerprint using a wax mold and gelatin presented by Austin Freeman in his crime novel "The Red Thumb Mark". The novel appeared in 1907, and the technique described is still used almost 100 years later to spoof fingerprint sensors. Note that this description appeared only four years after fingerprints were adopted by Scotland Yard, and long before the first fingerprint sensor appeared on the market.

Recent scientific studies and open challenges such as LivDet (`www.livdet.org`) suggest that presentation attacks are still an open problem in biometrics. Phan and Boulkenafet [6, 73] suggest that face recognition systems are vulnerable to presentation attacks with an equal error rate (related to distinguishing presentation attacks from genuine samples) reaching as high as 9%. Fingerprint-based recognition systems still face the same problem, with an average classification error rate achieving 2.9% [60]. Iris-based authentication, considered by many to be one of the most reliable biometrics, awaits efficient PAD methodology. Recent proposals in this area still report an average classification error rate around 1% [80].

Besides the laboratory testing of the biometric system's vulnerability to attack, a few real cases also confirm the problem. In the small city of Ferraz de Vasconcelos, in the outskirts of São Paulo, Brazil, a physician of the service of mobile health care and urgency was caught red-handed by the police in a scam that used silicone fingers to bypass an authentication system and confirm the presence several colleagues at work [49]. A similar case has been investigated by the Brazilian Federal Police in 2014, when workers at the

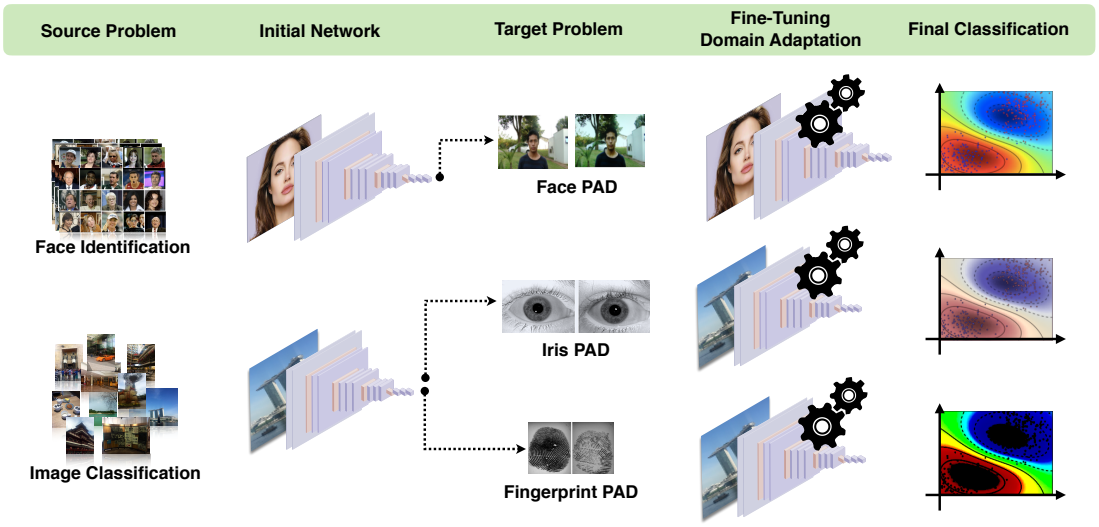

Figure 1: General pipeline exploited in this work. Initial network architectures, originally proposed for other problems, are independently fine-tuned with appropriate PAD examples from different datasets leading to discriminative features. Ultimately, classifiers are trained to separate between authentic images of faces, irises and fingerprints from their presentation attack versions.

Paranaguá Harbor in the Brazilian southern state of Paraná, were suspected of using silicone fingers to circumvent a time attendance biometric system [8]. In Germany, the biometric hacking team in the Chaos Computer Club managed to hack Apple's iPhone Touch ID [3] a few days after its launch, demonstrating that a biometric system without an adequate protection is unsuitable as a reliable access control method. Other cases of spoofing surveillance systems with 3-D masks to change their apparent age or race can also be found in [84, 85].

Considering the three aforementioned modalities, when we look at the literature and analyze the algorithms to prevent presentation attacks, we observe that the most promising in terms of errors and minimum effort of implementation or cost often share an interesting feature: they belong to a group of algorithms referred to as data-driven characterization algorithms. According to Pinto *et al.* [75], methods based on data-driven characterization exploit only the data that comes from a standard biometric sensor looking for evidence of artifacts in the already acquired biometric sample. Such approaches are preferable in practice because they are easily integrable with the existing recognition systems, as there is no extra requirement in terms of hardware nor is there the need of human interaction to detect attempted attacks.

Although the existing methods following this idea have led to good detection rates,

we note that some aspects still need to be taken into account when evaluating a PAD approach, *e.g.*, different types of attack, variety of devices to perform attempted attacks, and attacks directed to different sensors. Another aspect that is normally overlooked is that most detection methods are custom-tailored to specific types of presentation attacks, in what we refer to as hand-crafting of the features. With the emergence of deep learning methods and their success in tasks such as image classification, voice recognition and language translation, in this chapter, we set forth the objective of exploiting deep learning solutions for detecting presentation attacks, using data-driven solutions. In these cases, the biometric designer is responsible for choosing an appropriate architecture for PAD and training solely from the existing data available. We believe that this type of solution is the next step when designing robust presentation attack detectors and also that they can, if carefully designed, better deal with the challenging cross-dataset scenario. The cross-dataset scenario arises when the system is trained with a dataset from one sensor or one scenario, and then later tested on data from a different sensor or scenario. Figure 1 depicts the general pipeline we exploit in this chapter. We start with pre-trained deep neural networks and tune them independently for each modality (face, iris and fingerprint) with different datasets before building the final classifiers to distinguish between authentic images of faces, irises and fingerprints from their static counterparts.

We organize the rest of this Chapter as follows. Section 2 discusses state-of-the-art methods for PAD considering the three modalities considered in this chapter (face, iris and fingerprint). Section 3 details the data-driven PAD solution that we advocate as very promising for this problem, while Section 5 shows the experiments and validations for different biometric spoofing datasets. We close the chapter with some final considerations in Section 6.

2 Related Work

In this section, we review some of the most important presentation-attack detection methods published in the literature for iris, face and fingerprint.

2.1 Face Presentation Attack Detection

The existing face anti-spoofing techniques can be categorized into four groups [90]: user behavior modeling [67, 109] (*e.g.*, eye blinking, small face movements), methods that require additional hardware [18] (*e.g.*, infrared cameras and depth sensors), methods based on user cooperation (*e.g.*, challenge questions) and, finally, data-driven characterization approaches, which is the focus of our work herein.

We start this section reviewing frequency-based approaches, which are methods that rely on analyzing artifacts that are better visible in the frequency domain. Early studies followed this idea [48], and nowadays we have several works that support the effectiveness of this approach in detecting face spoofing. In [48], Li *et al.* proposed a face spoofing

detection that emerged from the observation that the faces in photographs are smaller than the real ones and that the expressions and poses of the faces in photographs are invariant. Based on these observations, the authors devised a threshold-based decision method for detecting photo-based attempted attacks based on the energy rate of the high frequency components in the 2-D Fourier spectrum. The major limitation of the technique proposed by Li *et al.* is that the high frequency components are affected by illumination, which makes this frequency band too noisy [48,102]. To reduce that effect, Tan *et al.* [102] exploited the difference of image variability in the high-middle band. This is done using Difference of Gaussian (DoG) bandpass filtering, which keeps as much detail as possible without introducing noisy or aliasing artifacts.

In [74], Pinto *et al.* introduced an idea seeking to overcome the illumination effect when working in the frequency domain. In that work, the authors proposed a face anti-spoofing method for detecting video-based attempted attacks based on Fourier analysis of the noise signature extracted from videos, instead of using the image pixel values directly. Basically, after isolating the noise signal present in the video frames, the authors transformed that information to the Fourier domain and used the visual rhythm technique to capture the most important frequency components to detect an attempted attack, taking advantage of the spectral and temporal information. In a more recent work [76], the same authors expanded upon this technique taking advantage of the spectral, temporal and spatial information from the noise signature, using the concept of visual codebooks. According to the authors, the new method enabled them to detect different types of attacks such as print- and mask-based attempted attacks as well.

Lee *et al.* [47] proposed an anti-spoofing technique based on the cardiac pulse measurements using video imaging [78]. The authors extended upon previous work proposed by Poh *et al.* [78] by adding a threshold-based decision level based on the entropy measure. It was calculated from the power spectrum obtained from normalized RGB channels after eliminating the cross-channel noise, caused by the environment interference, using the Independent Component Analysis (ICA).

Another expressive branch of face anti-spoofing algorithms reported in the literature consists of texture-based approaches. In general, those algorithms exploit textural cues inserted in the fake biometric samples during its production and presentation to the biometric sensor under attack (*e.g.*, printing defects, aliasing and blurring effects). Tan *et al.* [102] proposed a texture-based approach to detect attacks with printed photographs based on the difference of the surface roughness of an attempted attack and a real face. The authors estimate the luminance and reflectance of the image under analysis and classify them using Sparse Low Rank Bilinear Logistic Regression methods. Their work was extended upon by Peixoto *et al.* [70], who incorporated measures for different illumination conditions.

Similar to Tan *et al.* [102], Kose *et al.* [41] evaluated a solution based on reflectance to detect attacks performed with printed masks. To decompose the images into components of illumination and reflectance, the Variational Retinex [1] algorithm was applied.

Määttä *et al.* [52,53] relied on micro textures for face spoofing detection, inspired by the characterization of printing artifacts and by differences in light reflection when comparing real samples and presentation attack samples. The authors proposed a fusion scheme based on the Local Binary Pattern (LBP) [61], Gabor wavelets [15], and Histogram of Oriented Gradients (HOG) [13]. Similarly, to find a holistic representation of the face able to reveal an attempted attack, Schwartz *et al.* [90] proposed a method that employs different attributes of the images (*e.g.*, color, texture and shape of the face).

Chingovska *et al.* [9] investigated the use of different variations of the LBP operator used in [52]. The histograms generated from these descriptors were classified using χ^2 histogram comparison, Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM).

Face spoofing attacks performed with static masks have also been considered in the literature. Erdogmus *et al.* [17] explored a database with six types of attacks using facial information of four subjects. To detect attempted attacks, the authors used two algorithms based on Gabor wavelet [46] with a Gabor-phase based similarity measure [32].

Pereira *et al.* [71] proposed a score-level fusion strategy for detecting various types of attacks. The authors trained classifiers using different databases and used the Q statistics to evaluate the dependency between classifiers. In a follow-up work, Pereira *et al.* [72] proposed an anti-spoofing solution based on the dynamic texture, which is a spatiotemporal version of the original LBP.

Garcia *et al.* [24] proposed an anti-spoofing method based on detection of the Moiré patterns, which appear due to the overlap of the digital grids. To find these patterns, the authors used a peak-detector algorithm based on maximum-correlation thresholding, in that strong peaks reveal an attempted attack. Similar to [24], Patel *et al.* [69] proposed a presentation attack detection technique also based on the Moiré pattern detection, which uses the the multi-scale version of the LBP descriptor (M-LBP).

Tronci *et al.* [106] exploited the motion information and clues that are extracted from the scene by combining two types of processes, referred to as static and video-based analysis. The static analysis consists of combining different visual features such as color, edge, and Gabor textures, whereas the video-based analysis combines simple motion-related measures such as eye blink, mouth movement, and facial expression change.

Anjos *et al.* [2] proposed a method for detecting photo-based attacks assuming a stationary facial recognition system. According to the authors, the intensity of the relative motion between the face region and the background can be used as a clue to distinguish valid access of attempted attacks, since the motion variations between face and background regions exhibit greater correlation in the case of attempted attacks.

Wen *et al.* [108] proposed a face spoof detection algorithm based on image distortion analysis (IDA), describing different features such as specular reflection, blurriness, chromatic moment, and color diversity. These features are concatenated in order to generate feature vectors, which are used to generate an ensemble classifier, each one specialized to detect a type of attempted attack.

Kim *et al.* [38] proposed a method based on the diffusion speed of a single image to detect attempted attacks. The authors define the local patterns of the diffusion speed, namely local speed patterns via Total Variation (TV) flow [86], which are used as feature vectors to train a linear classifier, using the SVM, to determine whether the given face is fake. In turn, Boulkenafet *et al.* [7] proposed an anti-spoofing technique using a color texture analysis. Basically, the authors perform a micro-texture analysis considering the color-texture information from the luminance and the chrominance channels by extracting feature descriptions from different color spaces.

Different from the previous methods, which focus on defining a presentation attack detection that does not leverage the identity information present in the gallery, Yang *et al.* [111] proposed a person-specific face anti-spoofing approach, in which a classifier was built for each person. According to the authors, this strategy minimizes the interferences among subjects.

Virtually all previous methods exploit handcrafted features to analyze possible clues related to a presentation attack attempt. Whether these features are related to texture, color, gradients, noise or even reflection, blurriness, and chromatic moment, they always come down to the observation of specific artifacts present in the images and how they can be captured properly. In this regard, LBP stands out as the staple of face-based spoofing research thus far. Departing from this hand-crafted characterization modeling strategy, a recent trend in the literature has been devoted to designing and deploying solutions able to directly learn, from the existing available training data, the intrinsic discriminative features of the classes of interest, the so-called data-driven characterization techniques, probably motivated by the huge success these approaches have been showing in other vision-related problems [42, 100]. Out of those, the ones based on deep learning solutions stand out right away as very promising for being highly adaptive to different situations.

Menotti *et al.* [58] aimed at hyperparameter optimization of network architectures [4, 77] (architecture optimization) and on learning filter weights via the well-known back-propagation algorithm [45] (filter optimization) to design a face spoofing detection approach. The first approach consists of learning suitable convolutional network architectures for each domain, whereas the second approach focuses on learning the weights of the network via back propagation.

Manjani *et al.* [54] proposed an anti-spoofing solution based on a deep dictionary learning technique originally proposed in [104] to detect attempted attacks performed using silicone masks. According to the authors, deep dictionary learning combines concepts of two most prominent paradigms for representation learning, deep learning and dictionary learning, which enabled the authors to achieve a good representation even using a small data for training.

2.2 Fingerprint Presentation Attack Detection

Fingerprint PAD methods can be categorized into two groups: hardware-based and software-based solutions [27]. Methods falling into the first group use information provided from additional sensors to gather artifacts that reveal a spoofing attack that is outside of the fingerprint image. Software-based techniques rely solely on the information acquired by the biometric sensor of the fingerprint authentication system.

Based on several quality measures (*e.g.*, ridge strength or directionality, ridge continuity), Galbally *et al.* [20, 21] proposed a set of features aiming at fingerprint presentation attack detection, which were used to feed a Linear Discriminant Analysis (LDA) classifier.

Gragnaniello *et al.* [29] proposed an anti-spoofing solution based on Weber Local Descriptor (WLD) operating jointly with other texture descriptors such as Local Phase Quantization (LPQ) and Local Binary Pattern Descriptor (LBP). The experimental results suggest that WLD and LPQ complement one another, and their joint usage can greatly improve their discriminating ability, even when compared individually or combined with LBP.

Inspired by previous works based on LBP descriptor, Jia *et al.* [37] proposed a spoofing detection scheme based on Multi-scale Block Local Ternary Patterns (MBLTP) [103]. According to the authors, the computation of the LTP descriptor is based on average values of block subregions rather than individual pixels, which makes it less sensitive to noise, since the computation is based on a 3-value code representation and on average values of block subregions, rather than on individual pixels.

Ghiani *et al.* [25] proposed the use of Binarized Statistical Image Features (BSIF), a textural binary descriptor whose design was inspired by the LBP and LPQ methods. Basically, the BSIF descriptor learns a filter set by using statistics of natural images [34], leading to descriptors better adapted to the problem. The same authors also explored the LPQ descriptor to find a feature space insensitive to blurring effects [26].

In [28], Gottschlich proposed another idea based on filter learning convolution comparison pattern. To detect a fingerprint spoofing, the authors compute the discrete cosine transform (DCT) from rotation invariant patches, and compute their binary patterns by comparing pairs of DCT coefficients. These patterns are gathered in a histogram, which was used to feed a linear SVM classifier.

Rattani *et al.* [82] introduced a scheme for automatic adaptation of a liveness detector to new spoofing materials in the operational phase. The aim of the proposed approach is to reduce the security risk posed by new spoof materials on an anti-spoofing system. The authors proposed a novel material detector specialized to detect new spoof materials, pointing out the need for retraining the system with the new material spotted.

Similar to [82], Rattani *et al.* [83] proposed an automatic adaptation anti-spoofing system composed of an open-set fingerprint spoofing detector and by a novel material detector, both based on Weibull-calibrated SVM (W-SVM) [88]. The novel material detector was built with a multi-class W-SVM, composed by an ensemble of pairs of 1-Class and binary

SVMs, whereas the open set fingerprint spoofing detector was trained with features based on textural [26], physiological [55] and anatomical [101] attributes.

Gragnaniello *et al.* [31] proposed a fingerprint spoofing detection based on both spatial and frequency information, in order to extract local amplitude contrast, and local behavior of the image, which were synthesized by considering the phase of some selected transform coefficients generated by the short-time Fourier transform (STFT). This information generates a bi-dimensional contrast-phase histogram, which was used to train a linear SVM classifier.

Kumpituck et al. [44] exploited an anti-spoofing schema based on wavelet decomposition and LBP operator. In this work, the authors extract LBP histograms from several wavelet sub-band images, which were concatenated and used to feed an SVM classifier. The authors also evaluated a more conventional approach that consists of calculating the energy from wavelet sub-bands instead of the LBP histograms. Experimental results show that wavelet-LBP descriptor achieved a better discrimination than wavelet-energy and LBP descriptors used separately, besides achieving competitive results with the state-of-the-art methods.

Finally, also departing from the traditional modeling, which uses basically texture patterns to characterize fingerprint images, Nogueira *et al.* [60] proposed a fingerprint anti-spoofing technique based on the concept of pre-trained convolutional neural networks. Basically, the authors use well-known CNN architectures in the literature such as AlexNet [42] and VGG [95] as their starting point for learning the network weights for fingerprint spoofing detection.

Marasco *et al.* [56] investigated two well-known CNN architectures, the GoogLeNet [99], CaffeNet [42], in order to analyze their robustness in detecting unseen spoof materials and fake samples from new sensors. As mentioned before, Menotti *et al.* [58] also proposed hyperparameter optimization of network architectures along with filter optimization techniques for detecting fingerprints presentation attacks.

2.3 Iris Presentation Attack Detection

Early work on iris spoofing detection dates back to the 1990's, when Daugman [14] discussed the feasibility of some attacks on iris recognition systems. In that work, he proposed to detect such attempts using the Fast Fourier Transform to verify the high frequency spectral magnitude.

According to Czajka [12], solutions for iris liveness detection can be categorized into four groups, as Cartesian product of two dimensions: type of measurement (passive or active) and type of model of the object under test (static / dynamic). Passive solutions mean that the object is not stimulated more than it is needed to acquire an iris image for recognition purpose. Hence, it typically means that no extra hardware is required to detect an attempted attack. Active solutions try to stimulate an eye and observe the response to that stimuli. It means that typically some extra hardware elements are required. In turn, the classification between static and dynamic objects means that the algorithm can detect

an attempted attack using just one (static) image from the biometric sensor or needs to use a sequence of images to observe selected dynamic features. In this section, we review only passive and static methods, which is the focus of this chapter.

In [66], Pacut *et al.* introduced three iris liveness detection algorithms based on the analysis of the image frequency spectrum, controlled light reflection from the cornea and pupil dynamics. These approaches were evaluated with paper printouts produced with different printers and printout carriers, and shown to be able to spoof two commercial iris recognition systems. A small hole was made in the place of the pupil, and this trick was enough to deceive commercial iris recognition systems used in their study. The experimental results obtained on the evaluation set composed of 77 pairs of fake and live iris images showed that the controlled light reflections and pupil dynamics achieve zero for both False Acceptance Rate and False Rejection Rate. In turn, two commercial cameras were not able to detect 73.1% and 15.6% of iris paper printouts and matched them to biometric references of authentic eyes.

Galbally *et al.* [23] proposed an approach based on 22 image quality measures (*e.g.*, focus, occlusion, and pupil dilation). The authors use sequential floating feature selection [79] to single out the best features, which were used to feed a quadratic discriminant classifier. To validate the proposed approach, the authors used the BioSec [19, 87] benchmark, which contains print-based iris spoofing attacks. Similarly, Sequeira *et al.* [91] also exploited image quality measures [23] and three different classification techniques, validating the work on BioSec [19, 87] and Clarkson [89] benchmarks and introducing the MobBIOfake benchmark comprising 800 iris images. Sequeira *et al.* [92] extended upon previous work using a feature selection step to obtain a better representation to detect an attempted attack. The authors also applied iris segmentation [59] to obtain the iris contour and adapted the feature extraction processes to the resulting non-circular iris regions.

In [107], Wei *et al.* addressed the problem of iris liveness detection based on three texture measures: iris edge sharpness (ES), iris-texton feature for characterizing the visual primitives of iris texture (IT) and using selected features based on co-occurrence matrix (CM). In particular, they used fake iris wearing color and textured contact lenses. The experiments showed that the ES feature achieved comparable results to the state of the art methods at that time, and the IT and CM measures outperformed the state of the art algorithms.

Czajka [10] proposed a solution based on frequency analysis to detect printed irises. The author associated peaks found in the frequency spectrum to regular patterns observed for printed samples. This method, tuned to achieve close-to-zero false rejection rate (*i.e.,* not introducing additional false alarms to the entire system), was able to detect 95% of printed irises. This paper also introduced the Warsaw LivDet-Iris-2013 dataset containing 729 fake images and 1,274 images of real eyes.

Texture analysis has also been explored for iris spoofing detection. In the MobILive [93] iris spoofing detection competition, the winning team relied upon three texture descriptors: LBP [62], LPQ [64] and Binary Gabor Pattern (BGP) [112]. Sun *et al.* [98] recently

proposed a general framework for iris image classification based on a Hierarchical Visual Codebook (HVC). The HVC encodes the texture primitives of iris images and is based on two existing bag-of-words models. The method achieved a state-of-the-art performance for iris spoofing detection, among other tasks related to iris recognition.

Doyle *et al.* [16] proposed a solution based on modified Local Binary Patterns (mLBP) [63] descriptor. In this work, the authors show that although it is possible to obtain good classification results using texture information extracted by the mLBP descriptor, when lenses produced by different manufacturers are used, the performance of this method drops significantly. They report 83% and 96% of correct classification when measured on two separated datasets, and a significant drop in accuracy when the same method was trained on the one dataset and tested on the other dataset: 42% and 53%, respectively. This cross-dataset validation has been shown to be very challenging and seems to be recommended in several validation setups for presentation attack detection. Yadav *et al.* [110] extended upon the previous work by analyzing the effect of soft and textured contact lenses on iris recognition.

In [81], Raja *et al.* proposed an anti-spoofing method based on Eulerian Video Magnification (EVM) [5], which was applied to enhance the subtle phase information in the eye region. The authors proposed a decision rule based on cumulative phase information, which was applied by using a sliding window approach upon the phase component for detecting the rate of the change in the phase with respect to time.

Raghavendra *et al.* [80] proposed a novel spoofing detection scheme based on a multi-scale version of the Binarized Statistical Image Features (BSIF) and linear Support Vector Machine (SVM). Gupta *et al.* [33] proposed an anti-spoofing technique based on local descriptors such as LBP [61], HOG [13], and GIST [65], which provide a representation space by using attributes of the images such as color, texture, position, spatial frequency, and size of objects present in the image. The authors used the feature vectors produced by the three descriptors to feed a nonlinear classifier and decide whether an image under analysis is fake.

Czajka [11] proposed an iris spoofing detection based on pupil dynamics. In that work, the author used the pupil dynamics model proposed by Kohn and Clynes [39] to describe its reaction after a positive light stimuli. To decide whether the eye is alive, the author used variants of the SVM to classify feature vectors that contain the pupil dynamic information of a target user. This work has been further extended to a mixture of negative and positive light stimuli [12] and presented close-to-perfect recognition of objects not reacting to light stimuli as expected for a living eye.

Finally, Lovish *et al.* [50] proposed a cosmetic contact lens detection method based on Local Phase Quantization and Binary Gabor Patterns, which combines the benefits of both LBP and Gabor filters [112]. The histograms produced for both descriptors were concatenated and used to build a classification model based on SVM algorithm.

Similarly to the approaches tackling the presentation attack problem in fingerprint and faces, handcrafted texture features seem to be the preferred choice in iris spoofing

detection. Methods inspired by LBP, visual codebooks and quality metrics are the most popular methods so far. In this sense, the works of Menotti *et al.* [58] and Silva *et al.* [94], which exploit data-driven solutions for this problem, are sufficiently different from the previous methods and present very promising results.

2.4 Unified Frameworks to Presentation Attack Detection

Galbally *et al.* [22] proposed a general approach based on 25 image quality features to detect attempt attacks in face, iris and fingerprint biometric systems simultaneously. Evaluations performed upon popular benchmarks for three modalities show that this approach is highly competitive, considering the state-of-the-art methods dedicated for single modalities.

In [30], Gragnaniello *et al.* evaluated several local descriptors for face-, fingerprint- and iris-based biometrics in addition to the investigation of promising descriptors using the Bag-of-Visual-Word (BoVW) model [97], Scale-Invariant Feature Transform (SIFT) [51], DAISY [105], and the Shift-Invariant Descriptor (SID) [40].

Menotti *et al.* [58] (mentioned earlier in this section) showed that the combination of architecture optimization and filter optimization provides better comprehension of how these approaches interplay for face, iris and fingerprint PAD, and also outperforms the best known approaches for several benchmarks.

In this chapter, we decided to explore data-driven solutions for spoofing detection in different modalities based on deeper architectures than the one used in [58] and evaluate the effects of such decision. Our objective is to show the potential of this approach but also highlight its limitations, especially related to cross-dataset experiment.

3 Methodology

In this section, we present the convolutional neural network that we adopted to PAD for face, fingerprint and iris. Our objective is simply to show that this new trend in the literature is also relevant for the task of presentation attack detection and that research in this direction needs to be considered. At the same time, we also show that even when adopting a powerful image classification technique such as deep neural networks, we still cannot deal effectively with the very challenging cross-dataset problem. As a result, it is clear that the research community now needs to shift its attention to cross-dataset validation setups (or, more general, open-set classification) as they are closer to real-world operational conditions when deploying biometric systems.

3.1 Network Architecture

For this work, we adopted the VGG network architecture proposed by [96]. However, that network was first proposed for object recognition and not presentation attack detection. Therefore, for each problem of interest (PAD in face, iris and fingerprint), we adapt the

network's architecture as well as fine-tune its weights to our two-class problem of interest. Training the network from scratch to our problem is also a possibility if enough training samples (normal and presentation attack samples) are available. However, as this is not often the case in this area, it is recommended to start the network weights with a related (source) problem and then adapt these weights with training examples of a target problem.

Figure 2 depicts the network architecture we adopted in this work. During training, the network's input consists of fixed-size 224×224 RGB images which go through a stack of convolutional layers comprising filters with a very small receptive field (3×3). In this network, the convolution stride is fixed to one pixel and the spatial zero-padding for convolutional operation is also of one pixel. There are five max-pooling layers in this network (carefully placed after some convolution layers). The max-poolings are performed over a 2×2 pixel window, with stride 2.

The stack of convolutional layers is followed by three fully-connected (FC) layers: the first two have 4,096 units each, while the the third layer performs the 2-way spoofing classification problem of our interest (originally this was an FC layer with 1,000 units for the ImageNet 1,000-way classification problem). The final layer is the soft-max layer translating the outputs of 2-unit layer into a posterior probabilities of class membership. Each unit in the hidden layers has a rectified linear (ReLU) activation function [43]. The depth of convolution layers or, in other words, their number of channels, starts with 64 and is iteratively doubled after each max-pooling layer to a maximum of 512.

3.2 Training and Testing

For training, we start with the network trained to a source problem whenever it is possible. To detect presentation attacks with faces, we initialize the network with the weights learned for face recognition [68]. However, the closest problem we had for iris and fingerprints was general image classification. Therefore, presentation attack detection for iris and fingerprints is performed by the network initialized with the weights pre-computed for the ImageNet classification problem. The first convolutional layers act mostly as a generic feature detectors (such as edges) and are suitable for different computer vision tasks. However, each next convolutional layer is more context-focused and extracts features that are task-related. Hence, using last layers trained for general object recognition in visual spoofing detection is not optimal, and a large improvement may be achieved by specializing the network. Certainly, a preferable solution is to initialize the weights with those used in networks solving iris- and fingerprint-related tasks, as the network would have been specialized to this type of imagery. However, since training of such networks from the scratch requires a lot data and effort, it is still a good move to initialize own network with image-related weights than just purely at random and tune its weights if there is not enough available training data.

Once a source set of weights to initialize the network is chosen, the fine-tuning follows a standard procedure: selects the training set of the target domain and uses it to perform

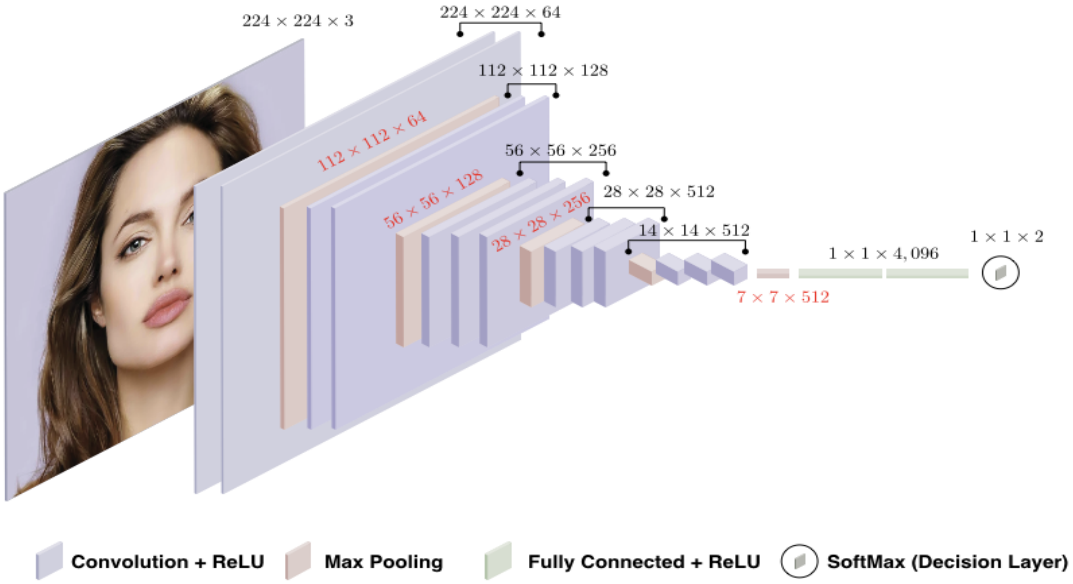

Figure 2: Adopted network architecture, originally proposed for object recognition by the Visual Geometry Group and thus referred to as VGG network [96].

forward passes and back-propagation in the network. The test for an input image is straightforward. Just resize it to the network's input size and feed it to the network. As the network has been fully adapted to the target problem of interest, it will already produce a two-class output.

More specifically, for the cases of fingerprints, the input images in a dataset are center-cropped and resized to 224 × 224 pixels, which is the standard input size of the VGG network. The centering happens through calculating the average of black pixels in the binary fingerprint image and keeping all the rows/columns with a density of black pixels greater than the global image average plus or minus 1.8 standard deviations of each respective row/column. This is used to eliminate the borders without any useful information. For optimizing the network in the fingerprint case, we use the standard SGD solver implemented in Caffe with the following hyperparameters: base learning rate of 0.0001, step lr policy, step size of 2,000, momentum of 0.9, weight decay of 0.0002, gamma of 0.5 and maximum of 2,001 iterations.

In the case of faces, we center-cropped the images based on the eye coordinates calculated with the aid of Face++[1]. Upon center-cropping, the image is resized to 224 × 224 pixels. For optimizing the network in the face case, we use the standard SGD solver im-

[1] `http://www.faceplusplus.com/`

plemented in Caffe with the following hyper parameters: base learning rate of 0.001, step lr policy, step size of 1,000, momentum of 0.9, weight decay of 0.0005, gamma of 0.001 and maximum number of iterations of 4,000.

For irises, we resize the images to the network's standard input size of 224×224 pixels and employ the same parameters as for the face optimization problem.

3.3 Memory Footprint

The chosen network has an average size of 140 MB. Most of its parameters (and memory) are in the convolution and fully-connected layers. The first FC layer contains 100M weights, out of a total of 134M for the entire adapted network.

4 Metrics and Datasets

In this section, we describe the benchmarks (datasets) and selected accuracy estimators considered in this work. All datasets used in this chapter were freely available to us and we believe that it is the case for other researchers upon request sent directly to their creators. Datasets composing our testing environment are the most commonly used benchmarks to evaluate presentation attack detection for face, iris and fingerprints. Since all the benchmarks have been already divided by their creators into training and testing subsets, we decided to follow these divisions. Each training subset was divided by us into two disjoint subsets multiple times to perform cross-validation-based training to increase generalization capabilities of the winning model and to minimize an overfitting. The results reported further in this chapter are those obtained on testing sets. The next subsections characterize briefly all datasets and Table 1 shows their major features, in particular the number of samples in each benchmark and their assignment to training and testing subsets.

4.1 Video-based Face Spoofing Benchmarks

In this chapter, we use two benchmarks used to evaluate the performance of PAD algorithms for face modality, Replay-Attack [9] and CASIA Face Anti-Spoofing [113] datasets. These datasets contain five types of attempted attacks performed with fake samples presenting different qualities.

4.1.1 Replay-Attack [9]

This benchmark contains short video recordings of both valid accesses and video-based attacks of 50 different subjects. To generate valid access videos, each person was recorded in two sessions in a controlled and in an adverse environment with a regular webcam. Then, spoofing attempts were generated using three techniques:

- *print attack*: hard copies of high-resolution digital photographs were presented to the acquisition sensor; these samples were printed with a Triumph-Adler DCC 2520 color laser printer;
- *mobile attack*: videos displayed on an iPhone screen were presented to the acquisition sensor; these videos were taken also with the iPhone;
- *high-definition attack*: high resolution photos and videos taken with an iPad were presented to the acquisition sensor using the iPad screen.

4.1.2 CASIA [113]

This benchmark was based on samples acquired from 50 subjects. Genuine images were acquired by three different sensors presenting different acquisition quality (from low to high): "long-time-used USB camera", "newly bought USB camera", and Sony NEX-5. Pixel resolution of images was either 640×480 (both webcams) or 1920×1080 (Sony sensor). Sony images were cropped to 1280×720 by the authors. During the acquisition, subjects were asked to blink. Three kinds of presentation attacks were carried out:

- *warped photo attack*: high quality photos were printed on a copper paper and videos were recorded by Sony sensor; the printed images were intentionally warped to imitate face micro-movements,
- *cut photo attack*: eyes were cut from the paper printouts and an attacker hidden behind an artifact imitated the blinking behavior when acquiring the video by the Sony sensor,
- *video attack*: high quality genuine videos were displayed on an iPad screen of 1280×720 pixel resolution.

The data originating from 20 subjects was selected for a training set, while remaining samples (acquired for 30 subjects) formed the testing set.

4.2 Fingerprint Spoofing Benchmarks

Two datasets used in Liveness Detection Competitions (LivDet, `www.livdet.org`) were employed in this chapter. LivDet is a series of international competitions that compare presentation attack methodologies for fingerprint and iris using a standardized testing protocol and large quantities of spoof and live samples. All the competitions are open to all academic and industrial institutions which have software-based or system-based biometric liveness detection solutions. For fingerprints, we use datasets released in 2009 and 2013.

The LivDet2009 benchmark [57] consists of three subsets of samples acquired by Biometrics FX2000, CrossMatch Verifier 300 LC and Identix DFR2100. Both the spatial scanning resolution and pixel resolution vary across subsets, from 500 DPI to 686 DPI,

and from 312×372 to 720×720 pixels, respectively. Three different materials were used to prepare spoofs: Play-Doh, gelatin and silicone.

The LivDet2013 benchmark [27] contains four subsets of real and fake fingerprint samples acquired by four sensors: Biometrika FX2000, Italdata ET10, Crossmatch L Scan Guardian, and Swipe. Inclusion of samples from the Swipe sensor is especially interesting, since it requires – as the name suggests – swiping a finger over the small sensor. This makes the quality of spoofs relatively different when compared to the regular, flat sensors requiring only touching the sensor by the finger. For a more realistic scenario, fake samples acquired by Biometrika and Italdata were generated without user cooperation, while fake samples acquired by Crossmatch and Swipe were generated with user cooperation. Several materials for creating the artificial fingerprints were used, including gelatin, silicone, latex, among others. The spatial scanning resolution varies from a small 96 DPI (the Swipe sensor) to 569 (the Biometrika sensor). The pixel resolution is also heterogeneous: from relatively non-standard 208×1500 to pretty large 800×750. This makes the cross-subset evaluation quite challenging.

4.3 Iris Spoofing Benchmarks

To evaluate our proposed method in detecting iris presentation attack, we used two benchmarks: AVTS [87] and a new dataset Warsaw LivDet2015, which is an extension of Warsaw LivDet2013 [10]. These datasets contain attempted attacks performed with printed iris images, which were produced using different printers and paper types.

4.3.1 AVTS [87]

This benchmark was based on live samples collected for 50 volunteers under the European project BioSec (Biometrics and Security). To create spoofing attempts, the authors tested two printers (HP Deskjet 970cxi and HP LaserJet 4200L), various paper types (*e.g.*, cardboard as well as white, recycle, photo, high resolution and butter papers), and a number of pre-processing operations. The combination that gave the highest probability of image acquisition by the LG IrisAccess EOU3000 sensor used in the study was selected for a final dataset collection. The authors printed their samples with the inkjet printer (HP Deskjet 970cxi) on a high resolution paper and applied an Open-TopHat pre-processing to each image prior printing. The pixel resolution of each image was 640×480, which is recommended by ISO/IEC as a standard resolution for iris recognition samples.

4.3.2 Warsaw LivDet2015

This dataset is an extension of the LivDet-Iris 2013 Warsaw Subset [10] and was used in 2015 edition of LivDet-Iris competition (`www.livdet.org`). It gathers 2854 images of authentic eyes and 4705 images of the paper printouts prepared for almost 400 distinct eyes. The photographed paper printouts were used to successfully forge an example commercial

Table 1: Main features of the benchmarks considered herein.

Modality	Benchmark	Color	Dimension	# Training			# Testing		
			$cols \times rows$	Live	Fake	Total	Live	Fake	Total
Face	Replay-Attack	Yes	320×240	600	3000	3600	4000	800	4800
	CASIA	Yes	1280×720	120	120	240	180	180	360
Iris	Warsaw LivDet2015	No	640×480	852	815	1667	2002	3890	5892
	AVTS	No	640×480	200	200	400	600	600	1200
Fingerprint	LivDet2009: CrossMatch	No	640×480	500	500	1000	1500	1500	3000
	LivDet2009: Identix	No	720×720	375	375	750	1125	1125	2250
	LivDet2009: Biometrika	No	312×372	500	500	1000	1500	1500	3000
	LivDet2013: Biometrika	No	312×372	1000	1000	2000	1000	1000	2000
	LivDet2013: CrossMatch	No	800×750	1250	1000	2250	1250	1000	2250
	LivDet2013: Italdata	No	640×480	1000	1000	2000	1000	1000	2000
	LivDet2013: Swipe	No	208×1500	1250	1000	2250	1250	1000	2250

iris recognition system (*i.e.*, samples used in real and successful presentation attacks). Two printers were used to generate spoofs: HP LaserJet 1320 and Lexmark C534DN. Both real and fake images were captured by an IrisGuard AD100 biometric device with liveness detection functionality intentionally switched off. To get a free copy of this dataset follow the instructions given at Warsaw's lab webpage `http://zbum.ia.pw.edu.pl/EN/node/46`.

4.4 Error Metrics

In this chapter we use the error metrics that are specific to presentation attack detection, and partially considered by ISO/IEC in their PAD-related standards [35].

Attack Presentation Classification Error Rate (APCER): proportion of *attack presentations* incorrectly classified as *bona fide (genuine) presentations* at the PAD subsystem in a specific scenario. This error metric is analogous to false match rate (FMR) in biometric matching, that is related to false match of samples belonging to two different subjects. As FMR, the APCER is a function of a decision threshold τ.

Bona Fide Presentation Classification Error Rate (BPCER): proportion of *bona fide (genuine) presentations* incorrectly classified as *presentation attacks* at the PAD subsystem in a specific scenario. This error metric is analogous to false non-match rate (FNMR) in biometric matching, that is related to false non-match of samples belonging to the same subject. Again, the BPCER is a function of a decision threshold τ.

Half Total Error Rate (HTER): combination of APCER and BPCER in a single error rate with a decision threshold as an argument:

$$\text{HTER}(\tau) = \frac{\text{APCER}(\tau) + \text{BPCER}(\tau)}{2} \tag{1}$$

5 Results

In this section, we present and discuss the experimental results of the proposed method. Sections 5.1, 5.2 and5.3 show the performance results and the experimental protocols employed to validate the performance of the proposed methodology.

5.1 Face

In this section, we present the results of our proposed PAD for face modality. The experiments are conducted considering the original protocol of the datasets used in this chapter (cf., Section 4), as well cross-dataset protocol, hereafter referred to as same-dataset and cross-dataset protocols, respectively. In general, a prime requirement of most machine learning algorithms is that both training and testing sets are independent and identically distributed. But, unfortunately, it does not always happen in practice – subsets can be identically distributed (*e.g.*, captured using the same sensor and in the same environment conditions), but totally dependent due to adding of bias in the data (*e.g.*, some dirt in the biometric sensor used to capture both subsets, identities present in two subsets, artifacts added during the attack simulations, etc.). In addition, the effects of the closed-world assumption [88] may mislead us to believe that a given approach is perfect when in fact its performance can be disastrous when deployed in practice for unknown presentation attacks. In this context, both same-dataset and cross-dataset are key experimental protocols in determining more accurate detection rates of an anti-spoofing system when operating in less controlled scenarios with different kinds of attacks and sensors.

Same-dataset results. Table 2 shows the results for Replay-Attack and CASIA datasets, considering that training and testing is performed on the same dataset. The VGG network was able to detect all kinds of attempted attacks present in the Replay-Attack dataset, and also to detect two methods of attempted attacks (hand-based and fixed-support attacks), which were confirmed by the perfect classification result (HTER of 0.0%). Considering the CASIA dataset, the proposed method obtained an HTER of 6.67%. The performance achieved by the proposed method on this dataset can be explained by the high degree of variability present in the CASIA dataset (*e.g.*, different kinds of attack and resolution) that makes this dataset more challenging. In both datasets, we use the k-fold cross-validation technique ($k = 10$) to built a classification model using the training set, and also the development set whether it is available. Figures 3 and 4 present empirical

distributions of the difference between two CNN output nodes and the corresponding ROC curves.

Table 2: Performance results obtained in the **same-dataset** evaluations of the **face PAD**. Pointers to plots presenting Receiver Operating Characteristics (ROC) and empirical Probability Distribution Functions (ePDF) are added in the last column.

	APCER (%)	**BPCER (%)**	**HTER (%)**	**ROC and ePDF**
Replay-Attack	0.00	0.00	0.00	Fig. 3
CASIA	0.00	13.33	6.67	Fig. 4

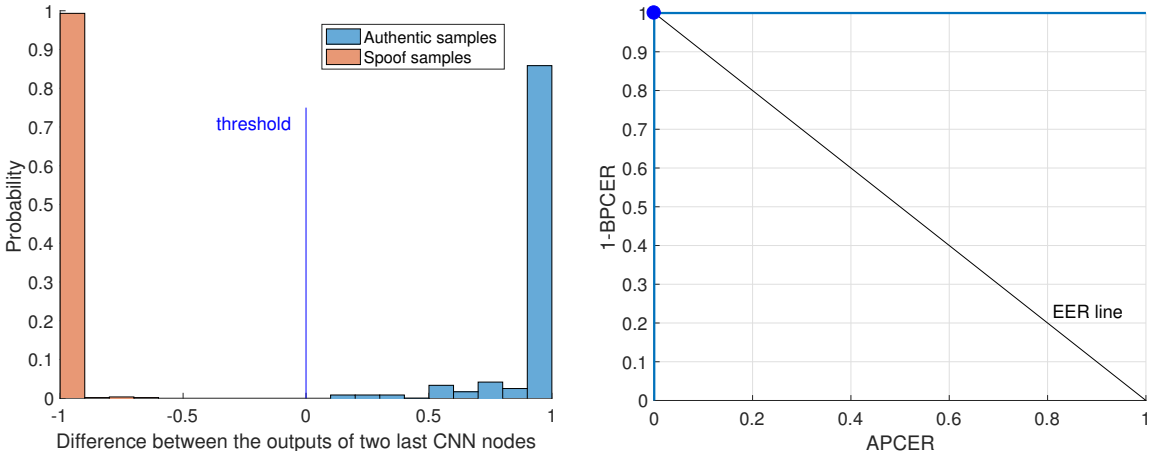

Figure 3: **Left:** Empirical probability distributions (ePDF) of the difference between two CNN output nodes (after softmax) obtained separately for authentic and spoof **face** samples. **Right:** ROC curve. Variant: training on **Replay-Attack**, testing on **Replay-Attack**. The threshold shown in blue color on the left plot and the blue dot on the ROC plot correspond to the approach when the predicted label is determined by the node with the larger output.

Cross-dataset results. Table 3, Fig. 5 and Fig. 6 show the results obtained in cross-dataset evaluation protocol. We can clearly see a dramatic drop in the performance when we train and test on different datasets. Several sources of variability between the datasets may contribute to this result. The first one is that the datasets contain different kinds of attack. The Replay-Attack dataset contains three kinds of attacks (high definition-based, mobile-based and video-based attacks) while the CASIA dataset includes additional two kinds of attack (warp-based and cut-based photo attacks). Another source is the fact

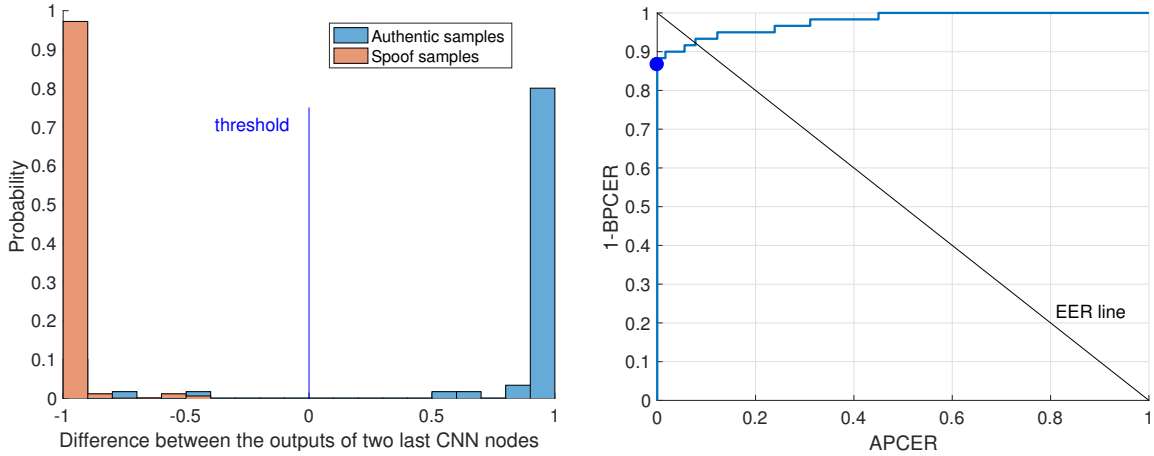

Figure 4: Same as Fig. 3 except the variant: training on **CASIA**, testing on **CASIA**.

that data comes from different sensors, which potentially produce samples with different resolutions, color distributions, backgrounds, etc. The VGG architecture finds very specific features and even when it is tuned to the specific problem, it does not generalize well to be agnostic to specific properties of data acquisition process.

Table 3: Performance results obtained with the **cross-dataset** evaluations of **face PAD** and using the overall testing set of each dataset.

Training	Test	APCER (%)	BPCER (%)	HTER (%)	ROC and ePDF
Replay-Attack	CASIA	42.67	51.67	47.16	Fig. 5
CASIA	Replay-Attack	89.44	10.0	49.72	Fig. 6

5.2 Fingerprints

This section presents how our VGG-based approaches perform in detection of fingerprint attack presentation. As for experiments with face benchmarks, we used the training subsets (as defined by dataset creators) to make a cross-validation-based training, and separate testing subsets in final performance evaluation. Fingerprint benchmarks are composed of subsets gathering mixed attacks (for instance glue, silicone or gelatin artifacts) and acquired by different sensors (cf. Table 1).

Same-sensor results. In this scenario, samples acquired by different sensors are not mixed together. That is, if the classifier is trained with samples acquired by sensor X, only sensor X samples are used in both the validation and final testing. As in previous

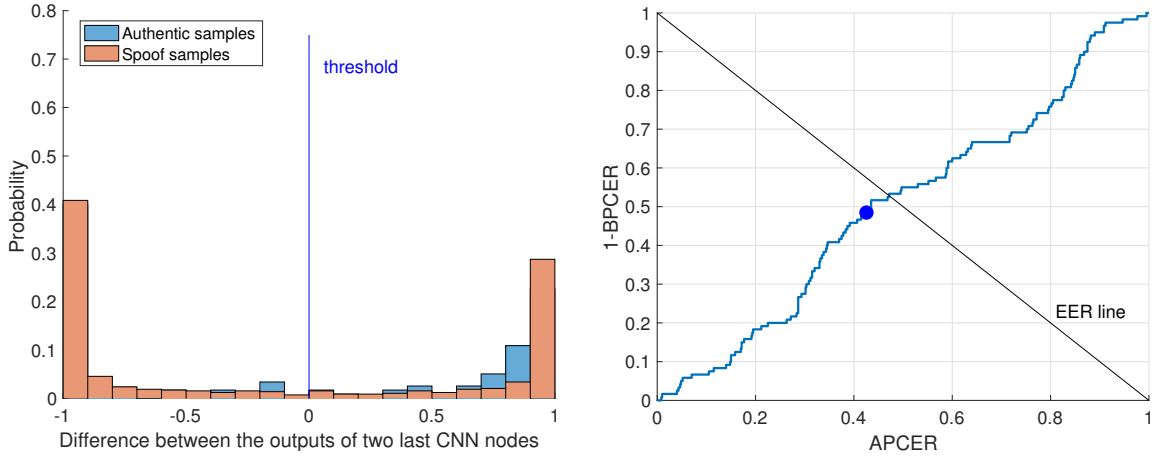

Figure 5: Same as Fig. 3 except the variant: training on **Replay-Attack**, testing on **CASIA** (cross-dataset testing).

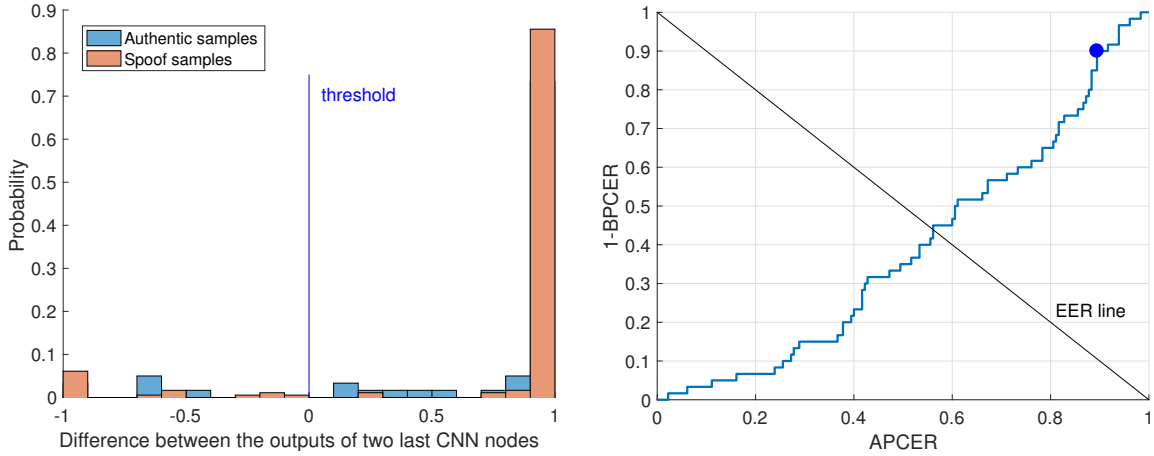

Figure 6: Same as Fig. 3 except the variant: training on **CASIA**, testing on **Replay-Attack**.

experiments, 10 statistically independent estimation-validation pairs of non-overlapping subsets were created, and the solution presenting the lowest HTER over ten validations was selected for testing. Table 4 as well as Figures 7 and 8 show the same-sensor testing results averaged over all sensors (used to build a given dataset) and presented for each benchmark separately. These results suggest that the older benchmark (LivDet2009) is relatively difficult since almost 20% of spoofing samples were falsely accepted in a solution that falsely rejects only 3.45 % of authentic examples.

Table 4: Performance results obtained in **same-dataset** evaluations of **fingerprint PAD** using a part of testing samples acquired by the same sensor as in the training procedure. Results are averaged over all subsets representing different sensors.

Training	Testing	APCER (%)	BPCER (%)	HTER (%)	ROC and ePDF
LivDet2009	LivDet2009	19.37	3.45	11.4	Fig. 7
LivDet2013	LivDet2013	6.8	2.79	4.795	Fig. 8

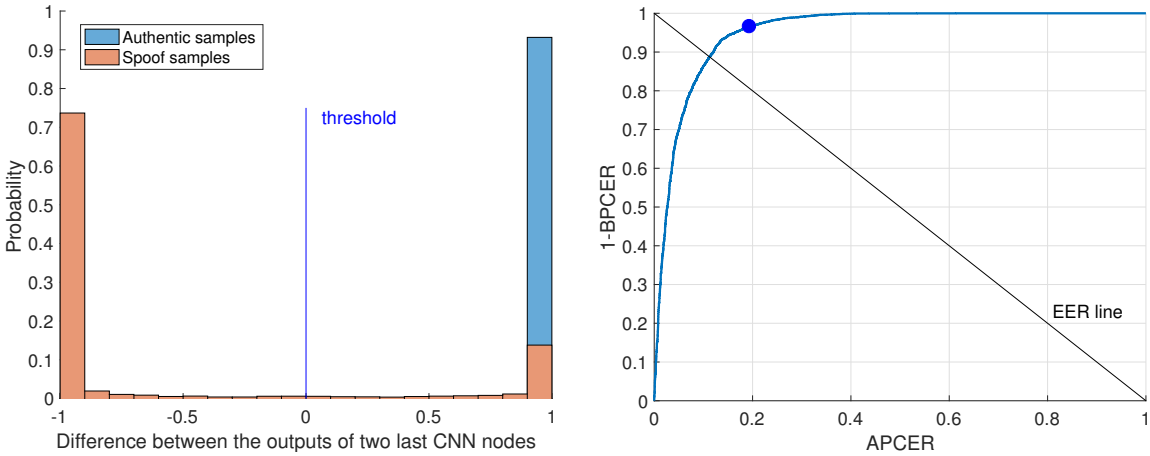

Figure 7: **Left:** Empirical distributions of the difference between two CNN output nodes (after softmax) obtained separately for authentic and spoof **fingerprint** samples. **Right:** ROC curve. Variant: training on **LivDet2009**, testing on **LivDet2009**. As in previous plots, the threshold shown in blue color on the left plot and the blue dot on the ROC plot correspond to the approach when the predicted label is determined by the node with the larger output.

Cross-sensor results. For cross-sensor analysis, the newer benchmark (LivDet2013) was selected. Each subset (estimation, validation and testing) was divided into two disjoint subsets of samples: a) acquired by ItalData and Swipe sensors, and b) acquired by Biometrika and CrossMatch sensors. Table 5 shows that, as with the other modalities, we can observe serious problems with recognition of both artifacts or genuine samples (two first rows of Table 5). Figures 10 and 9, illustrating these results, suggest that a better balance between APCER and BPCER can be found if there is a possibility to adjust the acceptance threshold.

For completeness, same-sensor results are also presented on this dataset in two last rows of Table 5, and in Figs. 12 and 11. As expected, a solution based on deep network achieves much better accuracy when the type of sensor in known.

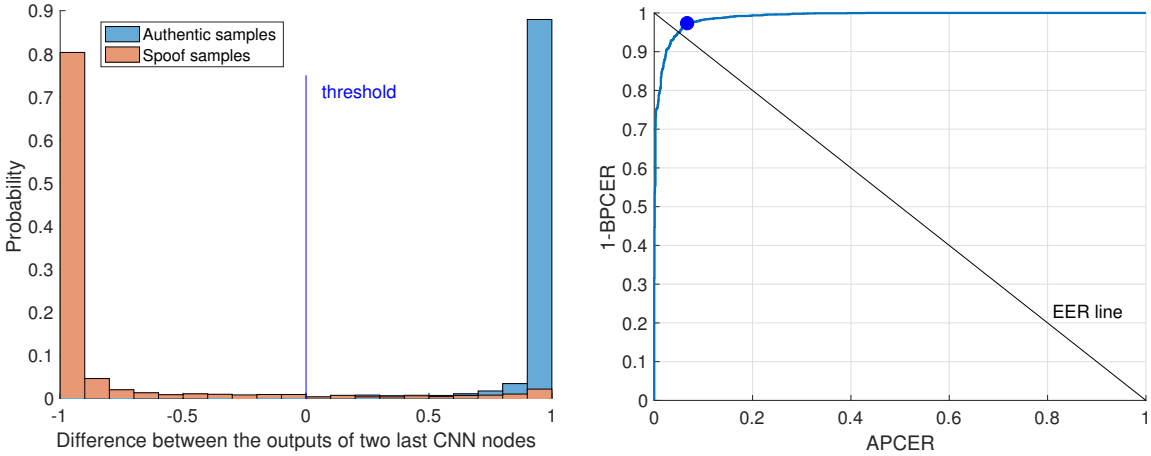

Figure 8: Same as Fig. 7 except the variant: training on **LivDet2013**, testing on **LivDet2013**.

Table 5: Performance results obtained in **cross-dataset** evaluations of **fingerprint PAD** using a part of testing samples acquired by different sensor as in the training procedure. All data comes for LivDet2013 fingerprint benchmark. IS = Italdata+Swipe, BC = Biometrika+CrossMatch.

Training	Testing	APCER (%)	BPCER (%)	HTER (%)	ROC and ePDF
IS	BC	24.9	4.01	14.1	Fig. 9
BC	IS	2.8	75.6	39.18	Fig. 10
IS	IS	3.4	2.37	2.88	Fig. 11
BC	BC	2.65	13.1	7.87	Fig. 12

5.3 Iris

This last section presents the results of iris presentation attacks detection. Two iris PAD benchmarks were used, as described in Section 4), and both same-dataset and cross-dataset experiments were carried out. Each dataset (Warsaw LivDet2015 and AVTS) are already split by their creators into training and testing subsets. We followed this split and used the testing subset only in final performance evaluation. The training subset, used in method development, was randomly divided 10 times into estimation and validation disjoint subsets used in cross-validation when training the classifiers.

The average HTER's over 10 splits calculated for validation subsets were approx. 0.0001 and 0.0 for Warsaw and AVTS datasets, respectively. HTER = 0.0 for 5 out of 10 splits of Warsaw training dataset. This means that the VGG-based feature extractor followed by a classification layer trained on our data was perfect on the AVTS dataset, and also it was

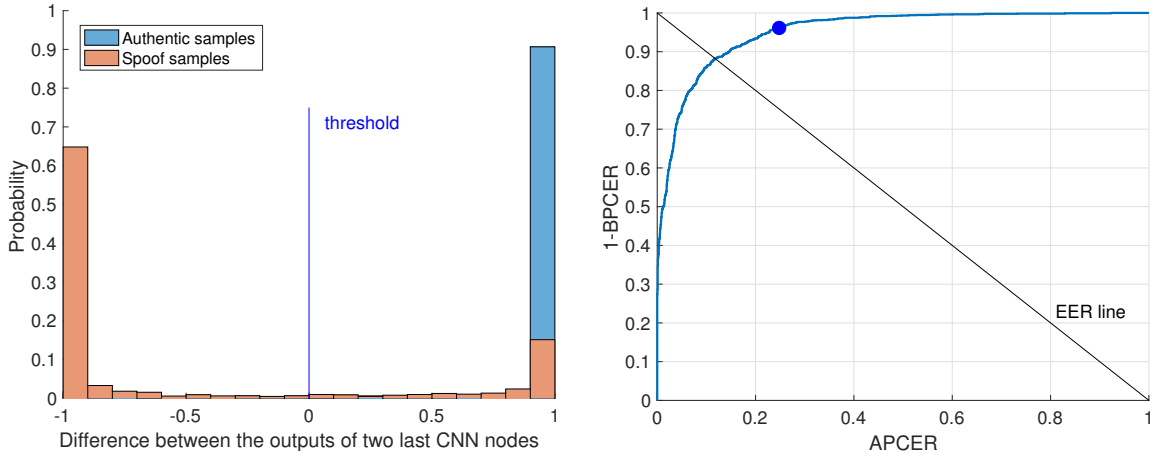

Figure 9: Same as Fig. 7 except the variant: training on **Italdata+Swipe**, testing on **Biometrika+CrossMatch**.

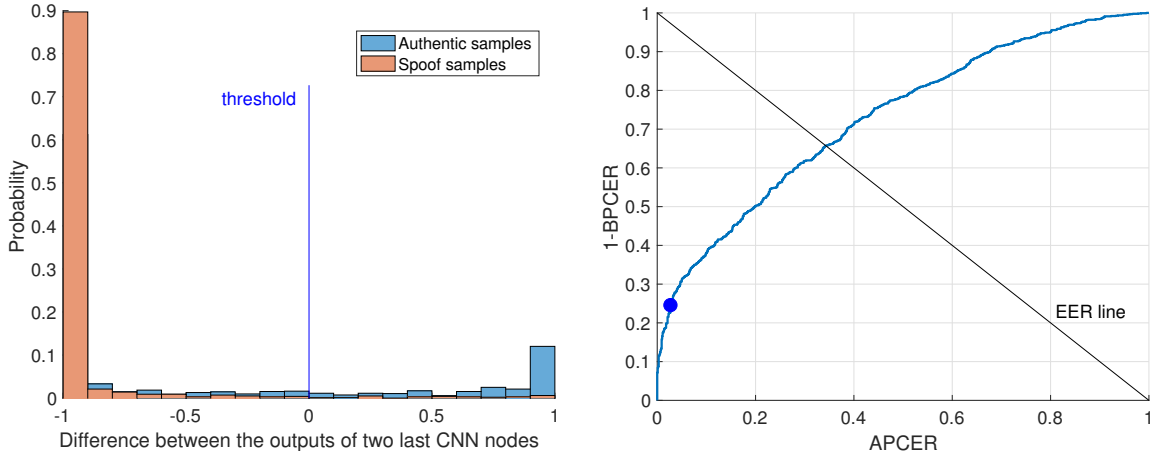

Figure 10: Same as Fig. 7 except the cross-sensor that training is realized on samples composed **Biometrika+CrossMatch**, testing on **Italdata+Swipe**.

perfect on half of the splits of the Warsaw benchmark. Since there is no "best split" for either of two datasets, we picked one trained solution presenting perfect performance on the training subsets to evaluate them on the test sets.

Same-dataset results. Table 6 presents the testing results obtained in the scenario when both training and testing sets come from the same benchmark. APCER and BPCER refer to classification task, that is each sample belonging to the testing set was classified to one of two classes (authentic or presentation attack) based on posteriori probabilities of

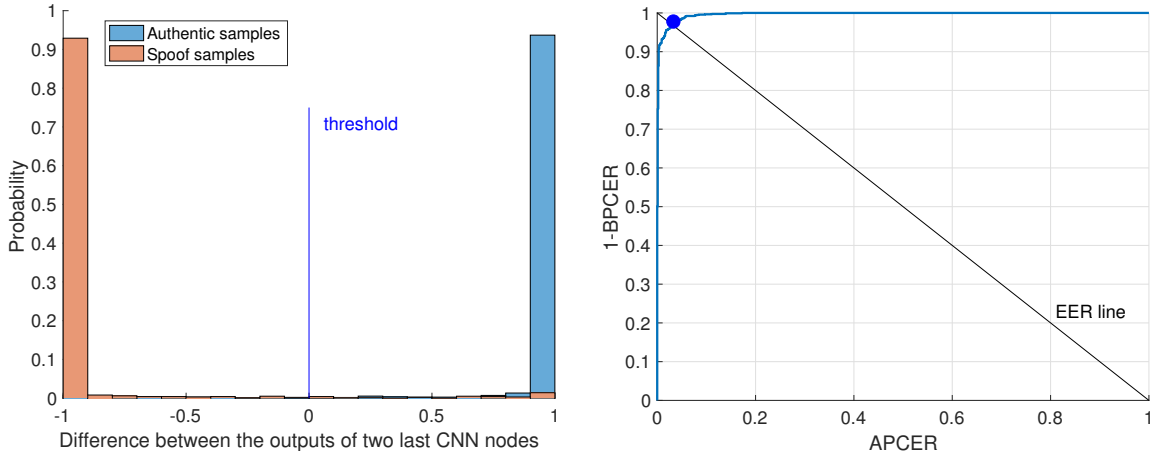

Figure 11: Same as Fig. 7 except the variant: training on **Italdata+Swipe**, testing on **Italdata+Swipe**.

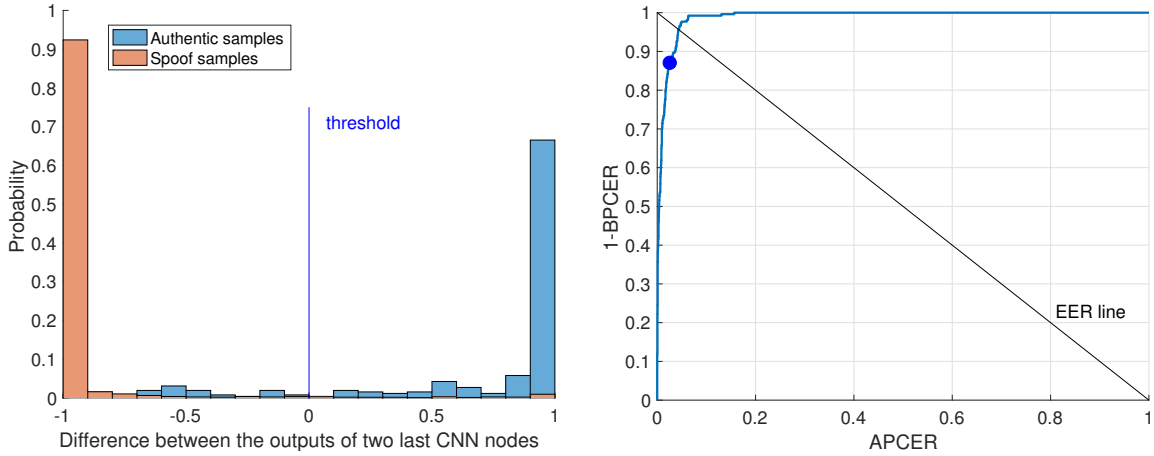

Figure 12: Same as Fig. 7 except the variant: training on **Biometrika+CrossMatch**, testing on **Biometrika+CrossMatch**.

class membership estimated by the softmax layer of the trained network. Hence, single APCER and BPCER (point estimators) are presented since this protocol is equivalent to a single acceptance threshold. The results obtained in this scenario are astonishing: the classifiers trained on disjoint subsets of samples originating from the same dataset are either perfect (ATVS benchmark) or close to perfect (a perfect recognition of spoofing samples of Warsaw benchmark with only 0.15% of authentic samples falsely rejected). Figures 13 and 14 present empirical distributions of the difference between two CNN output nodes and the corresponding ROC curves. The distributions are well separated for both

benchmarks, suggesting high performance of the VGG-based solution applied for known spoofing samples.

Table 6: Performance results obtained in **same-dataset** evaluations of **iris PAD** using the overall testing set of each dataset.

Training	Testing	APCER (%)	BPCER (%)	HTER (%)	ROC and ePDF
Warsaw	Warsaw	0.0	0.15	0.075	Fig. 13
ATVS	ATVS	0.0	0.0	0.0	Fig. 14

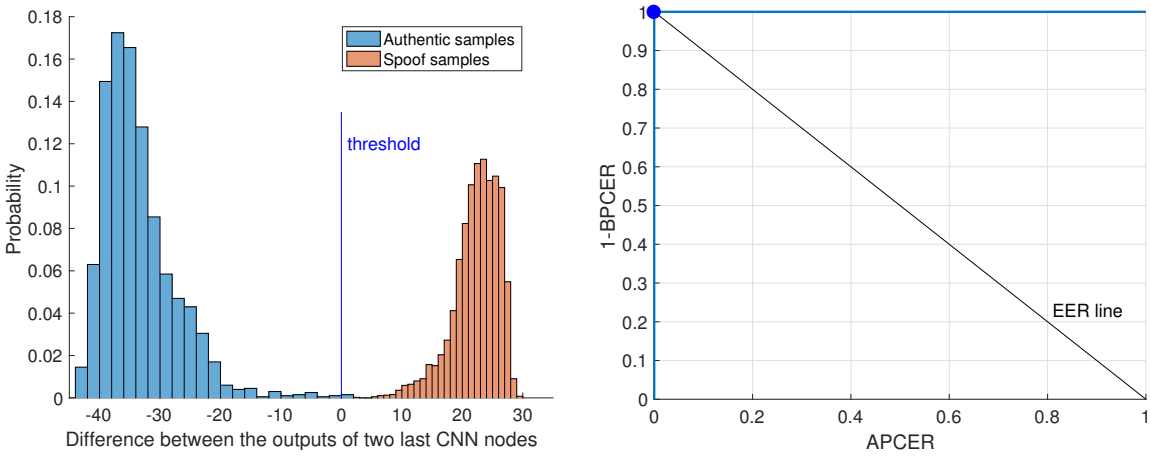

Figure 13: **Left:** empirical distributions of the difference between two CNN output nodes (before softmax) obtained separately for authentic and spoof **iris** samples. **Right:** ROC curve. Variant: training on **Warsaw LivDet2015**, testing on **Warsaw LivDet2015**. The threshold shown in blue color on the left plot and the blue dot on the ROC plot correspond to the approach when the predicted label is determined by the node with the larger output.

Cross-dataset results. Table 7 shows how catastrophically bad this method may be if tested on **cross-dataset** samples. ATVS and Warsaw samples differ significantly in terms of image properties such as contrast and visibility of iris texture. Especially, all the printouts used to produce Warsaw fake samples were able to spoof an example commercial iris recognition system, which is not the case in the ATVS benchmark. Hence, due to non-accidental quality of Warsaw samples, this database seems to be more realistic and more difficult to process than the ATVS. Indeed, training on Warsaw (the "difficult" benchmark) and testing on ATVS (the "easier" benchmark) yields good results. Figure 15 presents well

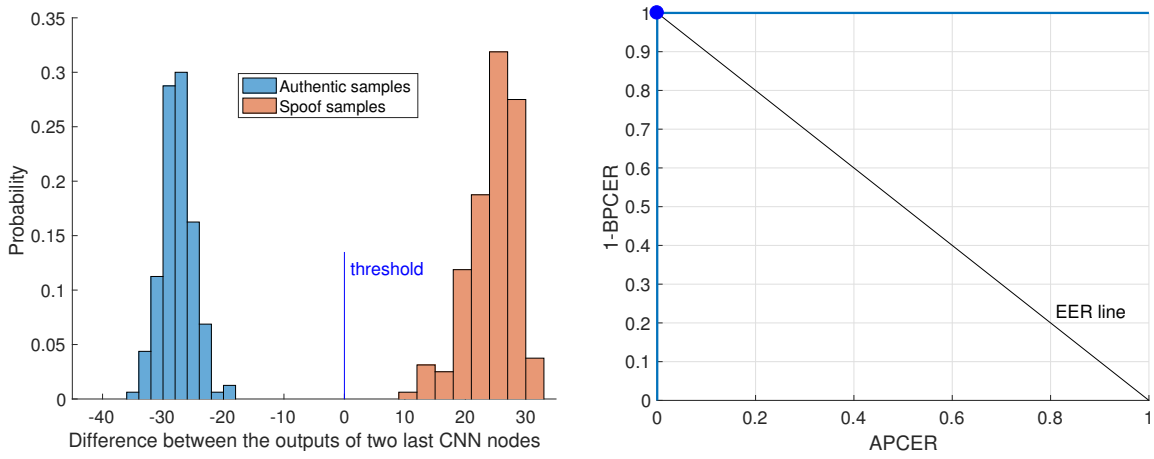

Figure 14: Same as Fig. 13 except the variant: training on **ATVS**, testing on **ATVS**.

separated empirical distributions of the difference between the output nodes of the network obtained for authentic samples and spoofs.

Table 7: Performance results obtained in **cross-dataset** evaluations of **iris PAD** using the overall testing set of each dataset.

Training	Testing	APCER (%)	BPCER (%)	HTER (%)	ROC and ePDF
Warsaw	ATVS	0.0	0.625	0.312	Fig. 15
ATVS	Warsaw	99.9	0.0	49.99	Fig. 16

However, training on ATVS and testing on Warsaw yields almost null abilities to detect spoofs (APCER = 99.9%). This may suggest that exchanging a single layer put on top of the VGG-based feature extraction (trained for a different problem than spoofing detection) is not enough to model various qualities of iris printouts prepared independently by different teams and using different acquisition hardware. Fig. 15 confirms that almost all scores obtained for spoofing samples are on the same side of the threshold as for authentic samples. Certainly, if the threshold can be adapted (which is not typically done in the tests), one can find other proportion between APCER and BPCER, for instance a threshold shifted from 0 to -21.9 results in the EER 13.2%.

6 Conclusions

In this chapter, we proposed a PAD solution for three modalities widely employed for designing biometric systems (*i.e.*, face, iris and fingerprint) based on VGG network archi-

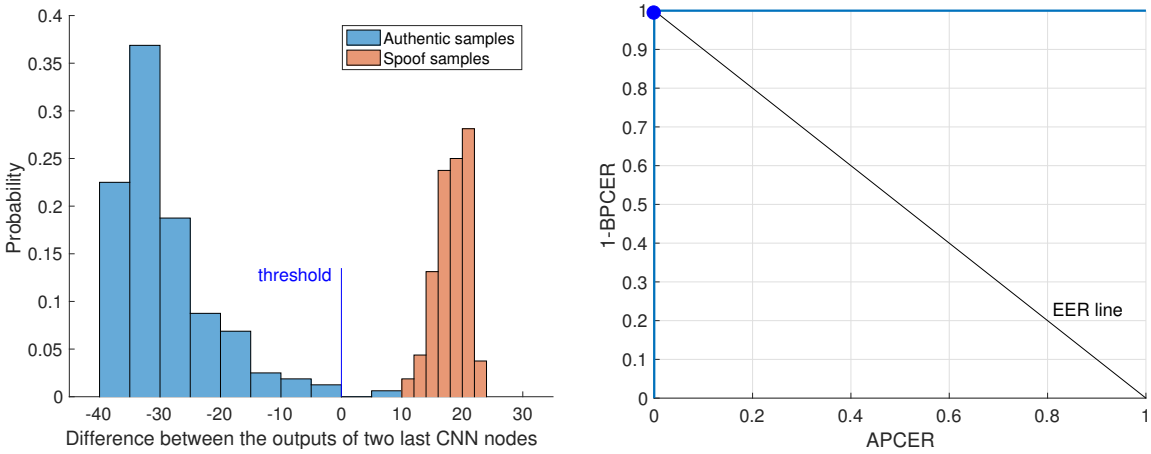

Figure 15: Same as Fig. 13 except the variant: training on **Warsaw LivDet2015**, testing on **ATVS**.

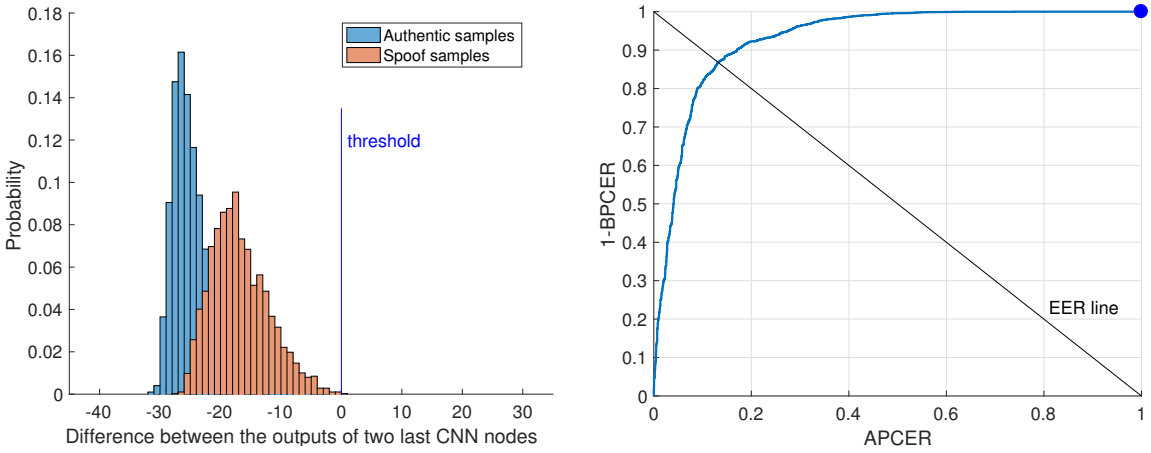

Figure 16: Same as Fig. 13 except the variant: training on **ATVS**, testing on **Warsaw LivDet2015**.

tecture, a deep network architecture originally proposed for object recognition. We showed a methodology to adapt the VGG network to the two-class spoofing classification problem, which was evaluated using six benchmarks available for scientific purposes. The experiments were conducted taking into account the main challenges existing in this research field such as classification across different types of attempted attacks, biometric sensors, and qualities of samples used during attack. In this section, we discuss two main takeaways observed after the analysis presented in this chapter.

The first conclusion is that deep learning is an astonishingly powerful approach to

detect image-based presentation attacks in three considered modalities. Note that the final solution is a subtle modification of the VGG network, trained for a different task, not related to presentation attack detection. In the case of iris and fingerprints, the starting network is not even related to the same object recognition task. The results showed that we can use deep learning to detect spoofing attacks in some cases (AVTS iris benchmark) even perfectly. In this simple approach, we have changed only the last layer, connected strictly to the classification task performed by the VGG network. However, one can consider replacing two or all fully connected layers and utilize the output of the convolutional part of the network more efficiently.

The second takeaway comes from the cross-dataset and cross-sensor experiments. These exceptionally poor results seem to be related to the flexibility that characterizes convolutional networks. The flexibility allows them to "decide" which discovered properties of the input data they use in the classification task. But if they are not trained on data that contains a reasonable sampling of the situations present during testing, then they fail terribly, since most of the features no longer correspond to the new data.

This, however, is not a surprising result and simply calls for solutions that take prior knowledge about the modeled phenomenon into account. Apparently the current fascination with deep learning has brought back an old debate: should we use models that are based on our understanding of the problem, which is neither full nor accurate (called feature engineering or "hand-crafted" solutions) or rather flexible models that learn everything from the data (called feature learning or "data-driven" solutions)? It seems that a reasonable mixture of both approaches should present the best reliability. We firmly believe the solution to this problem is in taking the *best of both worlds*.

Acknowledgment

We thank Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES) through the DeepEyes project, the São Paulo Research Foundation (FAPESP) through the DéjàVu project (Grant #2015/19222-9), and Microsoft Research for the financial support.

References

[1] N. Almoussa. Variational retinex and shadow removal. Technical report, University of California, Department of Mathematics, 2009.

[2] A. Anjos and S. Marcel. Counter-measures to photo attacks in face recognition: A public database and a baseline. In *IEEE Int. Joint Conference on Biometrics (IJCB)*, pages 1–7, Oct. 2011.

[3] C. Arthur. iPhone 5S fingerprint sensor hacked by germany's chaos computer club. `http://tinyurl.com/pkz59rg`, Sept. 2013. Accessed: 2016/02/20.

[4] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research (JMLR)*, 13:281–305, 2012.

[5] S. Bharadwaj, T. Dhamecha, M. Vatsa, and R. Singh. Computationally efficient face spoofing detection with motion magnification. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 105–110, June 2013.

[6] Z. Boulkenafet, J. Komulainen, X. Feng, and A. Hadid. Scale space texture analysis for face anti-spoofing. In *IAPR Int. Conference on Biometrics (ICB)*, pages 1–6, June 2016.

[7] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security (TIFS)*, 11(8):1818–1830, Aug 2016.

[8] E. Carazzai. Paranaguá Harbor employees used silicone fingers to circumnvent biometric system in Paraná. `http://tinyurl.com/hkoj2jg`, Feb. 2014. Accessed: 2016/02/20.

[9] I. Chingovska, A. Anjos, and S. Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *Int. Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–7, Sept. 2012.

[10] A. Czajka. Database of iris printouts and its application: Development of liveness detection method for iris recognition. In *Int. Conference on Methods and Models in Automation and Robotics (ICMMAR)*, pages 28–33, Aug. 2013.

[11] A. Czajka. Pupil dynamics for iris liveness detection. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(4):726–735, Apr. 2015.

[12] A. Czajka. Iris liveness detection by modeling dynamic pupil features. In K. W. Bowyer and M. J. Burge, editors, *Handbook of Iris Recognition*, pages 439–467. Springer London, London, 2016.

[13] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893, June 2005.

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

[15] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2(7):1160–1169, Jul 1985.

[16] J. S. Doyle, K. W. Bowyer, and P. J. Flynn. Variation in accuracy of textured contact lens detection based on sensor and lens pattern. In *IEEE Int. Conference on Biometrics: Theory Applications and Systems (BTAS)*, pages 1–7, Sept 2013.

[17] N. Erdogmus and S. Marcel. Spoofing 2D face recognition systems with 3D masks. In *Int. Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–8, 2013.

[18] N. Erdogmus and S. Marcel. Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect. In *IEEE Int. Conference on Biometrics: Theory Applications and Systems (BTAS)*, pages 1–6, Sept. 2013.

[19] J. Fierrez-Aguilar, J. Ortega-garcia, D. Torre-toledano, and J. Gonzalez-rodriguez. Biosec baseline corpus: A multimodal biometric database. *Pattern Recognition (PR)*, 40:1389–1392, 2007.

[20] J. Galbally, F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia. Fingerprint liveness detection based on quality measures. In *Int. Conference on Biometrics, Identity and Security (BIdS)*, pages 1–8, Sept. 2009.

[21] J. Galbally, F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia. A high performance fingerprint liveness detection method based on quality related features. *Future Generation Computer Systems (FGCS)*, 28(1):311–321, 2012.

[22] J. Galbally, S. Marcel, and J. Fierrez. Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. *IEEE Transactions on Image Processing (TIP)*, 23(2):710–724, Feb. 2014.

[23] J. Galbally, J. Ortiz-Lopez, J. Fierrez, and J. Ortega-Garcia. Iris liveness detection based on quality related features. In *IAPR Int. Conference on Biometrics (ICB)*, pages 271–276, 2012.

[24] D. Garcia and R. de Queiroz. Face-spoofing 2D-detection based on moiré-pattern analysis. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(4):778–786, Apr. 2015.

[25] L. Ghiani, A. Hadid, G. Marcialis, and F. Roli. Fingerprint liveness detection using binarized statistical image features. In *IEEE Int. Conference on Biometrics: Theory Applications and Systems (BTAS)*, pages 1–6, Sept. 2013.

[26] L. Ghiani, G. Marcialis, and F. Roli. Fingerprint liveness detection by local phase quantization. In *Int. Conference on Pattern Recognition (ICPR)*, pages 537–540, Nov. 2012.

[27] L. Ghiani, D. Yambay, V. Mura, S. Tocco, G. Marcialis, F. Roli, and S. Schuckcrs. LivDet 2013 – fingerprint liveness detection competition. In *IAPR Int. Conference on Biometrics (ICB)*, pages 1–6, 2013.

[28] C. Gottschlich. Convolution comparison pattern: An efficient local image descriptor for fingerprint liveness detection. *PLoS ONE*, 11(2):12, Feb. 2016.

[29] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva. Fingerprint liveness detection based on weber local image descriptor. In *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS)*, pages 46–50, Sept. 2013.

[30] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva. An investigation of local descriptors for biometric spoofing detection. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(4):849–863, Apr. 2015.

[31] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva. Local contrast phase descriptor for fingerprint liveness detection. *Pattern Recognition (PR)*, 48(4):1050–1058, 2015.

[32] M. Günther, D. Haufe, and R. Würtz. Face recognition with disparity corrected gabor phase differences. In *Int. Conference on Artificial Neural Networks and Machine Learning (ICANN)*, pages 411–418, 2012.

[33] P. Gupta, S. Behera, M. Vatsa, and R. Singh. On iris spoofing using print attack. In *Int. Conference on Pattern Recognition (ICPR)*, pages 1681–1686, Aug. 2014.

[34] A. Hyvrinen, J. Hurri, and P. Hoyer. *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision.* Springer Publishing Company, Incorporated, 1st edition, 2009.

[35] ISO/IEC. DIS (Draft International Standard) 30107–3, Information technology – Biometric presentation attack detection – Part 3: Testing and reporting, 2016.

[36] A. Jain, A. Ross, and K. Nandakumar. *Introduction to Biometrics*, chapter Introduction, pages 1–49. Springer US, Boston, MA, 2011.

[37] X. Jia, X. Yang, Y. Zang, N. Zhang, R. Dai, J. Tian, and J. Zhao. Multi-scale block local ternary patterns for fingerprints vitality detection. In *IAPR Int. Conference on Biometrics (ICB)*, pages 1–6, June 2013.

[38] W. Kim, S. Suh, and J.-J. Han. Face liveness detection from a single image via diffusion speed model. *IEEE Transactions on Image Processing (TIP)*, 24(8):2456–2465, Aug. 2015.

[39] M. Kohn and M. Clynes. Color Dynamics of the Pupil. *Annals of the New York Academy of Sciences*, 156(2):931–950, 1969.

[40] I. Kokkinos and A. Yuille. Scale invariance without scale selection. In *IEEE Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, June 2008.

[41] N. Kose and J.-L. Dugelay. Reflectance analysis based countermeasure technique to detect face mask attacks. In *Int. Conference on Digital Signal Processing (ICDSP)*, pages 1–6, 2013.

[42] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2012.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[44] S. Kumpituck, D. Li, H. Kunieda, and T. Isshiki. Fingerprint spoof detection using wavelet based local binary pattern. In *International Conference on Graphic and Image Processing (ICGIP)*, volume 10225, pages 102251C–102251C–8, 2017.

[45] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[46] T. S. Lee. Image representation using 2d gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 18(10):959–971, Oct. 1996.

[47] T.-W. Lee, G.-H. Ju, H.-S. Liu, and Y.-S. Wu. Liveness detection using frequency entropy of image sequences. In *IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 2367–2370, 2013.

[48] J. Li, Y. Wang, T. Tan, and A. Jain. Live face detection based on the analysis of fourier spectra. In *Biometric Technology for Human Identification (BTHI)*, volume 5404, pages 296–303. Proc. SPIE, 2004.

[49] F. Lourenço and D. Pires. Video shows Samu's medical using silicone fingers, in Ferraz. `http://tinyurl.com/akzcrgw`, mar. 2013. Accessed: 2016/02/20.

[50] Lovish, A. Nigam, B. Kumar, and P. Gupta. Robust contact lens detection using local phase quantization and binary gabor pattern. In G. Azzopardi and N. Petkov, editors, *Computer Analysis of Images and Patterns (CAIP)*, pages 702–714. Springer International Publishing, 2015.

[51] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision (IJCV)*, 60(2):91–110, 2004.

[52] J. Määttä, A. Hadid, and M. Pietikäinen. Face spoofing detection from single images using micro-texture analysis. In *IEEE Int. Joint Conference on Biometrics (IJCB)*, pages 1–7, Oct. 2011.

[53] J. Maatta, A. Hadid, and M. Pietikäinen. Face spoofing detection from single images using texture and local shape analysis. *IET Biometrics*, 1(1):3–10, Mar. 2012.

[54] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, and A. Majumdar. Detecting silicone mask based presentation attack via deep dictionary learning. *IEEE Transactions on Information Forensics and Security (TIFS)*, PP(99):1–1, 2017.

[55] E. Marasco and C. Sansone. Combining perspiration- and morphology-based static features for fingerprint liveness detection. *Pattern Recognition Letters (PRL)*, 33(9):1148–1156, 2012.

[56] E. Marasco, P. Wild, and B. Cukic. Robust and interoperable fingerprint spoof detection via convolutional neural networks. In *IEEE Symposium on Technologies for Homeland Security (HST)*, pages 1–6, May 2016.

[57] G. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli, and S. Schuckers. *Image Analysis and Processing (IAP)*, chapter First International Fingerprint Liveness Detection Competition – LivDet 2009, pages 12–23. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[58] D. Menotti, G. Chiachia, A. Pinto, W. Schwartz, H. Pedrini, A. Falcao, and A. Rocha. Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(4):864–879, Apr. 2015.

[59] J. Monteiro, A. Sequeira, H. Oliveira, and J. Cardoso. Robust iris localisation in challenging scenarios. In *Communications in Computer and Information Science (CCIS)*. Springer-Verlag, 2004.

[60] R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado. Fingerprint liveness detection using convolutional neural networks. *IEEE Transactions on Information Forensics and Security (TIFS)*, 11(6):1206–1213, June 2016.

[61] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 24(7):971–987, July 2002.

[62] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 24(7):971–987, 2002.

[63] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51 – 59, 1996.

[64] V. Ojansivu and J. Heikkilä. *Image and Signal Processing (ISP)*, chapter Blur Insensitive Texture Classification Using Local Phase Quantization, pages 236–243. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[65] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. Journal of Computer Vision (IJCV)*, 42(3):145–175, 2001.

[66] A. Pacut and A. Czajka. Aliveness detection for iris biometrics. In *IEEE Int. Carnahan Conferences Security Technology (ICCST)*, pages 122–129, Oct. 2006.

[67] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *IEEE Int. Conference on Computer Vision (ICCV)*, pages 1–8, Oct. 2007.

[68] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, number 3 in 1, page 6, 2015.

[69] K. Patel, H. Han, A. Jain, and G. Ott. Live face video vs. spoof face video: Use of moiré; patterns to detect replay video attacks. In *IAPR Int. Conference on Biometrics (ICB)*, pages 98–105, May 2015.

[70] B. Peixoto, C. Michelassi, and A. Rocha. Face liveness detection under bad illumination conditions. In *IEEE Int. Conference on Image Processing (ICIP)*, pages 3557–3560, Sept. 2011.

[71] T. Pereira, A. Anjos, J. de Martino, and S. Marcel. Can face anti-spoofing countermeasures work in a real world scenario? In *IAPR Int. Conference on Biometrics (ICB)*, pages 1–8, 2013.

[72] T. Pereira, J. Komulainen, A. Anjos, J. de Martino, A. Hadid, M. Pietikäinen, and S. Marcel. Face liveness detection using dynamic texture. *EURASIP Journal on Image and Video Processing (JIVP)*, 2014(1):2, 2014.

[73] Q. T. Phan, D. T. Dang-Nguyen, G. Boato, and F. G. B. D. Natale. Face spoofing detection using ldp-top. In *IEEE Int. Conference on Image Processing (ICIP)*, pages 404–408, Sept 2016.

[74] A. Pinto, H. Pedrini, W. Schwartz, and A. Rocha. Video-based face spoofing detection through visual rhythm analysis. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 221–228, Aug. 2012.

[75] A. Pinto, H. Pedrini, W. Schwartz, and A. Rocha. Face spoofing detection through visual codebooks of spectral temporal cubes. *IEEE Transactions on Image Processing (TIP)*, 24(12):4726–4740, Dec. 2015.

[76] A. Pinto, W. Schwartz, H. Pedrini, and A. Rocha. Using visual rhythms for detecting video-based facial spoof attacks. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(5):1025–1038, May 2015.

[77] N. Pinto, D. Doukhan, J. DiCarlo, and D. Cox. A high-throughput screening approach to discovering good forms of biologically-inspired visual representation. *PLoS ONE*, 5(11):e1000579, 2009.

[78] M.-Z. Poh, D. J. McDuff, and R. W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express*, 18(10):10762–10774, May 2010.

[79] P. Pudil, J. Novovičová, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters (PRL)*, 15(11):1119–1125, 1994.

[80] R. Raghavendra and C. Busch. Robust scheme for iris presentation attack detection using multiscale binarized statistical image features. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(4):703–715, Apr. 2015.

[81] K. Raja, R. Raghavendra, and C. Busch. Video presentation attack detection in visible spectrum iris recognition using magnified phase information. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(10):2048–2056, Oct. 2015.

[82] A. Rattani and A. Ross. Automatic adaptation of fingerprint liveness detector to new spoof materials. In *IEEE Int. Joint Conference on Biometrics (IJCB)*, pages 1–8, Sept 2014.

[83] A. Rattani, W. Scheirer, and A. Ross. Open set fingerprint spoof detection across novel fabrication materials. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(11):2447–2460, Nov. 2015.

[84] D. M. Reporter. Face off: Man arrested after boarding plane as an old man - only to land as youthful refugee. `http://tinyurl.com/33l9laz`, Nov. 2010. Accessed: 2016/02/20.

[85] D. M. Reporter. The white robber who carried out six raids disguised as a black man (and very nearly got away with it). `http://tinyurl.com/2cvuq59`, Dec. 2010. Accessed: 2016/02/20.

[86] M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based feature space. In *IEEE Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 699–704, June 2003.

[87] V. Ruiz-Albacete, P. Tome-Gonzalez, F. Alonso-Fernandez, J. Galbally, J. Fierrez, and J. Ortega-Garcia. Direct attacks using fake images in iris verification. In *First European Workshop on Biometrics and Identity Management (BioID)*, volume 5372 of *Lecture Notes in Computer Science*, pages 181–190. Springer, 2008.

[88] W. Scheirer, A. Rocha, A. Sapkota, and T. Boult. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 35(7):1757–1772, July 2013.

[89] S. Schuckers, K. Bowyer, A.C., and D. Yambay. LivDet 2013 - liveness detection iris competition, 2013.

[90] W. Schwartz, A. Rocha, and H. Pedrini. Face spoofing detection through partial least squares and low-level descriptors. In *IEEE Int. Joint Conference on Biometrics (IJCB)*, pages 1–8, Oct. 2011.

[91] A. Sequeira, J. Murari, and J. Cardoso. Iris liveness detection methods in mobile applications. In *Int. Conference on Computer Vision Theory and Applications (VISAPP)*, volume 3, pages 22–33, Jan. 2014.

[92] A. Sequeira, J. Murari, and J. Cardoso. Iris liveness detection methods in the mobile biometrics scenario. In *Int. Joint Conference on Neural Network (IJCNN)*, pages 3002–3008, July 2014.

[93] A. Sequeira, H. Oliveira, J. Monteiro, J. Monteiro, and J. Cardoso. MobILive 2014 - mobile iris liveness detection competition. In *IEEE Int. Joint Conference on Biometrics (IJCB)*, pages 1–6, Sept. 2014.

[94] P. Silva, E. Luz, R. Baeta, H. Pedrini, A. X. Falcao, and D. Menotti. An approach to iris contact lens detection based on deep image representations. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 157–164. IEEE, 2015.

[95] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *arXiv technical report*, 2014.

[96] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[97] J. Sivic and A. Zisserman. Video Google: a text retrieval approach to object matching in videos. In *IEEE Int. Conference on Computer Vision (ICCV)*, pages 1470–1477, Oct. 2003.

[98] Z. Sun, H. Zhang, T. Tan, and J. Wang. Iris image classification based on hierarchical visual codebook. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 36(6):1120–1133, 2014.

[99] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *IEEE Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, June 2015.

[100] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *IEEE Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1701–1708, June 2014.

[101] B. Tan and S. Schuckers. Spoofing protection for fingerprint scanner by fusing ridge signal and valley noise. *Pattern Recognition (PR)*, 43(8):2845 – 2857, 2010.

[102] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *European Conference on Computer Vision (ECCV)*, pages 504–517, 2010.

[103] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing (TIP)*, 19(6):1635–1650, June 2010.

[104] S. Tariyal, A. Majumdar, R. Singh, and M. Vatsa. Deep dictionary learning. *IEEE Access*, 4:10096–10109, 2016.

[105] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(5):815–830, May 2010.

[106] R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, and F. Roli. Fusion of multiple clues for photo-attack detection in face recognition systems. In *IEEE Int. Joint Conference on Biometrics (IJCB)*, pages 1–6, Oct. 2011.

[107] Z. Wei, X. Qiu, Z. Sun, and T. Tan. Counterfeit iris detection based on texture analysis. In *Int. Conference on Pattern Recognition (ICPR)*, pages 1–4, 2008.

[108] D. Wen, H. Han, and A. Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(4):746–761, Apr. 2015.

[109] C. Xu, Y. Zheng, and Z. Wang. Eye states detection by boosting local binary pattern histogram features. In *IEEE Int. Conference on Image Processing (ICIP)*, pages 1480–1483, Oct. 2008.

[110] D. Yadav, N. Kohli, J. Doyle, R. Singh, M. Vatsa, and K. Bowyer. Unraveling the effect of textured contact lenses on iris recognition. *IEEE Transactions on Information Forensics and Security (TIFS)*, 9(5):851–862, 2014.

[111] J. Yang, Z. Lei, D. Yi, and S. Li. Person-specific face antispoofing with subject domain adaptation. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(4):797–809, Apr. 2015.

[112] L. Zhang, Z. Zhou, and H. Li. Binary gabor pattern: An efficient and robust descriptor for texture classification. In *IEEE Int. Conference on Image Processing (ICIP)*, pages 81–84, Sept. 2012.

[113] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Li. A face antispoofing database with diverse attacks. In *IAPR Int. Conference on Biometrics (ICB)*, pages 26–31, Apr. 2012.