

Towards realistic fingerprint presentation attacks: the ScreenSpooF method

Roberto Casula^a, Marco Micheletto^a, Giulia Orrù^a, Gian Luca Marcialis^a and Fabio Roli^b

^aUniversity of Cagliari, Italy

^bUniversity of Genoa, Italy

ARTICLE INFO

Keywords:
fingerprint
PAD
liveness
in the wild

ABSTRACT

Using mobile devices, people leave latent fingerprints on the screen that can be photographed and used as a mold to fabricate high-quality spoofs. In this paper, we analysed the threat level of realistic attacks using snapshot pictures of latent fingerprints against presentation attack detection systems. Our evaluation compares a semi-consensual acquisition, given by a voluntary pressure of the finger on the screen, with a completely non-consensual one, given by the normal use of the device. ù

1. Introduction

Fingerprint presentation attack detectors (FPADs) are systems able to detect whether a fingerprint image comes from a Presentation Attack Instrument (PAI) [1], namely, an artificial replica of a finger. In the last years, FPADs have shown an increasing accuracy, thanks to deep learning-based approaches especially [4, 8]. However, presentation attacks (PAs) detection is commonly considered an open problem due to its “arms race” nature. Current FPAD are machine learning models trained over pre-collected alive and fake images. Training an FPAD by a sufficiently large set of fingerprint images from PAIs consensually obtained, as commonly done, should lead to an effective detector. Actually, it would be difficult for attackers to obtain such conditions in real scenarios. In other words, semi-consensual or unconsensual PAIs should be relatively worse than those obtained by the consensual approach, and easily blocked, as suggested in [5, 9].

However, previous works showed no significant correlation between image quality score and “liveness” score, i.e., the probability that the image comes from an alive finger given the feature set, as usual outcome of an FPAD [9] viewed as a two-class classifier. To the best of our knowledge, the claim is still to be confirmed that more realistic attacks, leading to low-quality images, are ineffective.

This is the goal of the present paper, which is the follow up of [3], where a novel approach to provide the finger’s mold was presented and called “ScreenSpooF”. It consists of digitally processing the snapshot of the smartphone screen of the targeted client. This picture can retain the latent fingerprints released in daily device use. In [3], we tested fingerprint images provided by this method against the best three FPADs submitted to the 6th edition of the International Fingerprint Liveness Detection Competition (LivDet2019 [11]), which can be considered as good state of the art (SOTA) rep-


resentatives. The ScreenSpooF data set of [3] was captured under strict, constrained conditions: the user’s smartphone screen was always cleaned and the user puts her/his finger on it under our control. Thus, it could not yet be considered a full realistic test and, for this reason, we refer to it as “in vitro” ScreenSpooF data set. This paper provides a novel set of ScreenSpooF-based images completely acquired “in the wild”, since the screenshots were taken from daily-used smartphone screens without our intervention. This replicates the scenario where the attacker must process latent fingerprints, never analyzed in other contributions to the SOTA. The new data is collected fully unsupervised. Moreover, we compared the “in vitro” and the new “in the wild” acquisition methods, testing the attacker ability to deceive an FPAD. For this purpose, we used the same algorithms of [3], namely, the LivDet2019 winners.

Behind this contribution, we carried out a large set of experiments, which allowed us to confirm the effectiveness of the SOTA FPADs against “in the wild” attacks; however, several deviations from this general statement are worthy of being put into the research community’s attention. We propose a possible explanation for those deviations, that represents a further step forward from our previous work [3]. Analyzing the qualities and distributions of gray levels allowed us to hypothesize why “in the wild” PAs made with some materials are more dangerous than others.

The paper is structured as follows. Section 2 reports a brief summary of existing approaches to PAI fabrication based on realistic attacks. Section 3 describes the ScreenSpooF method adapted to “in the wild” scenarios. Section 4 details experimental protocols and results on the proposed data set. Section 5 concludes the paper.

2. Realistic approaches to PAI fabrication

While the consensual method can be part of a coercion scenario for the targeted client or a very insidious cheat that leads the user to press a finger on a plasticine-like material, non-consensual methods rely on developing the latent fingerprint left on a smooth or non-porous surface. The digitized version of the mark can be processed to obtain the mold,

 roberto.casula@unica.it (R. Casula); marco.micheletto@unica.it (M. Micheletto); giulia.orrù@unica.it (G. Orrù); marcialis@unica.it (G.L. Marcialis); roli@unica.it (F. Roli)

ORCID(s): 0000-0003-3810-5935 (R. Casula); 0000-0003-1196-7869 (M. Micheletto); 0000-0002-7802-2483 (G. Orrù); 0000-0002-8719-9643 (G.L. Marcialis); 0000-0003-4103-9190 (F. Roli)

printed on a transparent sheet by a laser printer or photolithographic techniques. Methods of this kind require a high level of precision and image processing capabilities, besides an excellent manual ability and time [9].

In the 2013 LivDet competition [5], two data sets were created by developing latent fingerprints left by volunteers on a paper sheet using magnetic powders. Since the finger mark was left under the laboratory staff's control and the whole fingers were clearly visible, the process was semi-consensual. Nevertheless, a high accuracy rate suggested the ineffectiveness of those semi-consensual PAIs. In particular, the winner of LivDet 2013 achieved on average 95.9% accuracy on samples obtained by the semi-consensual method based on powder and 77.4% on data obtained by the consensual method. FPADs were approximately 20% more accurate in detecting PAs obtained from latent fingerprints. Therefore, that early attempt was not a real threat to biometric systems equipped with FPADs.

Another method for taking latent fingerprints consists of photographing a phone screen, firstly proposed by [6]. They argued that the typical device use leaves on the glass intact or partial fingerprints, which can be detected very well in optimal light conditions. Captured finger marks were processed using PCB techniques. Several attacks against the fingerprint sensor of various smartphones were simulated. An IAPRM of 9.08% was obtained over 17,100 attacks on five devices[1]¹). However, as well-known, the smartphones do not have integrated software PADs, and the user cannot set the acceptance threshold.

Therefore, to the best of our knowledge, the literature lacks a clear reporting of the effectiveness of realistic PAs on systems equipped with presentation attack detection ability.

3. The ScreenSpooF method

The ScreenSpooF method (SS) [3] simplifies the idea presented in [6]. The basic assumption is that the regular smartphone use during the day generates several finger marks on the screen surface, which can be photographed and digitally processed in order to create a mold for a PAI. This is potentially a concrete and more realistic threat since an attacker could obtain the fingerprints from the screen without costly resources and technical skills. A clever attacker could also take a snapshot on a smartphone left temporarily unattended.

Moreover, ScreenSpooF bypasses the expensive techniques, in time and skills especially, based on 3d printing or PCB as done in [6]. To be as realistic as possible, we used the standard 2d laser printing process on a transparent foil, exploiting the 3d differences in height between ridges and valleys introduced by the inkjet on the foil.

The PAI creation process of the ScreenSpooF technique can be outlined in four steps (Fig. 1):

1. Acquisition: a camera is used to snap a picture of the smartphone screen, after identifying an ideal set of candidates for creating the molds. An external light

source can be employed to better differentiate it from the background.

2. Binarization: the image is converted from RGB to gray scale and inverted, then the contrast between valleys and ridges is increased in order to create a very accurate cast for the fake fabrication.
3. Pagination: each fingerprint image is cropped and resized to the real fingerprint dimensions by analyzing proportions with respect to the device in the original photo. Finally, they are printed on a transparent sheet to simulate the distance between ridges and valleys through the thickness between sheet and ink.
4. PAI fabrication: the previous phase provides a set of casts that can be used to create several PAIs by dripping different materials on the sheet as done in the non-consensual approach. Once dried, the fakes can be removed and acquired with the sensors.

Based on this starting point, we collected two data sets based on the degree of realism in PAI fabrication: the first one is called *in vitro* ScreenSpooF, already employed in [3]; the second one is called *in the wild* ScreenSpooF because no real cooperation is asked to the volunteer except for its consent to lend the device for the finger marks development. The term *in the wild* comes from biometric recognition, in particular from face detection, referring to data collected «in unconstrained conditions» [13]².

3.1. *In vitro* ScreenSpooF data collection

The *in vitro* procedure is detailed as follows: after thoroughly cleaning the screen, the smartphone is placed horizontally, and the user is asked to place the left index, middle, and ring fingers in the lower part of the screen, while the same fingers from the right hand are placed above the previous ones. The smartphone is then placed on a vertical support and properly lit to display the fingerprint details better, while a high-resolution photo of the entire screen is taken. The same procedure is repeated to capture both hands' thumb and little finger. We then followed the 2-4 steps, described above.

The *in vitro* approach was the first to be compared to the SOTA consensual methods in terms of PA detection. Overall, reported results [3] showed comparable performances in the two cases, confirming the effectiveness of PAIs realized through snapshot pictures. Moreover, some tested algorithms have experienced a significant drop in performance, clearly proving the actual threat of the new fabrication method. In particular, we showed in [3] that ScreenSpooF-based attacks can be a threat comparable to that where PAIs are fabricated by the consensual method.

Although these notable results, it should be noted that this scenario represented the “worst-case”, i.e. the best chance for an attacker to steal the most apparent latent print of the victim, since he/she collaborates during the first phase. This paper aims to face this aspect, simulating a real and stealthy acquisition of finger marks from the user's smartphone; the details will be discussed in the next Section.

¹Impostor Attack Presentation Match Rate (IAPMR), in some publications also reported as SpooF-False Acceptance Rate or SFAR

²See the LFW data set at: <http://vis-www.cs.umass.edu/lfw/index.html>

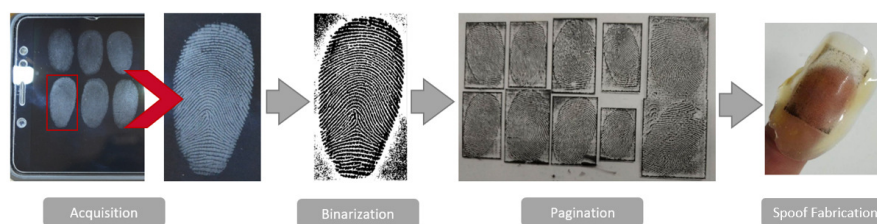


Figure 1: The four steps of the ScreenSpooF method, in accordance with the *in vitro* acquisition protocol: (1) acquisition of the latent fingerprint, (2) pre-process and binarization, (3) resize and pagination and (4) PA creation with different materials.



Figure 2: Differences between the first two steps of ScreenSpooF method using different acquisition protocols: *in vitro* acquisition (a) and binarization (b); *in the wild* acquisition (c) and binarization (d).

3.2. In the wild ScreenSpooF data collection

Figure 2c shows that, in an uncontrolled scenario, the dirty surface of the screen is the leading cause of noisy latent finger marks, mostly partials or overlapped to other impressions. However, it is still possible to find complete fingerprints with sufficient ridge information, representing optimal candidates for realizing the cast and consequently a viable threat. Accordingly, we captured the smartphone screen by simulating the total unawareness of the victim.

We collected a new set of images called *in the wild* ScreenSpooF data set, without imposed restrictions on the fingers' position or screen cleanness. In this scenario, finding clear latent marks can be challenging and time-consuming due to the poor ridge quality and considerable background noise.

By varying the shooting modes, we obtained very different results in terms of ridge clearness, according to the characteristics of the impressions left on the user's smartphone screen and the light conditions in the room. By employing an external light source, such as portable led lights, smartphone flashlight, etc., we further highlighted the finger marks; moreover, this is an optimal solution in cases where the display of latent fingerprints cannot be done satisfactorily by varying the shooting mode only. We subdivided the potential molds in "Full finger marks", that is, finger marks fully deploying a complete mold, and "Partial finger marks", where only a fraction of them was useful to deploy a mold. Some examples are shown in Figure 3. The amount of collected images is reported in Table 2.

In our opinion, the *in the wild* data set is a significant step ahead with respect to what done in [3] and described in Section 3.1. This task is challenging for the attacker because the related skill and ability in designing the best strategy to capture the targeted finger mark dramatically increase. Nevertheless, nothing assures that a good snapshot can lead to



Figure 3: Examples of acquired PAs of ScreenSpooF *in the wild* data set. Green Bit: full (a); partial (b). DigitalPersona: full (c); partial (d).

Table 1

Device characteristics for ScreenSpooF data sets.

Scanner	Model	Res.[dpi]	Img Size	Format	Type
Green Bit	DactyScan84C	500	500x500	BMP	Optical
DigitalPersona	U.are.U 5160	500	252x324	PNG	Optical

an effective PAI. Thus the attack is expensive, as we experienced that 50% of snapshots were useless, due to overlapping or distorted fingerprints. Moreover, the failure probability is potentially high once the PAI is fabricated, as we showed in the next Sections.

4. Experimental results

4.1. Experimental protocol

In order to compare the *in the wild* ScreenSpooF with the "*in vitro*" ScreenSpooF data set, we employed the same materials (Body Double, Mix1 and Mix2) for the PAIs' fabrication as well as the same sensors, namely, the Green Bit DactyScan84C and DigitalPersona U.are.U 5160 (Tab. 1). The total number of fingerprints collected with the new approach is 67, 56 of which are complete and 11 partial. Each of them was acquired ten times, varying rotation, displayed section and pressure exerted on the spooF when placed on the

Table 2

ScreenSpooft *in vitro* and *in the wild* data sets composition. The latter consists of two subsets: one composed of full (F) and one composed of partial (P) latent fingerprints.

Data set	ScreenSpooft <i>in vitro</i> data set				ScreenSpooft <i>in the wild</i> data set							
	Live	BD	Mix1	Mix2	Live (F)	BD (F)	Mix1 (F)	Mix2 (F)	Live (P)	BD (P)	Mix1 (P)	Mix2 (P)
Green Bit	1500	600	600	600	1250	560	560	560	250	110	110	110
DigitalPersona	1500	600	600	600	1250	560	560	560	250	110	110	110

Table 3

APCER of the *in the wild* approach compared to the *in vitro* method, for the three most accurate FPADs of LivDet 2019.

	Green Bit				DigitalPersona			
	<i>in vitro</i>	wild(T)	wild(F)	wild(P)	<i>in vitro</i>	wild(T)	wild(F)	wild(P)
PAD [7]	32.83%	1.04%	1.25%	0.00%	15.06%	0.05%	0.06%	0.00%
ZJUT [14]	0.61%	9.35%	8.15%	15.45%	19.67%	0.00%	0.00%	0.00%
JLW	0.72%	9.45%	8.27%	15.45%	19.67%	0.00%	0.00%	0.00%

sensor. Table 2 reports the amount of useful PAs obtained by *in vitro* and *in the wild* ScreenSpooft. We clearly indicated the amount of PAs derived from full (F) and partial (P) finger marks detected in the smartphone screen, as explained in Section 3.2.

In the experiments, we focused on the Attack Presentation Classification Error Rates (APCER) and Bona fide Presentation Classification Error Rates (BPCER), which are error rates strictly referred to the FPAD performance and False Non-Match Rate (FNMR), False Match Rate (FMR) and Imposter Attack Presentation Match Rate (IAPMR), related to an integrated system [1].

We put ourselves in the same experimental conditions of [3], namely, the three winning FPADs submitted to LivDet 2019 competition were selected as good representatives of SOTA FPADs: 1) *PAD*, a handcrafted system that uses combinations of local and global features [7], 2) *ZJUT*, based on Slim-Res CNN [14] and 3) *JLW*, also based on deep learning. These algorithms had previously been trained with the LivDet 2019 training set, consisting of fingerprint images collected by the consensual method. Therefore, this analysis will also outline the effectiveness of SOTA solutions against cross-method as well as cross-material threats.

4.2. Threat evaluation

Tables 3 and 4 describe the outcomes of each algorithm on all data sets analyzed by reporting the accuracy and the BPCER @1%APCER and APCER @1%BPCER values, respectively. As can be clearly noticed in Table 3, a trend reversal characterizes the *in the wild* method compared to the *in vitro* one: for the Green Bit sensor, the deep learning-based algorithms are more vulnerable than the handcrafted one, while the DigitalPersona sensor keeps a high level of security. The same tendency can be observed in Table 4. To show more precisely the relationship between error rates at different decision thresholds, the detection error tradeoff (DET) curves are shown in Figure ???. In general, our experiments show that an attack performed with an *in the wild* replica has a much lower incidence in terms of fake recognition, in fact, on average, the *in the wild* experiments are characterized by a lower APCER for all operating points.

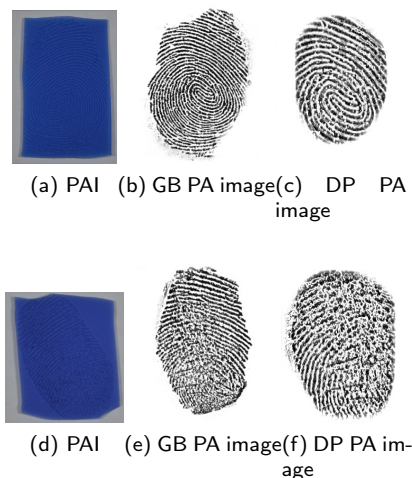


Figure 4: Examples of PAs *in vitro* (first row) and *in the wild* (second row).

To explain these performance differences, we firstly performed a visual analysis of the samples acquired. Figure 4 reports an example of Body Double fakes created using the two ScreenSpooft methods and the relative acquisitions with the two sensors. It is apparent that the *in vitro* (4a) replica has a higher quality, and this feature is also transmitted in the optical sensor acquisitions (4b-4c). The *in the wild* forgery (4d) is characterized by a considerable degree of noise, particularly pronounced in the DigitalPersona sample (4f). As a matter of fact, the screen cleaning in the *in vitro* data set allows a very sharp and detailed photo of the fingerprint; on the contrary, the new approach presents “noisy” images, due mainly to accumulated dirt and continuous overlapping of fingerprints on the screen.

Another aspect worth deepening regards each FPAD’s behavior over different materials used to counterfeit the fingerprints. This allows to figure out the riskiest or most suitable material classes for a specific detector type. Table 5 shows the outcome of this investigation, which can be summarized as follows:

- For the Green Bit sensor, Mix2 PAs fabricated *in the wild* appear to be a threat for the two deep learning-based FPADs, presenting a much higher APCER than those fabricated *in vitro*; the handcrafted-based detector turns out to be more robust with the *in the wild* data set. On the other hand, no significant differences are shown in terms of APCER in the Body Double and Mix1 cases. This suggests that the Mix2 material is responsible for the trend reversal highlighted above.

Table 4
BPCER@1%BPCER and APCER@1%BPCER for the top-three LivDet 2019 winners tested on the ScreenSpooF *in the wild* and *in vitro* data sets.

	Green Bit								DigitalPersona							
	BPCER@1%BPCER				APCER@1%BPCER				BPCER@1%BPCER				APCER@1%BPCER			
	in vitro	wild(T)	wild(F)	wild(P)	in vitro	wild(T)	wild(F)	wild(P)	in vitro	wild(T)	wild(F)	wild(P)	in vitro	wild(T)	wild(F)	wild(P)
PAD	40.80%	2.33%	1.84%	1.60%	48.39%	3.03%	3.15%	3.33%	8.93%	0.13%	0.16%	0.00%	20.61%	0.05%	0.06%	0.00%
ZJUT	0.07%	2.80%	2.40%	6.00%	0.17%	2.84%	2.26%	5.76%	18.07%	0.60%	0.80%	0.40%	27.67%	0.30%	0.36%	0.00%
JLW	0.07%	2.87%	2.24%	6.80%	0.17%	2.99%	2.44%	5.76%	18.07%	0.60%	0.80%	0.40%	27.56%	0.30%	0.36%	0.00%

Table 5
APCER for each individual material composing the *in the wild* and the *in vitro* data sets, for the three most accurate liveness detectors of LivDet 2019 competition.

	Alg.	Green Bit			DigitalPersona				
		in vitro	wild (T)	wild (F)	wild (P)	in vitro	wild (T)	wild (F)	wild (P)
BD	PAD	37.67%	1.64%	1.96%	0.00%	17.50%	0.00%	0.00%	0.00%
	ZJUT	0.67%	1.04%	1.25%	0.00%	19.00%	0.00%	0.00%	0.00%
	JLW	0.00%	1.05%	1.25%	0.00%	25.33%	0.00%	0.00%	0.00%
Mix1	PAD	29.00%	0.60%	0.71%	0.00%	14.83%	0.00%	0.00%	0.00%
	ZJUT	0.00%	0.90%	1.07%	0.00%	15.34%	0.00%	0.00%	0.00%
	JLW	0.00%	1.05%	1.25%	0.00%	25.33%	0.00%	0.00%	0.00%
Mix2	PAD	31.83%	0.90%	1.07%	0.00%	12.83%	0.15%	0.18%	0.00%
	ZJUT	1.17%	26.12%	22.14%	46.36%	14.67%	0.00%	0.00%	0.00%
	JLW	1.17%	16.27%	22.32%	26.36%	14.67%	0.00%	0.00%	0.00%

- For the DigitalPersona sensor, all detectors are unanimous on the ineffectiveness of *in the wild* spoofs, achieving 0% APCER in almost all considered materials, although it was strongly susceptible to *in vitro* attacks. As pointed up in the visual analysis (Figure 4d), this evidence could be due to a certain degree of noise in the spoof, which results in being more marked in images acquired with this sensor.

4.3. Spoof images analysis

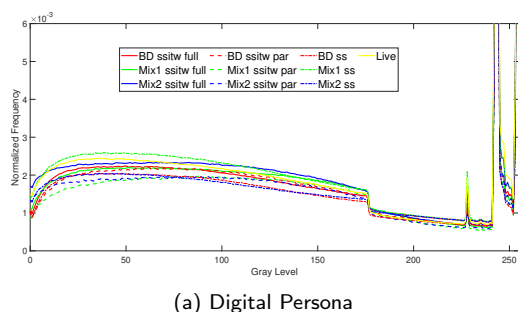
Reported results allowed us to identify the Mix2 material as a severe threat to the security of the Green Bit sensor if equipped with deep learning-based FPADs. Such accuracy decrease can not be explained with only the different FPADs' nature, given the excellent response against the other materials. To explain why PAIs *in the wild* in Mix2 are more difficult to classify for deep-learning based methods than other PAs, we carried out two additional investigations: (1) the quality assessment and (2) the gray level analysis of all the images of the *in vitro* and *in the wild* ScreenSpooF data sets. The goal is to understand if these images are of higher quality than the others or if they have particular characteristics that confuse the classifiers.

For the first analysis, we used the NIST Fingerprint Image Quality (NFIQ) algorithm [12], which assigns a score based on five levels of quality: poor (5), fair (4), good (3), very good (2), and excellent (1). We plot the quality score and the liveness score associated to each image in Figures 8 - 9. These Figs. point out two mainly common aspects: first, live fingerprint samples are generally characterized by high quality, regardless of the considered sensor; second, the quality trend in the PAI case is strongly affected by the acquisition protocol.

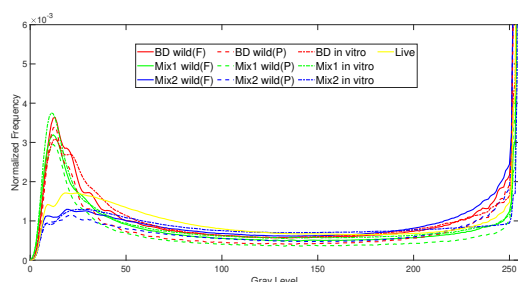
Considering the analysis of the ScreenSpooF *in vitro* data set acquired with the Green Bit sensor reported in Fig. 8,

we can observe that although the PAIs' quality is comparable to that of live samples, the PAs are, overall, correctly recognized by the detectors, achieving a much lower liveness score. This lack of correlation was already observed in [9]. Worth noting, deep learning-based FPADs tend to squeeze scores towards extremes, while the detector based on handcrafted features shows a more uniform and concentrated distribution in a range close to the 50% threshold. As a matter of fact, this behavior is synonymous of worse performance (Table 3). On the other hand, the "*in the wild*" ScreenSpooF data are characterized by low-quality PAs, explicitly confirming the findings of the visual analysis carried out in the previous section. However, despite this, the Mix2 fakes' distribution for the two deep learning-based PADs is equally spread in the liveness score axis; therefore, performance is not related to quality, which does not help us understanding the drop in this specific case. The difference between the *in vitro* and *in the wild* approaches is more evident for the Digital Persona sensor (Figure 9). In the first case, a significant percentage of high-quality PAs presents a liveness score greater than 50, leading to an inadequate level of security, especially for JLW and ZJUT detectors. On the contrary, the PAIs created through the *in the wild* approach generate a well-separated distribution, where the poor quality of the sample implies a low score and, consequently, a significant growth in the accuracy.

The average gray level distribution for all fingerprint images gave us some more insights. The gray level profiles for the Digital Persona sensor, shown in Fig. 5a, are very similar for each type of fake and for the live ones. However, the gray-level profile of Mix2-based PAIs differs significantly from others when the GreenBit sensor is used. This is true for gray values in the range 0-50 and 220-255 especially (Fig. 5b). In other words, the *in vitro* images have more pixels pure white (values around 255) than the *in the wild* ones. These are more characterized by off-white pixels (values between 200 and 254), probably due to the dirty background typical of images of the kind. Therefore, to assess whether this range of gray level profiles affected the FPADs performance, we adapted the gray-level profiles of *in the wild* images to those of *in vitro* images as follows: Experiment 1) we modified the value of the gray levels of 10% of the image pixels chosen randomly within the range (201-254), whatever their position, to the value of 255; that is, if a pixel fell in the range (201-254) we squashed it to 255 with probability to 0.1. (Fig. 6a); Experiment 2) we modified the value of the gray levels of 12.5% of the pixels chosen randomly within the range (201-224) and 20% chosen randomly within the



(a) Digital Persona



(b) Green Bit

Figure 5: Differences between the mean grayscale histograms for the Green Bit (a) and the Digital Persona (b) sensors.



Figure 6: Example of the altered pixels, in green, in the three different experiments: (a) no constraints; (b) inside the ROI; (c) outside the ROI.

range (225-254), internal to the fingerprint, to the value of 255 (Fig. 6b); Experiment 3) we modified the value of the gray levels of 12.5% of the pixels chosen randomly within the range (201-224) and 20% chosen randomly within the range (225-254), external to the fingerprint, to the value of 255 (Fig.6c). In the above experiments, we modified only 10% or 12.5% of pixels for meeting two constraints: matching the *in the wild* profile on the *in vitro* profile the more as possible, and avoiding strong modifications of the images clearly detectable by visual inspection. We repeated each test ten times and reported the results in Table 6. All live samples were also included in this investigation to evaluate the impact of this alteration on the detectors' entire classification potential.

Experiment 1 led to the APCER decrease of deep learning-based algorithms, while leaving the BPCER almost unchanged. This means that the introduced alterations lowered the liveness scores of fake fingerprints previously misclassified, leading them under the 50% threshold. Therefore, the differ-

Table 6

APCER and BPCER analysis before and after pixel alterations introduced by Experiments 1, 2 and 3.

	BPCER			APCER Mix2		
	PAD	ZJUT	JLW	PAD	ZJUT	JLW
Original	2.33%	0.13%	0.13%	0.9%	26.12%	26.27%
Exp.1	2.33%	0.19%	0.19%	0.96%	18.75%	19.13%
Exp.2 (inside ROI)	2.50%	0.19%	0.18%	0.9%	18.82%	19.15%
Exp.3 (outside ROI)	2.09%	0.19%	0.13%	0.81%	23.99%	24.60%

Table 7

APCER of a CNN-based PAD trained both with PAs generated with only the consensual method and including ScreenSpooF *in vitro* samples and tested on ScreenSpooF *in the wild*.

Train data	APCER (%)
GB Consensual	30.26
GB Consensual+ <i>in vitro</i> ScreenSpooF	1.65
DP Consensual	32.58
DP Consensual+ <i>in vitro</i> ScreenSpooF	12.74

ent gray level profiles observed for Mix2 affected the deep learning-based FPAD performance. On the other hand, these changes don't affect the handcrafted method. Off-white pixel distributed over the whole image do not impact on images classification.

Experiments 2 and 3 aim to point out which pixels are more discriminative for deep learning-based FPAD classification, between the external and internal to the fingerprint ROI. In particular, the improved accuracy of *JLW* and *ZJUT* algorithms when tested on samples modified inside the ROI (Experiment 2), highlights the role of these pixels in the classification performance. In our opinion, this alteration is explained by the dirty surface of the smartphone. Therefore, a possible attacker able to know the gray-level profile given by the sensor may further proceed in "soiling" the images to increase the probability of circumvent the PAD algorithm. Since recent results reported in [10] show that specific changes in the gray level profiles may lead to much more effective PAIs, we found a further confirmation of the weaknesses of deep-learning methods, without relying on complex adversarial approaches.

To sum up, the Mix2 PAIs deviation can be related to the different gray level profiles. In particular, the threat that Mix2 PAIs represent for deep learning FPADs may be due to dirt inside the ROI, invisible to the eye, which neural networks associate instead to significant, "alive" features, such as secretions of the eccrine, sebaceous and apocrine glands of the finger skin.

4.4. Match analysis in an integrated system

Once we evaluated the actual threat posed by the two ScreenSpooF procedures, we completed our investigation by assessing the risk of ScreenSpooF *in vitro* PAs on an integrated system since it has proven to be the most critical one. For the sake of space, we present the trade-off between FMR and FNMR and between IAPMR and FNMR for *JLW* and *ZJUT* algorithms (Figure 7). For this evaluation, 6000 genuine comparisons (FNMR), 8700 zero-effort comparisons (FMR) and 9000 PA comparisons (IAPMR) were performed. For the licit scenario, in blue, the horizontal axis represents

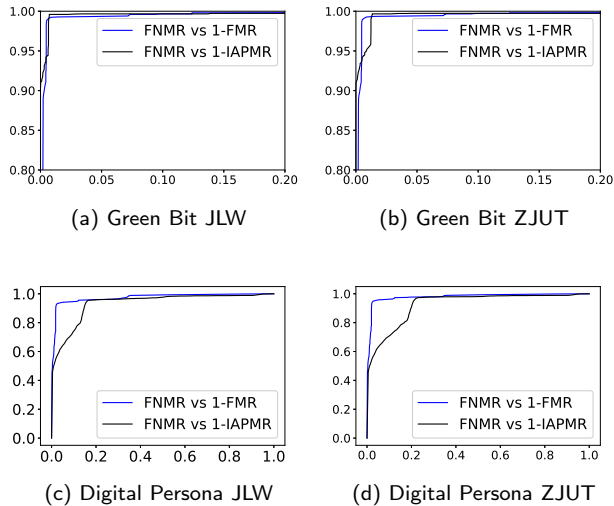


Figure 7: ROC curves for the top-two LivDet 2019 integrated systems tested on ScreenSpooft *in vitro* data set for the Green Bit (a,b) and the DigitalPersona (b,c) sensor.

FMR, while for the PA scenario, in black, it represents IAPMR. From the graphs, it is evident that in all the analyzed scenarios the PAs represent a higher risk than zero-effort attacks. This is particularly apparent for the DigitalPersona sensor, where the fakes' images preserve the information necessary to match the genuine template with a higher probability.

4.5. Results and discussions

The dangerousness of the *in vitro* ScreenSpooft technique for the replication of fingerprints has been also confirmed by the seventh edition of the Fingerprint Liveness Detection Competition [2]. In fact, part of the test set was acquired with this semi-consensual technique, and these data proved to be more difficult to classify than the consensual ones. Since the training set was entirely consensual, the data acquired with the ScreenSpooft *in vitro* technique represent a “never-seen-before” attack and a deterioration in performance was expected. Although these models should provide a good generalization ability, many proposed FPAD solutions exhibit high IAPMR rates on the ScreenSpooft data. This technique not only is able to circumvent the FPAD, but also can preserve information about the person's identity and thus represent a severe risk for system security. This is further confirmed by the match analysis reported in this work. In this scenario, a simple solution is to include, in the PAD's training set, samples generated with the *in vitro* ScreenSpooft protocol since it provides high-quality molds easy to acquire. For this purpose, we carried out further experimental analysis to show the benefit of training with data obtained with multiple acquisition techniques. In particular, we trained a simple CNN (made up of five layers: two convolutional layers alternated with the same number of average-pooling layers and a final dense layer) both with PAs generated only through the consensual method, and subsequently, we have also included ScreenSpooft *in vitro* samples. We then tested

both models on PAs fabricated with the new *in the wild* method. Reported results (Table 7) confirmed our hypothesis since the model trained with additional data discriminates the PAs with more significant accuracy. From these preliminary results, it has been shown that *in vitro* data is representative of *in the wild* data and a designer can protect a PAD system from the *in the wild* attack by collecting *in vitro* data.

5. Conclusions

Anticipating the capabilities of an adversary is crucial in computer security systems. In the design of the FPAD systems, it is therefore necessary to know the threat that new attacks can pose, in order to protect the biometric recognition systems in advance. In this paper, we deepened the analysis on a new technique of replicating latent fingerprints, called ScreenSpooft. It captures the latent fingerprints left on the smartphone screen surface by a simple snapshot.

In particular, snapshots were taken *in the wild*, i.e., better simulating an attack during the everyday use of the device. Accordingly, a novel data sets of PA images were collected. This new technique, completely non-consensual, realistically simulates an attack scenario with latent fingerprints captured from a screen. The LivDet 2019 competition's top three winning algorithms on this new data set has shown that the attack *in the wild* is less effective than the *in vitro* counterpart due to the noise in images. The exceptions required further investigation. This allowed us to examine better the characteristics of some sensors and materials that still constitute, especially in the worst scenario, an open issue for modern PADs. The analysis of the qualities and distributions of gray levels highlighted the vulnerability of deep learning-based methods to certain pixel configurations given to the dirt present in the sensor. This was confirmed by proposed experiments, where the pixel “whitening” inside and outside the fingerprint ROI improved the PA classification.

We believe this paper is a significant step ahead toward the realistic PA detection. In general, current FPADs appeared to be strong enough to face the proposed one. However, the pieces of evidence reported here showed that the problem is not yet solved and only a careful analysis of what the FPADs actually focused on may prevent significant classification errors.

References

- [1] 30107-3, I., 2017. Information technology-biometric presentation attack detection-part 3: Testing and reporting .
- [2] Casula, R., Micheletto, M., Orrù, G., Delussu, R., Concas, S., Panzino, A., Marcialis, G.L., 2021a. Livdet 2021 fingerprint liveness detection competition - into the unknown, in: 2021 IEEE Int. Joint Conf. on Biometrics (IJCB), pp. 1–6. doi:10.1109/IJCB52358.2021.9484399.
- [3] Casula, R., Orrù, G., Angioni, D., Feng, X., Marcialis, G.L., Roli, F., 2021b. Are spoofs from latent fingerprints a real threat for the best state-of-art liveness detectors?, in: 2020 25th Int. Conf. on Pattern Recognition (ICPR), IEEE, pp. 3412–3418.
- [4] Chugh, T., Cao, K., Jain, A.K., 2018. Fingerprint spoof buster: Use of

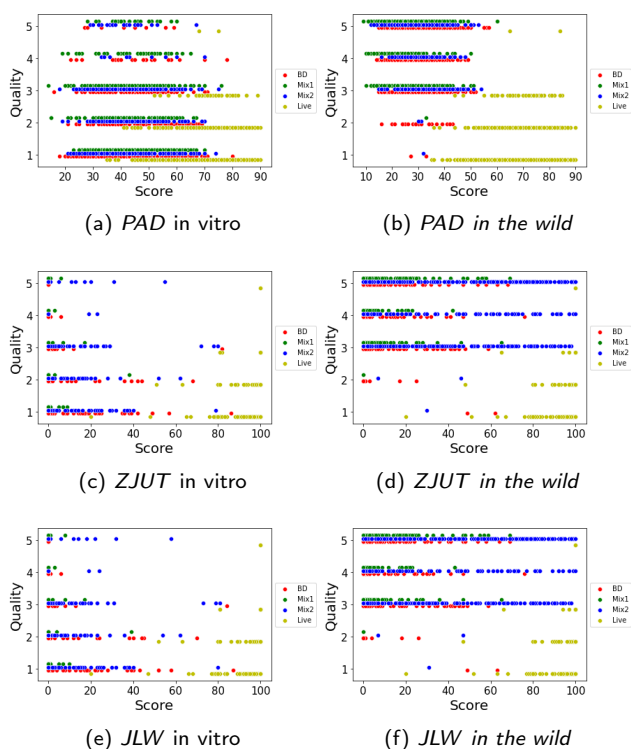


Figure 8: Liveness-Quality correlation for the Green Bit sensor.

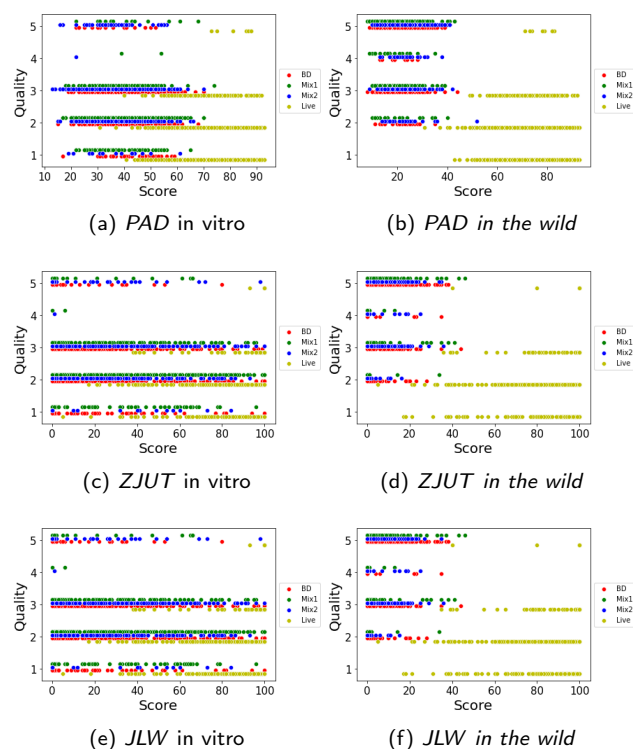


Figure 9: Liveness-Quality correlation for the DigitalPersona sensor.

minutiae-centered patches. *IEEE Transactions on Information Forensics and Security* 13, 2190–2202.

- [5] Ghiani, L., Yambay, D., Mura, V., Tocco, S., Marcialis, G.L., Roli, F., Schuckers, S., 2013. Livdet 2013 fingerprint liveness detection competition 2013, in: 2013 Int. Conf. on Biometrics, IEEE. pp. 1–6.
- [6] Goicoechea-Telleria, I., Garcia-Peral, A., Hussein, A., Sanchez-Reillo, R., 2018. Presentation attack detection evaluation on mobile devices: Simplest approach for capturing and lifting a latent fingerprint, in: 2018 Int. Carnahan Conf. on Security Technology (ICCST), pp. 1–5. doi:10.1109/CCST.2018.8585605.
- [7] González-Soler, L.J., Gomez-Barrero, M., Chang, L., Pérez-Suárez, A., Busch, C., 2021. Fingerprint presentation attack detection based on local features encoding for unknown attacks. *IEEE Access* 9, 5806–5820.
- [8] Liu, F., Liu, H., Zhang, W., Liu, G., Shen, L., 2021. One-class fingerprint presentation attack detection using auto-encoder network. *IEEE Transactions on Image Processing* 30, 2394–2407.
- [9] Marasco, E., Cando, S., Tang, L., Ghiani, L., Marcialis, G.L., 2018. A look at non-cooperative presentation attacks in fingerprint systems, in: 2018 Eighth Int. Conf. on Image Processing Theory, Tools and Applications (IPTA), IEEE. pp. 1–6.
- [10] Marrone, S., Casula, R., Orrù, G., Marcialis, G.L., Sansone, C., 2021. Fingerprint adversarial presentation attack in the physical domain, in: *Pattern Recognition. ICPR Proc., Part VI*, Springer. pp. 530–543.
- [11] Orrù, G., Casula, R., Tuveri, P., Bazzoni, C., Dessalvi, G., Micheletto, M., Ghiani, L., Marcialis, G.L., 2019. Livdet in action-fingerprint liveness detection competition 2019, in: 2019 Int. Conf. on Biometrics (ICB), IEEE. pp. 1–6.
- [12] Tabassi, E., Wilson, C., Watson, C., et al., 2004. Fingerprint image quality.
- [13] Zafeiriou, S., Zhang, C., Zhang, Z., 2015. A survey on face detection in the wild: past, present and future. *Computer Vision and Image Understanding* 138, 1–24.
- [14] Zhang, Y., Shi, D., Zhan, X., Cao, D., Zhu, K., Li, Z., 2019. Slim-rescnn: A deep residual convolutional neural network for fingerprint

liveness detection. *IEEE Access* 7, 91476–91487.