Multi-Modal Fingerprint Presentation Attack Detection: Evaluation On A New Dataset

Leonidas Spinoulas*, Hengameh Mirzaalian*, Mohamed Hussein, Member, IEEE, and Wael AbdAlmageed, Member, IEEE

Abstract—Fingerprint presentation attack detection is becoming an increasingly challenging problem due to the continuous advancement of attack preparation techniques, which generate realistic-looking fake fingerprint presentations. In this work, rather than relying on legacy fingerprint images, which are widely used in the community, we study the usefulness of multiple recently introduced sensing modalities. Our study covers front-illumination imaging using short-wave-infrared, near-infrared, and laser illumination; and back-illumination imaging using near-infrared light. Toward studying the effectiveness of each of these unconventional sensing modalities and their fusion for liveness detection, we conducted a comprehensive analysis using a fully convolutional deep neural network framework. Our evaluation compares different combination of the new sensing modalities to legacy data from one of our collections as well as the public LivDet2015 dataset, showing the superiority of the new sensing modalities in most cases. It also covers the cases of known and unknown attacks and the cases of intra-dataset and inter-dataset evaluations. Our results indicate that the power of our approach stems from the nature of the captured data rather than the employed classification framework, which justifies the extra cost for hardware-based (or hybrid) solutions. We plan to publicly release one of our dataset collections.

Index Terms—Fingerprint Biometrics, Presentation Attack Detection, Multi-Spectral Imaging, Short-Wave Infrared, Convolutional Neural Networks.

1 INTRODUCTION

Biometric authentication systems provide additional security and convenience as well as reduced cost, compared to conventional authentication methods. As a result, their use is widespread in different application domains, including law-enforcement or border and access control, for government, corporate or personal purposes. Nevertheless, such systems can be vulnerable to different types of attacks targeting different points of the underlying authentication pipeline. Arguably, the most vulnerable component of a biometric authentication system is the biometric sensor itself, due to the public accessibility of sensors, in many cases. An attack on a biometric sensor typically constitutes the presentation of a fake sample in order to either (1) impersonate a legitimate user or (2) conceal the true identity of a black-listed one. Automatic detection of this type of presentation attack (PA) has attracted significant research interest with a myriad of presentation attack detection (PAD) methods applied on different biometric modalities, such as fingerprint, iris or face [1]. However, due to the continuous advent of realistic presentation attack instruments (PAIs), PAD is still an increasingly challenging problem.

Fingerprint is perhaps the first modality to be used for biometric authentication, and hence, has been thoroughly studied by the biometrics and computer vision communities [2]. Despite its wide acceptability as a universal, distinctive, and permanent biometric characteristic [3], presentation attacks have been shown to successfully spoof fingerprint authentication systems [4], [5]. As a result, significant research work has been devoted to address the problem of fingerprint presentation attack detection (FPAD) [6], [7].

FPAD methods can be categorized into software-only or hybrid, based on the components they add to the biometric authentication system. Software-only or software-based techniques, which are the most abundant in the literature, only add a software module to augment existing fingerprint authentication systems with PAD functionality. Hence, they solely depend on the data used for enrollment and recognition to perform FPAD. Examples of software-based techniques include [8], [9], [10], [11], [12], [13], [14]. On the contrary, hybrid or hardware-based techniques employ additional hardware for FPAD along with the hardware used for fingerprint sensing. We refer to them as hybrid techniques since they still involve software modules that process the data captured by the additional sensing hardware to deliver FPAD functionality. Examples of hybrid techniques include [15], [16], [17].

Software-only FPAD techniques have attracted more interest in the research community owing to their cost-effectiveness and direct applicability on publicly available datasets. Nonetheless, we argue that hybrid techniques should gain more attention for the following reasons:

- Authentication and PAD are two fundamentally different problems. Hence, restricting them to rely on the same sensing hardware limits the progress that could be attained in each.
- With the continuous evolution of sophisticated PA techniques and the attackers’ deeper understanding of the intrinsics of biometric authentication, it is
becoming increasingly challenging to rely on a single sensing technology for both authentication and PAD. In fact, it has been shown that each of the major fingerprint sensing technologies (e.g., optical, capacitive or thermal) is vulnerable to at least one type of material used for PAIs (e.g., gelatin, silicone or glycerin) [18], [19], [20], [21].

- Not only attack fabrication technology improves but also the effect of attacks is becoming more devastating, especially when a single successful attack can be used to impersonate multiple individuals at the same time [22], [23]. Consequently, especially for security critical applications, the additional hardware cost of hybrid approaches is well justified.

- Hybrid PAD techniques can enhance the reliability of a biometric authentication system. Consider the system illustrated in Fig. 1 which employs a hybrid PAD subsystem with a parallel matching branch (similar to the one introduced in [24]). In this design, the separation between the matching and PAD pipelines enhances security since an attacker would have to simultaneously compromise both processing pipelines to succeed. Such task is more challenging compared to the single-point attacks typically found in software-only PAD approaches [25].

In this paper, we focus on hybrid techniques and evaluate a number of recently introduced sensing modalities for FPAD against legacy data from fingerprint authentication systems. Specifically, we investigate the performance of a novel fully-convolutional neural network (FCN) model for FPAD on images captured under different illumination conditions, namely:

1) Visible (VIS) and near-infrared (NIR), denoted together as $F_M$.
2) Short-wave-infrared (SWIR), denoted as $F_S$.
3) Laser speckle contrast imaging (LSCI), denoted as $F_L$.
4) Near-infrared back-illumination, denoted as $B_N$.

In this notation, the letter $F$ stands for front-illumination-based sensing, in which the illumination source and the camera are on the same side with respect to the finger; and the letter $B$ stands for back-illumination-based sensing, in which the illumination source and the camera are on two opposite sides of the finger. The two types of illumination are illustrated in Fig. 2. The subscripts in the aforementioned notation refer to the type of illumination used. To assess the value of hybrid techniques, the performance of our FPAD model using these modalities – individually or in different combinations with one another – is compared to the performance of the same model on legacy fingerprint images, used in software-only techniques. We will refer to the unconventional sensing modalities data as prototype data.

The performance evaluation in this paper is conducted over a new large dataset, named Presentation Attack Detection from Information Sciences Institute (PADISI). PADISI includes data from three different biometric modalities: fingerprint, face, and iris that were collected using the developed system in [26]. In this paper, we only study the fingerprint portion, which we will refer to as PADISI-Finger. PADISI was collected at two different sites. The first site is the University Park Campus of the University of Southern California (USC), in Los Angeles, California while the second is in Columbia, Maryland in a facility of the Applied Physics Laboratory (APL) of John Hopkins University. The data at the two sites were collected using two different replicas of the system in [26]. Additionally, the APL collection included legacy fingerprint data collected via a number of commercial fingerprint sensors. Most of the analysis in this paper is done on the USC data. This dataset will become publicly available upon the acceptance of this manuscript. The APL collection is used for inter-site evaluation and for comparison with legacy data. The public release of the APL data is beyond the control of the authors in this work.

The work presented in this paper builds on top of our prior work [27], [28], with the following additional notable contributions:

- This is the first study that analyzes the PAD capabilities of unconventional sensing modalities for fingerprint compared to legacy data. Our results highlight the power of the studied sensing modalities in tackling different types of attacks.
- The SWIR and LSCI data in the new datasets have significantly higher resolution ($\sim 192$ ppi – pixels per inch) compared to the data used in our prior work [27], [28]. The resolution in the old data was only $\sim 35$ for SWIR and $\sim 135$ for LSCI.
- Our evaluation covers each unconventional modality individually as well as their combinations and compares all that to the performance on legacy data. Such thorough analysis provides valuable information to practitioners on the complementary nature among
groups of sensing modalities and their power in comparison to legacy data.

- We employ a novel FCN model, which is more efficient and more powerful than our prior patch-based convolutional neural network (CNN) models [27], [28]. To validate the effectiveness of this model, we present its evaluation on the LivDet2015 dataset [29] achieving superior performance to state-of-the-art.
- Our evaluation protocols cover a wide spectrum of scenarios. We use 3-fold cross-validation on the USC collection to assess the power of different sensing modalities. Additionally, we follow an inter-collection evaluation (training on USC and testing on APL data) to assess the effect of changes in demographic and PAI distributions. Finally, we follow a leave-one-attack-category out cross-validation to assess the performance on completely unseen attacks.
- The data collected at USC will be publicly released upon the acceptance of this manuscript, which constitutes a valuable asset to the research community. To our knowledge, this will be the first public dataset covering such a broad range of sensing modalities for FPAD. The dataset is relatively large and covers a wide range of PAI species.

1.1 Related Work
Aside from alteration detection, in which the focus is on detecting a physical alteration to a real finger for the purpose of hiding real identity, FPAD methods typically attempt to find characteristics in the input presentation that can distinguish a live finger from a dead or fake one [7]. This explains why these methods are collectively referred to as liveness detection [7].

Numerous proposed approaches are based on physiological characteristics of the finger that are directly sensed as part of the biometric authentication system. These characteristics can be either static or dynamic. Static characteristics, such as odor [16], skin resistance [30], [31], perspiration [32], and internal finger structure measured by optical coherence tomography (OCT) [17], [33], [34], [35], are typically extracted from a single image of the fingerprint. Other methods exploit multiple images to extract static characteristics, such as techniques based on multi-spectral imaging (MSI) [36], [37], [38], [39], [40], [41], [42] or multi-view imaging [43]. On the other hand, dynamic characteristics are derived by nature when processing multiple fingerprint images, e.g., a time series of images to measure finger distortion and elasticity [13], [44], heartbeat [14] or blood flow [45], [46].

Most of the aforementioned approaches are hybrid techniques since they involve additional hardware components for detecting certain physiological characteristics. However, the distinction between bona-fides and attacks may not only rely on the lack of liveness-related physiological attributes but also on the existence of artifacts on a fake fingerprint, stemming from its fabrication process. Consequently, most software-only techniques attempt to differentiate between bona-fide and attack samples based on pattern recognition and signal processing techniques without making the distinction between liveness detection or artifact detection.

Traditionally, software-only FPAD methods apply conventional classification techniques (e.g., support vector machines - SVMs) using hand-crafted features such as wavelet and gray level co-occurrence matrix of optical images [8], [47], [48], [49], histogram and binarized statistical image features [50], [51], [52], scale-invariant feature descriptors [53], Weber local descriptor [54], frequency domain information [55], and pore location distribution [56].

More recently, many software-only FPAD approaches have been proposed utilizing CNNs [57], [58], [59], [60]. Nogueira et al. [58] fine-tuned AlexNet [61] and VGG [62] architectures to perform liveness detection of fingerprints. A classical CNN consisting of four 2D convolutional layers with a binary cross-entropy loss was used by Wang et al. [63], Bhau et al. [64] used triplet loss in their network to minimize the intra-class distances of the patches belonging to the same class while maximizing the inter-class distances. Chugh et al. [65], [66], [67] used MobileNet-v1 over the centered and aligned patches extracted around fingerprint minutiae to discriminate between fake and real fingerprints of optical images. Park et al. [68] included fire and gram modules within their network to learn the textures of bona-fide and PA samples. Kim et al. [69] employed deep belief networks and used contrastive divergence for FPAD.

There exist few CNN-based hybrid techniques for FPAD. For instance, recently, CNN models were used with LSCI [27], [28], [70] and SWIR [36], [42] imaging. MSI-based hybrid techniques are particularly relevant to our research. MSI can reveal distinguishing characteristics of real human skin compared to a multitude of materials used to create fingerprint PAs. In the simplest form of MSI, multiple images of a given target are captured while the target is actively illuminated by different wavelengths in each frame. Each wavelength exhibits varying penetration, absorption and reflection properties for different materials. These phenomena can be utilized in distinguishing real human skin from other materials. Row et al. first introduced MSI to fingerprint image acquisition in [38] but all wavelengths were in the visible spectrum; specifically 430nm-630nm and white. Response to visible spectrum illumination significantly varies between different skin tones, which limits its utility for FPAD. Very limited applications of MSI in FPAD beyond the visible domain exist, such as the work on the visible and near infrared spectrum (i.e., 400nm-850nm) [71], and on the visible to SWIR regimes (400nm-1650nm), whose dynamic characteristics were investigated as a means of distinguishing live fingers from cadavers [40]. In both [71] and [40], very few samples were used. In this paper, we present the most comprehensive study of MSI on FPAD. Not only, our study covers wavelengths in the broad range VIS-SWIR, but also includes front, back, LED-based and laser-based illumination. Furthermore, we use relatively large datasets containing a wide variety of attacks.

2 Finger Biometrics System Design
The utilized finger biometrics sensor suite has been improved and simplified compared to its previous version in [27]. In this section, we provide a brief overview of the system focusing on the rationale for using the selected sensing modalities, supported by relevant literature. The
interested reader can refer to [26] for more technical details about the design, selected hardware components and its operation. The schematic diagram of the system is shown in Fig. 3. A finger is placed on the slit and is illuminated from the bottom side (front-illumination) by multi-spectral LEDs and a laser as well as the top side (back-illumination) by NIR LEDs while being observed by two cameras from a distance of \( \sim 35 \text{ cm} \). Each camera is sensitive to the VIS/NIR or the SWIR spectrum, respectively. The sensor is touch-less in the sense that no platen covers the finger slit.

Each chosen sensing modality in our station provides indicative information for either the liveness of a presentation or the presence of a fabrication artifact, as follows:

- **Front VIS/NIR illumination** \( (F_{M}) \) data: NIR multispectral images can provide ample information about the spectral and textural characteristics of the material of a presentation [71]. Further, they can be used as legacy compatible data, as shown in [26].

- **Front SWIR illumination** \( (F_{S}) \) data: Human skin has a distinctive response in the SWIR spectrum that is independent of skin tone. In a recent study, conducted by the National Institute of Standards and Technology (NIST), the variability of skin response due to differences among people was found to be less significant than the variability due to instrument characteristics only beyond the 1100 nm wavelength [75]. The same observation was confirmed in a successful application of PAD using MSI in the SWIR domain, for face biometrics [76], and more recently, using the previous version of our data, on fingerprint biometrics [27], [36], [42], [77], [78].

- **Front LSCI** \( (F_{L}) \) data: When laser light illuminates a surface, a random interference pattern, known as the speckle pattern, is formed. The pattern is affected by the roughness and/or temperature of the surface and appears static for stationary objects. However, the pattern changes over time when there is motion on the illuminated object, such as the movement of blood cells under the skin surface. In fact, blood perfusion in a tissue can be visualized [79] using LSCI by collecting a sequence of images. Therefore, LSCI measurements can constitute a useful liveness signal for FPAD. LSCI has attracted very little attention in the FPAD literature despite its interesting properties. Chatterjee et al. [80] conducted a study on the hardware and physics of LSCI and showed that there is significant difference between the biospeckle patterns of bona-fide and fake fingers. However, the study involved a very small dataset and did not actually evaluate FPAD performance. Using data from the previous version of our system [27], Keilbach et al. [50], [70] performed LSCI-based FPAD by applying classical classification algorithms (e.g., SVMs), on a set of hand crafted features, such as intensity histograms or LBP features. Later on, Hussein et al. [27] and Mirzaalian et al. [28] utilized various spatio-temporal neural networks to perform LSCI-based FPAD which demonstrated promising performance.

- **Back NIR illumination** \( (B_{N}) \) data: The internal vascular pattern of a hand can be visualized using NIR illumination [71], [81]. In the presence of a finger between a light source and a camera, the collected image intensities represent the amount of light penetrating through the sample, as compared to front-illumination which measures the reflected light, instead [81]. Vessel structures appear darker since light is absorbed by the hemoglobin in the blood and can be used as an indicator of liveness for FPAD. Back-illuminated NIR images from our previous dataset were used for FPAD in [82].

It is worth noting that select combinations of the sensing modalities were already studied on the previous version of our dataset, as in [27], [53], [78], [83]. In this study, we evaluate FPAD performance on our new dataset using a comprehensive list of experiments.

A summary of the captured data for a single finger is presented in Table 1. In multiple cases, data is also captured without active illumination and such images can be used as reference frames for ambient illumination conditions. We developed two replicas of our sensor suite each one used at each collection site (USC and APL). Note that, despite

| Camera | Sensing Modality | Illumination | \( F_{M} \) | \( B_{N} \) | \( F_{S} \) | \( F_{L} \) |
|--------|------------------|--------------|-------------|-------------|-------------|-------------|
| VIS/NIR | Front | white | 406 nm | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SWIR | Front | 4591 nm | 720 nm | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SWIR | Back | 780 nm | 940 nm | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SWIR | Back | 1200 nm | 940 nm | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SWIR | Back | 1300 nm | 1000 nm | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

![Fig. 3. Finger biometric sensor suite. See [26] for more technical details.](image-url)
Fig. 4. Demographic information for the collected datasets. A summary of the collected samples per dataset can be found in Table 3.

Fig. 5. Images of select PAIs used in the PADISI-USC data collection. More details for each PAI can be found in Table 2.

3 Dataset

As discussed in Section 1, the PADISI dataset was collected at two different sites (USC and APL). The data at USC were collected in two separate sessions, referred to as USC-1 and USC-2, between which minor adjustments were applied on the system. During all collections, a participant presented 4 fingers from each hand (excluding little fingers). Each participant passed once or twice by the collection station either in the absence of any attack or in the presence of up to three attached PAIs to the fingers of one or both hands, while a few participants participated in multiple collection sessions. At the APL data collection, data was also collected from all participants using a series of commercial legacy sensors (see [26] for details). In this work, for comparison purposes, we chose to use data from the Optical-C sensor, since it has the highest resolution and provided the most reliable data in terms of match rates (see [26]).

The collected data was thoroughly reviewed by the research team and samples with defects, e.g., due to finger motion or hardware failure, were excluded. A summary of the collected data counts, after these revisions, is provided in Table 3 while the relevant demographic information is presented in Fig. 4. As observed, the APL dataset is larger but exhibits a big imbalance between bona-fide and PAI samples. At the same time, the USC dataset contains a much larger variety of PAI species. Moreover, there is a huge discrepancy in demographics especially in terms of age and race. The USC dataset contains mostly young people of Asian origin (since it was collected in a university environment) while the APL dataset exhibits a skewed distribution.

best efforts, some system adjustments could not be identical in the two replicas, such as lens shutter adjustments. Also, despite the closed-box design, ambient light might leak into the station and partially influence imaging conditions.
TABLE 2
PAI counts in the collected datasets. For each PAI code, we provide a general PAI description, the number of different species per dataset as well as the attributes used for grouping PAI codes in terms of material, species, transparency, and attack type. PAI categories whose appearance depends heavily on the participant and preparation method are marked with *.

A summary of the collected samples per dataset can be found in Table 3. Sponsor approval is required to release additional information about each PAI code.

| Description                         | Code | Total Samples | Total Species | Grouping Attributes |
|--------------------------------------|------|---------------|---------------|--------------------|
|                                      |      | USC  | APL  | USC | APL |
| Play-doh finger                      | 01   | 116  | -    | 4   | -   |
| Silly putty finger                   | 02   | 55   | -    | 3   | -   |
| Dental material finger               | 03   | 51   | -    | 1   | -   |
| Wax finger                           | 04   | 74   | -    | 1   | -   |
|                                      | 05   | 180  | -    | 5   | -   |
|                                      | 06   | 76   | 50   | 1   | 1   |
| Gummy material finger                | 07   | 78   | -    | 2   | -   |
|                                      | 08   | 550  | -    | 6   | -   |
|                                      | 09   | 59   | -    | 2   | -   |
| Gummy material finger with coating   | 10   | 195  | -    | 5   | -   |
|                                      | 11   | 69   | -    | 1   | -   |
| 2D printed fingerprint               | 12   | 49   | -    | 1   | -   |
|                                      | 13   | 64   | -    | 1   | -   |
|                                      | 14   | 22   | -    | 1   | -   |
|                                      | 15   | 37   | -    | 1   | -   |
| Gummy material overlay               | 16   | 265  | 282  | 7   | 2   |
|                                      | 17   | 18   | -    | 1   | -   |
|                                      | 18   | 164  | -    | 3   | -   |
|                                      | 19   | 90   | -    | 1   | -   |
|                                      | 20   | 77   | -    | 3   | -   |
| Gummy material overlay               | 21   | 61   | 314  | 1   | 4   |
|                                      | 22   | 21   | 86   | 1   | 1   |
| 3D printed finger                    | 23   | -    | 77   | -   | 1   |
| Conductive overlay                   | 24   | 48   | -    | 2   | -   |
|                                      | 25   | 24   | -    | 1   | -   |
| Staples/Band-aid                     | 26   | 72   | -    | 1   | -   |
|                                      | 27   | -    | 172  | -   | 1   |
|                                      | 28   | 98   | 100  | 1   | 1   |
|                                      | 29   | 2    | 4    | 1   | 1   |

TABLE 3
Collected datasets summary. Demographic information can be found in Fig. 4 while more details about PAIs are provided in Table 2.

| PADISI Datasets | USC | APL | APL-LEGACY |
|-----------------|-----|-----|------------|
| Participants    | 355 | 672 | 665        |
| Unique fingers  | 2490| 5371| 5308       |
| Total samples   | 6221| 11444| 8850      |
| Bona fide samples | 3596| 10359| 8043     |
| PAI Samples     | 2615| 1085| 807        |
| PAI Species     | 58  | 12  | 12         |

in Table 3 is due to the fact that participants did not always visit the legacy sensors multiple times due to their slow speed as well as the data revision process, which was independently performed per sensor. In this table, we present only the samples from the legacy sensor that correspond to participants and fingers available in our prototype data.

Snapshots of prepared and collected PAIs are illustrated in Fig. 5. Analytic information about all PAIs used during the data collections are provided in Table 2 where PAIs are categorized using different codes, each of which can contain one or more species (e.g., referring to a different color of a specific material or a slightly different preparation method). For each PAI code, we have provided a classification scheme using a set of grouping attributes based on their material, species, transparency, and attack type. Some of these at-
Fig. 6. RGB visualization of samples from the collected datasets for all types of captured data. For each image, the corresponding dark channel has been subtracted and each RGB channel has been normalized by dividing by its maximum value for visualization purposes, albeit introducing visible color artifacts in some cases. Note that, the resolution of the $F_S$ data is the same as the $F_M$ data but is presented smaller in the illustration.

3.1 Mean Intensity Analysis

Based on the appearance of the images in Fig. 6, one might argue that the average intensities of each spectral channel of the captured data could be sufficient for obtaining a reasonable FPAD performance, since the images of bona-fides and PAIs look drastically different. In order to understand the characteristics of the captured multi-spectral data in more detail, we conduct an analysis of the average intensity of each channel using $t$-SNE visualizations [84], where we use all available $F_M$ and $F_S$ data, 10 frames from
the \( F_L \) data and 3 frames from the \( B_N \) data (see Table 1). The visualizations are presented in Fig. 7 where samples are re-colored to distinguish bona-fides from specific PAI codes, bona-fides and PAIs in general, bona-fides and PAIs from each data collection session as well as bona-fides and PAIs with the grouping attributes introduced in Table 2. From the presented 2D distributions, we can make the following observations:

- PAI codes tend to form individual clusters, supporting the use of multi-spectral data for observing the distinctive response of different PAI materials.
- Bona-fide samples of the \textit{USC-1} and \textit{USC-2} data collections appear to have a large separation mainly due to the adjustments in our system (as discussed earlier). At the same time, bona-fide samples between the \textit{USC} and \textit{APL} data collections also exhibit significant separation possibly because of their vastly different demographics (as shown in Fig. 4).
- It is apparent that a large number of PAIs becomes separable from bona-fides simply using the average intensity features, strengthening the power of multi-spectral data for FPAD. However, the analysis demonstrates that certain PAI types might be particularly hard to detect, the majority of which are transparent overlays made of multiple materials.
The presented visualization can also provide an estimate of how challenging a dataset is and could be used for assigning weights to PAI codes, based on their distance from the bona-fide cluster.

4 FULLY-CONVOLUTIONAL NEURAL NETWORK MODEL

Deep neural networks (DNNs) have repeatedly delivered breakthroughs in multiple research disciplines. However, they are notoriously known for being data-hungry: substantial amounts of data have to be used to avoid overfitting during training. Despite the relatively large size of our dataset, it is still small for effectively training a DNN model. This problem has been addressed in prior work on our data in two different ways. In [27], [28], patch-based models were employed by extracting hundreds of $8 \times 8$ patches from an input sample and classifying it by taking the average score over all patches. This approach proved very effective but has two main drawbacks. First, it involves repeated computations when processing consecutive patches, which typically overlap. Second, it does not scale well as the input image size increases, in which case sparse sampling of patches becomes necessary for the approach to be feasible. The alternative approach of transfer learning was used in [42], [53], [78], where a pre-trained network is fine-tuned on our data instead of trained from scratch. The main issue with transfer learning is relying on existing models, which might be too complex for the task at hand.

In this paper, we introduce a novel model for FPAD that avoids the drawbacks of both the patch-based and transfer learning-based approaches. Our new model uses an FCN structure that maps a whole region of interest (ROI) of the input finger image to a map of classification scores, which is then averaged to produce the overall classification score of the input sample. The details of the network structure are shown in Fig. 8. The network consists of five convolutional blocks, with a batch-normalization layer after each group of two-blocks. Starting from a specified number of convolutional maps in the first block ($h$ in Fig. 8), the number of convolutional maps is doubled in the transition between each two consecutive blocks. The final block is followed by a convolutional layer that produces a single map using a sigmoid activation, which represents the patch-wise classification score. Note that, there are no explicitly cropped patches in the FCN model. The patch is an implicit concept referring to the receptive field of each score map element. The score map is then fed to a global average pooling (GAP) layer to produce the final score. Each score map entry is in the range $[0, 1]$, enforced by the sigmoid activation. Hence, the final model output also falls in $[0, 1]$.

The resulting PAD score can be interpreted as the probability of the presence of a PA. Hence, the ground truth score is set to 1 for PA samples and 0 for bona-fide samples. The loss function comprises of two components, one for the final classification score $L_{GAP}$, and one for the patch classification score $L_{patch}$, as:

$$L = L_{GAP} + \lambda L_{patch} ,$$

where $\lambda > 0$. Let the score map associated with an input image $x$ be $M : R^C \times W \times H \rightarrow R^{W_m \times H_m}$, where $C$ is the number of channels in $x$, $W \times H$ is the input 2D image size, and $W_m \times H_m$ is the score map size. Considering the binary cross entropy (BCE) as the loss function, the two loss components can be expressed as:

$$L_{GAP}(x) = BCE \left( \frac{1}{W_mH_m} \sum_i \sum_j M_{ij}(x), t \right) ,$$

$$L_{patch}(x) = \frac{1}{W_mH_m} \sum_i \sum_j BCE (M_{ij}(x), t) ,$$

where $t \in \{0, 1\}$ is the ground truth label of the input $x$.

For all experiments presented in this work, we use $h = 16$ and $\lambda = 10$, while $C$ is varying depending on the provided input image channels. Based on these parameters, the effective patch size (or receptive field) of the model is an area of $54 \times 54$ pixels for each score in $M$. Training is performed for a maximum of 100 epochs using a batch size of 16 and the Adam optimizer [85]. The initial learning rate is set to $2 \times 10^{-4}$ with a minimum learning rate of $1 \times 10^{-7}$ following a reduce-on-plateau by 0.5 strategy (with patience 10 epochs and threshold $1 \times 10^{-4}$) by monitoring the validation loss, when a validation set is available.

Fig. 8. Proposed FCN network architecture: Given parameter $h$, an input image of $C$ channels is first converted into a two-dimensional score map (corresponding to a PAD score per patch) from which the final PAD score is extracted by global average pooling.
5 EXPERIMENTAL EVALUATION

This section presents the experimental evaluation of this work. We first study the capabilities of the presented FCN model on an existing dataset and then run a comprehensive list of experiments on the presented dataset. The analysis uses various evaluation protocols on different combinations of input channels from the collected sensing modalities in our dataset and compares their performance with legacy data, whenever appropriate. Before proceeding with the evaluation, we summarize the data pre-processing steps and evaluation metrics used in our analysis.

Data Pre-Processing: The data, whether prototype or legacy, are first pre-processed:

- Prototype Data: From the captured data, summarized in Table 1, we use all frames from the \( F_M \) and \( F_S \) sensing modalities, frames 10 – 19 from the \( F_L \) data and frames 10 – 12 from the \( B_N \) data, in different combinations. If non-illuminated frames are available for any spectral channel, the time-averaged non-illuminated frame is first subtracted. The data is then normalized in \([0, 1]\) using the corresponding bit depth (see Table 1). Because of the fixed relative location of the finger slit in images captured by our equipment, it was possible to set fixed ROIs for each sensing modality. The ROI was chosen to cover most of the area of the top finger knuckle which almost always consists of skin in bona-fide samples, and PAI material in PA samples. ROIs of images of different sensing modalities are all resized to 160×80 pixels using bicubic interpolation and then stacked to form the multi-spectral data cube provided to our model.

- Legacy Data: Legacy finger images consist of a single grayscale image where the finger ridges appear dark. We detect the finger area by first binarizing the image using thresholding, applying dilation with a circular structuring element of diameter 7 pixels and finding the centroid of the largest connected component in the resulting binary image. The centroid is used as the center for extracting an ROI of pre-defined fixed size, for all images, depending on the average finger coverage in the images of each legacy sensor.

Evaluation Metrics: As discussed in Section 4, the output score of the FCN model in Fig. 8 represents the PA probability (in \([0, 1]\)) of an input sample. Therefore, using a threshold of 0.5 is a natural choice for obtaining the FPAD binary classification and, hence computing the accuracy. However, any choice of threshold without a specific target operating point is indeed arbitrary. Hence, we opt to use metrics that do not depend on a pre-set threshold. In particular, three metrics are used to compare different models in our experiments: i) area under the receiver operating characteristic (ROC) curve, denoted as AUC; ii) true positive rate at 0.2% false positive rate, denoted as TPR0.2%, which is the primary evaluation metric used in IARPA’s Odin program [86], through which this research has been sponsored; and iii) bona-fide presentation classification error rate (BPCER) at attack presentation classification error (APCER) of 5%, denoted as BPCER20 in the ISO standard [87].

5.1 Evaluation on LivDet dataset

We first evaluate the proposed FCN model on the LivDet2015 dataset [29]. The dataset contains images from 4 legacy sensors, namely, CrossMatch, GreenBit, Hi-Scan and Digital-Persona from which, following the pre-processing steps for legacy data, we extracted ROIs of size 320 × 256,
We now evaluate the performance of the FCN model on the PADISI-USC dataset. Our analysis uses the pre-processed data cubes and, following an early-fusion approach, considers different stacked combinations of $F_M$, $F_S$, $F_L$, and $B_N$ data as input channels to our model under a range of evaluation protocols. Hence for each experiment in each evaluation protocol, we vary the number of channels $C$, as presented in Fig. 8, to consider all possible 15 combinations of the captured sensing modalities as input data to our model.

### 5.2 Evaluation on PADISI dataset

We now evaluate the performance of the FCN model on the PADISI dataset. Our analysis uses the pre-processed data cubes and, following an early-fusion approach, considers different stacked combinations of $F_M$, $F_S$, $F_L$, and $B_N$ data as input channels to our model under a range of evaluation protocols. Hence for each experiment in each evaluation protocol, we vary the number of channels $C$, as presented in Fig. 8, to consider all possible 15 combinations of the captured sensing modalities as input data to our model.

#### 5.2.1 Evaluation Protocols

We employ three different evaluation protocols. In the first strategy, we use a 3-fold (3FOLD) partitioning to alleviate the bias resulting from a fixed division of the dataset into training, testing, and validation sets. In this strategy, samples of a collection are divided into three roughly-equal sets of samples such that data of each participant only appears in a single set. Then, 3FOLD cross-validation evaluation is done by performing three experiments, each time using two sets for training and validation, and the left-out set for testing. Each time, from the training/validation group, 80% of data is used for training and 20% for validation, while making sure all training/testing/validation sets are participant-disjoint. It is also important to note that in all splits, the distribution of bona-fide and different PAI category samples is approximately balanced among the training/testing/validation sets. The 3FOLD partitioning is con-
considered for intra-collection evaluations, which are applied either to the PADISI-USC or the PADISI-APL collections.

As a second partitioning strategy, we use a leave-one-attack-out (LOO) protocol, in order to evaluate the ability to detect unknown attacks (i.e., attacks not present in the training data). Using the PAI grouping attributes of Table 2, we create a partition for each color (excluding the special case), leading to 11 partitions for the PADISI-USC collection.

As mentioned in Sections 2 and 3, the sensor parameters, environmental settings as well as the demographics for PADISI were not identical in the two collection sites (USC and APL). To study the effect of these variations to the FPAD classification performance, we also perform inter-site evaluation for which PADISI-USC is used for training and validation, and PADISI-APL is used for testing.

FPAD classification performance is evaluated, per fold, for each of the settings and for all aforementioned metrics. The mean and standard deviation of each metric is computed over the total number of the folds for each partitioning strategy. In the following sections, we provide an analysis of our results under the different scenarios.

5.2.2 Evaluation Results

- 3FOLD evaluation: The 3FOLD evaluation results for the PADISI-USC dataset are presented in Fig. 10 and for the PADISI-APL in Fig. 11. For the PADISI-APL dataset, a comparison to the performance using the data of the selected legacy sensor is presented using two different pre-processing methods. Legacy-S refers to ROIs of $100 \times 80$ pixels, equal in size to our prototype data, while Legacy-L refers to larger ROIs of size $480 \times 384$ following the general pre-processing steps for legacy data described earlier. Even though the total number of samples for the APL-LEGACY data is not equal to our prototype data; the folds were consistent by using the same participant’s samples as in the 3FOLD partitions described in Section 5.2.1.

- PADISI-USC LOO evaluation: The average performance on all 11 LOO partitions of the PADISI-USC dataset is presented in Fig. 12 while analytic results for each partition are summarized in Table 5.

- Inter-site evaluation: The inter-site evaluation results are depicted in Fig. 13.
5.2.3 Discussion

Based on the results of Section 5.2.2 we can make the following observations:

- Some of the studied sensing modalities (particularly \( F_M \) and \( F_S \)) can achieve very high FPAD performance with \( F_S \) being less affected by cross-dataset variations (as seen in Fig. 11), which is consistent with the relevant literature on SWIR imaging.

- With the exception of \( F_L \) or \( B_N \) data alone, any other single sensing modality or combination proves superior to legacy data (see Fig. 11), especially at very low false positive rates.

- In the majority of the cases, combining more than
one sensing modalities leads to improved performance. This is particularly apparent in the unknown attack scenario where most models employing two or more sensing modalities consistently outperform the single ones for most metrics (see Table 5). This observation supports the use of multi-spectral data for FPAD and the power of hybrid methods.

- In multiple cases, increasing the number of channels provided to the model does not always lead to performance improvement, albeit not with significant loss. Such behavior may signify overfitting or non-optimal weighting of certain channels toward the final classification output and could open new research directions for incorporating channel attention techniques to our model [88], [89], [90].

Extending the t-SNE visualization analysis of Section 3, we compare the separation in t-SNE space between the features extracted by the FCN model (score maps) to the corresponding mean-intensity features, for select experiments, in Fig. 14. The illustrations include results for 1 fold of the 3FOLD evaluation protocol on the PADISI-USC collection and the inter-site evaluation protocol. The figure presents the two worst results on the left, the medial result in the middle and the best two results on the right for each evaluation protocol by using the equal error rate (EER) threshold to calculate the accuracy of each experiment. The visualization marks the misclassified bona-fide and PAI samples based on the EER threshold and shows the corresponding locations of these samples in the mean-intensity feature visualization. Finally, the codes of the misclassified PAs are presented in the legends. This analysis results in the following observations:

- The addition of sensing modalities leads to performance improvement, in most cases (compare right-most to leftmost input channels), consistent with the ROC curve observations.
- The mean-intensity feature visualization indeed provides an estimate of how challenging a dataset is. The majority of misclassified PAs by the FCN model are in most cases within the bona-fide cluster of the mean-intensity t-SNE visualization. This dictates that intensities of multi-spectral data could be playing an important role as classification features in the model.
- The majority of misclassified bona-fides lie at the border of the bona-fide cluster in the mean-intensity feature visualization, which further supports the aforementioned argument (this is mostly visible in the inter-site experiment).
- The misclassified PAI codes for the rightmost highest accuracy experiments agree with the analysis of Fig. 7, which demonstrated that the most challenging PAs are the transparent overlays (see Table 2).
- The left-most experiments also demonstrate the ability of deep features to achieve a reasonable classification performance even when large amounts of intensity features are intermingled for the two classes.
- Finally, the inter-site analysis clearly shows a shift in the EER threshold resulting from the vastly different demographics and collection system variations for the two datasets. This is consistent with the mean-intensity t-SNE separation of bona-fide samples from different datasets in Fig. 7.

6 Conclusions
This paper presented a comprehensive analysis of hybrid (hardware-based) FPAD, using a number of recently introduced sensing modalities, which are front-illuminated visible, NIR, and SWIR images; back-illuminated NIR images, and laser-speckle contrast imaging. The analysis was conducted on a new dataset, named PADISI-Finger. PADISI-Finger consists of two collections performed at two different sites, PADISI-USC and PADISI-APL, the former of which will be publicly released upon the acceptance of this manuscript. PADISI-Finger contains data from over a thousand participants and covers more than 60 PAI species. Our analysis employed a novel fully-convolutional network model, whose power was demonstrated by showing its state-of-the-art performance on the LivDet2015 dataset. FPAD performance using this FCN model revealed the advantages of some of the individual unconventional sensing modalities and all different combinations of them over legacy data. The power of combining multiple sensing modalities was further confirmed by the results of our rigorous evaluation protocols, which assessed the effects of testing on completely unseen attack categories (leave-one-attack out) as well as under different dataset characteristics (inter-collection). In such challenging protocols, front-illuminated multi-spectral images (visible, NIR, and SWIR) stood out as the most reliable sensing modalities, either individually or in combination with others. Low-dimensional data visualization of the raw average intensity values revealed a notable similarity to the equivalent visualization of the features learned by the FCN model, which upholds the role of the data in the obtained FPAD performance in our analysis. Finally, it was observed that the FPAD performance is not directly correlated with the number of employed sensing modalities. This is believed to be an artifact of the neural network model architecture, and its investigation is deferred to our future work.

Acknowledgments
This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2017-1702020005. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

The authors would also like to acknowledge the efforts of the APL team for organizing and managing the large scale data collection at their site in Columbia, Maryland, and the effort of our collaborators from Hochschule Darmstadt University of Applied Sciences for their help and support in collecting fingerprint data at the APL site. A final acknowledgement is due to Mian Zhong and Hanchen Xie for their efforts towards a very early draft of this manuscript.
Fig. 14. t-SNE visualization comparison for deep features (score maps) of the FCN model in Fig. 8 and the corresponding mean-intensity features of the utilized input channels. Each experiment presents the classification accuracy based on the EER threshold, the misclassified samples of the FCN model in both visualizations and the codes of misclassified PAIs in the legend. Experiments are sorted by their accuracy from left to right. The reader is referred to Section 5.2.3 for a detailed analysis.

REFERENCES

[1] S. Marcel, M. S. Nixon, J. Fierrez, and N. W. D. Evans, Eds., Handbook of Biometric Anti-Spoofing - Presentation Attack Detection, Second Edition, ser. Advances in Computer Vision and Pattern Recognition. Springer, 2019. [Online]. Available: https://doi.org/10.1007/978-3-319-92627-8

[2] A. K. Jain, K. Nandakumar, and A. Ross, “50 years of biometric research: Accomplishments, challenges, and opportunities,” Pattern Recognition Letters, vol. 79, pp. 80 –105, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865515004365

[3] A. Ross and A. Jain, “Human recognition using biometrics: An overview,” Annales des Telecommunications, vol. 62, no. 1-2, pp. 11–35, 1 2007.

[4] J. Galbally, J. Fierrez, and J. Ortega-Garcia, “Vulnerabilities in biometric systems: attacks and recent advances in liveness detection,” in Spanish Workshop on Biometrics, SWB, 2007.

[5] E. Bowden-Peters, R. C. W. Phan, J. N. Whitley, and D. J. Parish, Fooling a Liveness-Detecting Capacitive Fingerprint Scanner. Springer Berlin Heidelberg, 2012, pp. 484–490. [Online]. Available: https://doi.org/10.1007/978-3-642-28368-0_32

[6] E. Marasco and A. Ross, “A survey on antspoofing schemes for fingerprint recognition systems,” ACM Comput. Surv., vol. 47, no. 2, pp. 28:1–28:36, Nov. 2014. [Online]. Available: http://doi.acm.org/10.1145/2617756

[7] C. Sousedik and C. Busch, “Presentation attack detection methods for fingerprint recognition systems: a survey,” IET Biometrics, vol. 3, no. 4, pp. 219–233, 2014. [Online]. Available: https://doi.org/10.1049/iet-bmt.2013.0020

[8] S. B. Nikam and S. Agarwal, “Fingerprint liveness detection using curvelet energy and co-occurrence signatures,” in 2008 Fifth International Conference on Computer Graphics, Imaging and Visualisation, Aug 2008, pp. 217–222.

[9] C. Zaghetto, M. Mendelson, A. Zaghetto, and F. d. B. Vidal, “Liveness detection on touchless fingerprint devices using texture descriptors and artificial neural networks,” in 2017 IEEE International Conference on Biometrics (IJCB), Oct 2017, pp. 406–412.

[10] B. Tan and S. Schuckers, “Spoofing protection for fingerprint scanner by fusing ridge signal and valley noise,” Pattern Recognition, vol. 43, no. 8, pp. 2845 –2857, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320310000993

[11] S. Memon, N. Manivannan, and W. Balachandran, “Active pore detection for liveness in fingerprint identification system,” in 2011
Leonidas Spinoulas
Leonidas Spinoulas received his Diploma degree in Electrical and Computer Engineering from the National Technical University of Athens, Greece in 2010. In September 2010 he joined Northwestern University, Evanston, IL, USA and the Image and Video Processing Laboratory (IVPL) under the supervision of Prof. Aggelos K. Katsaggelos. He received the M. Sc. Degree in Electrical Engineering and Computer Science in 2012 and the Ph.D. degree from the same department in August 2016. Since 2017, he holds a Research Computer Scientist position with the Information Sciences Institute (University of Southern California), Marina del Rey, CA. He was previously a Research Scientist for Ricoh Innovations Corporation, Cupertino, CA, USA. He was the recipient of the best paper awards at EUSIPCO 2013 and SENSORCOMM 2015 and has 4 patents. His primary research interests include deep learning, biometrics, multispectral imaging, image processing, image restoration, inverse problems and compressive sensing.

Hengameh Mirzaalian
Hengameh Mirzaalian conducts research on image-processing, machine-learning, and deep-learning techniques to address different tasks on images with different modalities and applications. She received her Ph.D. in computing science from Simon Fraser University, BC. Prior to ISI, she worked with various research institutes including Siemens Corporate Research and Technologies, Harvard Medical School, Boston Children’s Hospital, Brigham and Womens Hospital and Nasa Jet Propulsion Laboratory. Her research appeared on top journals including Medical Image Analysis, NeuroImage and IEEE conferences.

Mohamed Hussein
Dr. Mohamed E. Hussein is a computer scientist at USC ISI and an associate professor (on leave) at Alexandria University, Egypt. Dr. Hussein obtained his Ph.D. degree in Computer Science from the University of Maryland at College Park, MD, USA in 2009. Then, he spent close to two years as an Adjunct Member Research Staff at Mitsubishi Electric Research Labs, Cambridge, MA, before moving to Alexandria University as a faculty member. Prior to joining ISI, he spent three years at Egypt-Japan University of Science and Technology (E-JUST), in Alexandria, Egypt. Dr. Hussein’s most recent research focus has been in securing biometrics and machine learning systems. He has over 30 published papers, and three issued patents.

Wael AbdAlmageed
Dr. AbdAlmageed is a Research Associate Professor at the Department of Electrical and Computer Engineering, and a research Team Leader and Supervising Computer Scientist with Information Sciences Institute, both being units of USC Viterbi School of Engineering. His research interests include representation learning, debiasing and fair representations, multimedia forensics and visual misinformation (including deepfake and image manipulation detection) and biometrics. Prior to joining ISI, Dr. AbdAlmageed was a research scientist with the University of Maryland at College Park, where he led several research efforts for various NSF, DARPA and IARPA programs. He obtained his Ph.D. with Distinction from the University of New Mexico in 2003 where he was also awarded the Outstanding Graduate Student award. He has two patents and over 70 publications in top computer vision and high performance computing conferences and journals. Dr. AbdAlmageed is the recipient of 2019 USC Information Sciences Institute Achievement Award.