A Novel and Responsible Dataset for Face Presentation Attack Detection on Mobile Devices

Nathan Ramoly[1] nathan.ramoly@idnow.io
Alain Komaty[2] akomaty@idiap.ch
Vedrana Krivokuća Hahn[2] vkrivokuca@idiap.ch
Lara Younes[1] lara.younes@idnow.io
Ahmad-Montaser Awal[1] montaser.awal@idnow.io
Sébastien Marcel[2] marcel@idiap.ch

[1]Research Center of Excellence, IDnow, Rennes, France
[2]Idiap Research Institute, Martigny, Switzerland

Abstract

Presentation Attack Detection (PAD) is essential for ensuring the security of face recognition (FR) systems, particularly in the context of mobile authentication in various sectors, such as online banking and government services. However, current PAD methods are often sensitive to the data domain, partly due to the limitations of training PAD datasets. In this paper, we introduce the SOTERIA dataset, which provides captures of bona-fide and diverse Presentation Attacks (PAs) recorded using smartphones. The dataset was collected ***responsibly*** *from 70 consenting individuals, as opposed to web scraping. It includes face videos, motion data, and depth information (when available) as well as a* ***novel projector-based replay attack****. To demonstrate the utility of the SOTERIA dataset, we evaluate the vulnerability of a SOTA FR model (IResNet100) to the PAs in the dataset. We also analyze the PAD capabilities of a SOTA PAD model (DeepPixBis) through cross-dataset experiments as well as on real attacks observed in an industrial application. Our findings show the effectiveness and versatility of the SOTERIA dataset in advancing PAD research, in particular toward generalization.*

1. Introduction

Relying on the advances of Deep Learning (DL) technologies, Face Verification (FV) has matured and is today widespread in various applications, including mobile authentication or remote identity verification. However, as their popularity grows, FV systems are increasingly targeted by Presentation Attacks (PA) from fraudsters aiming to spoof identities, as illustrated in Figure 1. Consequently, there is a growing need for PAD solutions [20].

PAD has progressed significantly over the past few years due to the advances in DL technologies. This was supported by dedicated PAD datasets captured by the community. While intra-dataset evaluations appear excellent, current PAD systems are under-performing in cross-dataset experiments and thus lacking generalization capabilities, which is key to widespread adoption. This is partly caused by the limits of PAD datasets, which have fewer samples than general-purpose face datasets, and which lack diversity in terms of attacks, devices, environments and identities, leading to bias in trained PAD models and/or over-fitting to a single domain.

(a) Reference (Enroll) (b) Genuine $[-0.175]$ (c) Print $[-0.265]$ (d) Projector $[-0.078]$

Figure 1. Cropped faces and their corresponding negative cosine distances [1, 4] to the reference cropped face, using IResNet100 model.

In this paper, we present the SOTERIA dataset, captured using multiple smartphones, which addresses the aforementioned dataset limitations. It provides high-quality yet noisy bona-fide (BF) and PA data, close to real-world applications, through diverse identities, recording scenarios, Presentation Attack Instruments (PAIs), and Capture Instruments (CIs). Demographic information is also provided, enabling bias evaluation and mitigation. It is composed mainly of color images and videos, but also includes motion (accelerometer and gyroscope) data and depth map for the compatible devices. In total, the dataset includes 8400

BF captures from 70 consenting volunteers, and more than 24000 PAs including a novel (to the best of our knowledge) projector-based replay attack, making it one of the largest and most diverse self-captured datasets in the literature. The remainder of this paper provides an overview of existing PAD datasets, describes the SOTERIA dataset, and presents a vulnerability analysis of a SOTA FV model and an evaluation of a SOTA PAD method on this dataset. For reprodicibility purposes, the source code is made publicly available[1].

2. Related work

Supporting the progress of DL based PAD, more than 40 face PAD datasets have been published since 2010 [20]. As we focus on general public and mobile use cases, we put aside multimodal datasets relying on specialized sensors [20, 13, 9], thus reducing the scope to 26 PAD datasets captured using RGB cameras. We provide an overview of the most recent ones in Table 1.

The lack of generalisation of PAD models is currently the main challenge faced by the PAD community: while intra-dataset performance is usually high, cross-dataset evaluations often show models to under-perform [8, 14, 18, 17, 15]. This can be partly explained by the lack of representativity in PAD datasets due to their acquisition cost [10, 13, 20]. Such weaknesses can manifest in various ways: **(1)** Several datasets, such as OULU-NPU [5], RECOD-MPAD [2] or CRMA [7], only provide a few tens of identities. Such lack of diversity may impact PAD algorithms by rendering the models sensitive to a subject's biometric traits, and potentially biased and unfair. Many PAD datasets are demographically imbalanced [13], yet some of them, such as CASIA-SURF CefA [12], provide multiple sensitive attributes, enabling bias quantification and mitigation. **(2)** The size of PAD datasets tends to be small. OULU-NPU [5] and RECOD-MPAD [2] only include a few thousand samples, well below the size of general purpose face datasets. This is detrimental to achieving robust and generic PAD models [10]. **(3)** PAD models need to be robust to every capture condition, environment, CI, and, importantly, any kind of PA. While datasets may offer diversity for one of these factors, such as PA and devices [21], PAI [16], and recording scenarios [9], almost no dataset provides an extensive and global variability, leading to potential bias and lack of model flexibility across different domains.

Overall, while some datasets try to address one issue, all self-captured datasets are subject to at least one weakness. To alleviate the volume and representativity challenges, researchers have resorted to extracting samples from the web, drastically reducing the cost of acquisition of a large number of samples. Celeba-spoof [21] and Flickr-PAD [16] belong in this category: BF samples are extracted from the web and the effort was put into capturing PAs. WFAS [19] is the only dataset to extract both BF and PA samples from the web. Face images extracted from photos, screens or toys, are labeled as PA. The quality and noise of such an approach is open to discussion, but no vulnerability analysis has been conducted. However, web extracted datasets are subject to legal constraints: as identities were extracted without their consent, such datasets are not compliant with most data protection laws, in particular the EU GDPR, rendering them unusable by part of the PAD community. Alternatively, following the recent breakthrough in generative models, SynthASpoof [8] explored their application to PAD, with promising yet unsatisfactory results.

3. SOTERIA Dataset

The publicly available SOTERIA dataset[2] was constructed with the aim of achieving robust face PAD in a mobile FV context. This was achieved through the acquisition of realistic samples matching the variability of industry captures, which enable trained models to be agnostic to context noise. Hence, our acquisition protocol, described in this section, emphasizes the capture of diverse BF and PA samples by considering a high pool of consenting volunteers, devices, PAIs, and loosely controlled captured scenarios.

3.1. Bona-fide Face Captures

BF face samples were acquired from 70 data subjects using a dedicated application installed on the capturing devices (smartphones). Each volunteer explicitly gave their written consent to have their face data captured and used, and the data collection process was both ethical and legal. Various demographic traits were recorded for the data subjects, to evaluate (in the future) bias in FV and PAD algorithms, including gender, age, skin color (Fitzpatrick scale), and the wearing of glasses. Our dataset boasts an almost perfect gender split (49% females and 51% males), a wide age range of about 20-80 years old (although most subjects were in the 20-30 age range), and skin colours across the whole Fitzpatrick-scale spectrum (with Types II and III being the most common). More information on the demographics is available in [11], which is a related work on personalised hygienic mask PAs.

To add variability to the data captures, the volunteers were asked to attend two capture sessions, which were generally separated by approximately 3 weeks. During a capture session, the data subject was tasked with recording themselves using the dedicated application, and using the front and main cameras of 5 smartphones (Apple iPhone 6s and 12, Xiaomi Redmi 6 Pro and 9A, and Samsung Galaxy S9) as CI in 3 scenarios, with and without hygienic

[1] https://gitlab.idiap.ch/bob/bob.paper.ijcb2024_soteria_database.git

[2] https://www.idiap.ch/dataset/soteria

Dataset	Year	#Id	#BF/#PA	#sbf	pa	#pai	#ci	#spa	form	Dem.	srcbf	gdpr
Replay-Mobile[6]	2016	40	550/640	5	print	1	2	2	vid	no	sc	yes
					replay	1						
OULU-NPU[5]	2017	55	720/2880	3	print	2	6	1	vid	no	sc	no
					replay	2						
RECOD-MPAD [2]	2020	45	450/1800	5	print	1	2	2	vid	yes	sc	yes
					replay	2		1				
Celeba-spoof[21]	2020	10k	156k/469k	-	print	6	>10	8	img	yes	ws	no
					replay	3						
					mask	1						
CASIA-S.CeFA[12]	2021	1607	6300/24k	1	print	1	1	6	vid	yes	sc	no
					replay	1						
					mask	2						
CRMA[7]	2022	47	423/13k	3	print	1	3	1	vid	no	sc	no
					replay	4						
Flickr-PAD[16]	2023	3000	3k/11k	-	print	2+2	2	2	img	no	ws	no
					replay	7+						
WFAS[19]	2023	148k	530k/828k	-	print	8	-	-	img	no	ws	no
					replay	4						
					mask	1						
					other	3						
SynthASpoof[8]	2023	25k	25k/79k	-	print	1	1	1	img	no	ge	yes
					replay	3	1		vid			
SOTERIA	2024	70	8400/24k	3	print	2	**8**	3	vid	yes	sc	yes
					replay	8			img			

Table 1. Overview of main RGB PAD datasets since 2020, and two commonly used mobile-based datasets (Replay-Mobile and OULU-NPU) where: # = number, sbf = BF capture scenario, ci = Capture Instrument, spa = PA capture scenario, form = Data format, Dem = Demographic data, srcbf = Source of BF samples (sc = Self Captured, ws = Web scraped, ge = Generated), gdpr = consensual and GDPR compliant.

masks, for a total of 60 captures. Three lighting scenarios were defined to ensure variability in the captures: indoor, indoor low, and outdoor lateral. These scenarios encompass an ideal/normal capture scenario, but also under- and over-lit environments, challenging the CIs and adding noise to the captures (back-light and glare were nevertheless avoided). The acquired dataset also includes a large variety of backgrounds, particularly for outdoor captures. Additionally, to emulate the facial motion challenge employed in various PAD products, volunteers were prompted to move their head side-to-side after the first 5 seconds of recording. Hence, the acquired BF videos contain both still, frontal face views, as well as faces turning left and right. Motion information and depth were also recorded when possible[3]. In total, the BF part of the SOTERIA dataset consists of 8400 diverse captures from 70 identities.

3.2. Presentation Attack Captures

The BF samples were used to create and capture 4 different types of attack: print, mobile replay, TV replay, and projector replay. For each type of PA, several indoor scenarios were considered in order to add variety and noise. A sample of the 24000 captured PAs is provided in Figure 2.

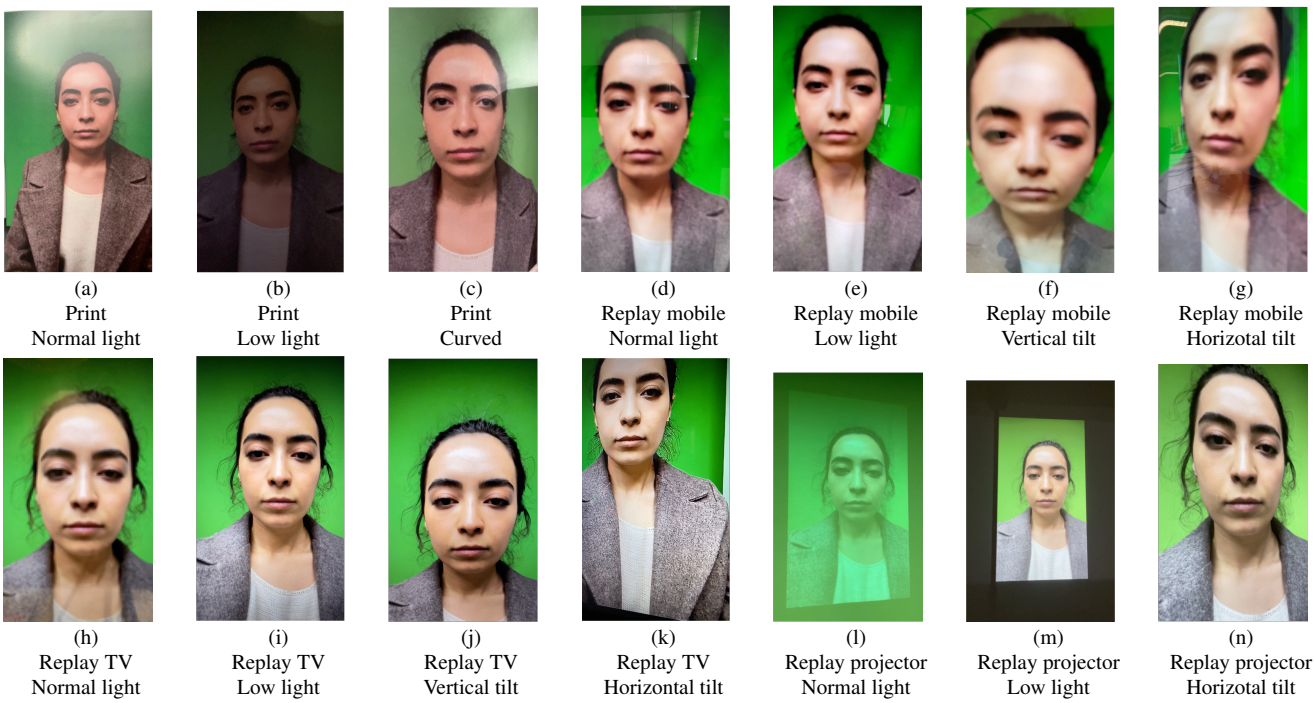
Print attacks: Best-effort print attacks were targeted. Face photos were printed on glossy paper using a laser printer, and matte paper using an inkjet printer. The print attacks were then recorded using both the front and main cameras of CIs in 3 scenarios: (a) normal light, (b) low light, and (c) curved. Some print attacks were recorded from the side with an unconstrained angle, including frontal views. The printed photos were also curved and recorded front-on.

Mobile replay attacks: Smartphones were used as both PAI (attacking device) and CI (attacked device). We defined a set of pairs of attack*ed*/attack*ing* devices (Table 2) so that each CI is attacked via two display technologies (OLED and LCD-Based).

Two configurations of mobile replay were performed: (1) best-effort PA, where for each identity, only the best BF capture was used to perform PAs through multiple attacked/attacking device pairings; and (2) medium-effort PA, where all BF captures were used to perform a single PA each, with rolling pairings.

Four scenarios were applied: (d) normal light, (e) low light (nominal), (f) vertical tilted low light, (g) horizontal

[3]Can be exploited for future work on multimodal PAD systems. This paper considers only the RGB data for benchmarking.



(a) Print Normal light
(b) Print Low light
(c) Print Curved
(d) Replay mobile Normal light
(e) Replay mobile Low light
(f) Replay mobile Vertical tilt
(g) Replay mobile Horizotal tilt
(h) Replay TV Normal light
(i) Replay TV Low light
(j) Replay TV Vertical tilt
(k) Replay TV Horizontal tilt
(l) Replay projector Normal light
(m) Replay projector Low light
(n) Replay projector Horizotal tilt

Figure 2. Examples of captures included in the SOTERIA dataset

tilted low light. The low light scenario minimizes reflection in the recording and is considered to be the best-effort capture scenario.

Attack*ed* Device	**Attack*ing* Devices**
iPhone 12	iPhone 6S & S9
iPhone 6S	iPhone 12 & Redmi 6
S9	iPhone 12 & Redmi 9A
Redmi 6 pro	iPhone 12 & Redmi 9A
Redmi 9A	iPhone 12 & S9

Table 2. Pairing attack*ed*(CI)/attack*ing*(PAI) devices considered for mobile replays attacks.

TV replay attacks: For each data subject, two captures were selected and replayed on a wide HD TV. Replays were captured using the front CIs in 4 scenarios: (h) normal light, (i) low light (nominal), (j) vertical tilted low light, and (k) horizontal tilted low light.

Projector replay attacks: A lamp-based projector (IBM iLM300) was used to project the videos on two screens: white and green. The projector was placed 2m away from the projection screen. The attacks were recorded with the front CIs by replaying two captures per identity in 3 scenarios: (l) normal light, (m) low light, and (n) tilt.

4. Vulnerability Analysis

This section presents a vulnerability analysis of the FV model IResNet100 [4] to PAs in different scenarios. We chose IResNet100 as it is one of the best off-the-shelf SOTA FV models in use today. The vulnerability to four different PAs (print and replay mobile/TV/projector) is evaluated by analysing the relationship between the BF genuine (G), zero-effort impostor (ZEI), and PA scores. The decision threshold was computed on IJB-C @FMR=0.1%. As mentioned in Section 3.1, data subjects participated in two recording sessions. For the vulnerability analysis, the first and second session data were used for enrollment and probing, respectively.

The vulnerability analysis provides insight into the ability of the IResNet100 FV model to differentiate between PAs and BF presentations, and therefore the ability of the SOTERIA PA dataset to spoof this model. We analyze the results in terms of the distribution of scores obtained from the BF face comparisons (G and ZEI) and each PA type (Figure 3). We also quantify the vulnerability to each PA in terms of Impostor Attack Presentation Match Rate (IAPMR) (Table 3), which indicates the percentage of PAs that are falsely "accepted" as G users of the FV system. So, the higher the IAPMR, the more vulnerable the system is to the PA.

In Figure 3, it is evident that the clearest separation is be-

[4] https://github.com/deepinsight/insightface

	Print		Replay mobile		Replay TV	Replay projector		
	matte-inkjet	glossy-laser	LCD	OLED	LCD	Green	White	
Normal light	99.5	99.6	92.1	93.1	91.9	83.9	95.0	
Low light	96.7	99.8	91.5	90.3	92.2	82.8	96.8	
Curved	99.7	99.5	-	-	-	-	-	
Tilted	-	-	84.3	85.4	88.5	58.0	89.1	
Avg	99.3		88.3		90.3	81.9		**89.2**

Table 3. IAPMR (%) for different PA scenarios. The threshold was set on IJB-C @FMR=0.1%. Columns represent the attack device: paper & printer types for print attacks, replay device for replay mobile and TV attacks, background wall colour for projector attacks. Rows represent different recording conditions: normal & low light for all attacks, tilted recording angle for replay attacks, and curved paper for print attacks.

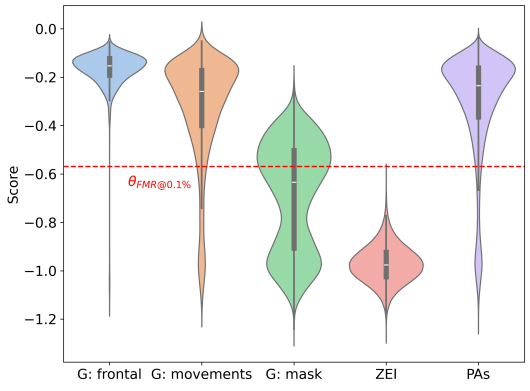


Figure 3. BF G scores when face data was acquired only frontal-on (G: frontal), with side-to-side head movements (G: movements), and with a hygienic mask (G: mask). ZEI distribution corresponds to BF ZEI scores (both frontal and side-to-side head movement face captures). PAs distribution represents the combined scores for print, replay-mobile, replay-tv, and replay-projector PAs. Decision threshold is set @FMR=0.1% on IJB-C.

tween the G: frontal face and ZEI scores. This would result in the best FV accuracy. Adding variations like side-to-side head movements and hygienic masks has the effect of dispersing the G scores and pushing the distribution closer to the ZEI distribution. This would result in a decrease in FV accuracy. Regarding the PAs score distribution, for the most part there is a clear overlap with the genuine scores (particularly the frontal face and side-to-side head movement distributions). This is confirmed by the high average IAPMR of 89.2% across all PA types in Table 3, which implies that 89.2% of the print and replay attacks considered in this analysis would succeed in fooling the IResNet100 FV model (i.e., the model would classify these attacks as genuine BF face samples). Table 3 also shows that, on average, the IResNet100 FV model is most vulnerable to print attacks (99.3% IAPMR), least vulnerable to projector-based replay attacks (81.9% IAPMR), and almost equally vulnerable to mobile-based and TV-based replay attacks.

Figure 4 presents the score distributions for the four different PA types separately. For all PAs, it appears that lighting does not have a significant effect on the comparison scores obtained by the IResNet100 FV model. This is confirmed by Table 3, where the IAPMR between the "normal light" and "low light" scenarios differs by a maximum of only about 3% across all PAs. For print attacks, curving the paper on which the face has been printed when presenting it to the attacked device, also has no perceivable effect on the obtained scores. The IAPMR is 99.5% – 99.7% depending on the type of paper used for the PA (Table 3). For mobile-based replay attacks, changing the screen technology (i.e., LCD versus OLED) also does not affect the vulnerability of IResNet100 to a high degree (Table 3 shows a maximum IAPMR difference of about 1%). Tilting (horizontally or vertically) the attacked device (CI) when attempting a mobile- or TV-based replay attack seems to lower the resulting scores, pushing more of them below the match threshold @FMR of 0.1%. This makes it more likely that a score would be classified as a ZEI – consequently, the IAPMR (vulnerability) is reduced, making the attacks less likely to fool the IResNet100 FV system. This effect is most pronounced in projector-based replay attacks, when the background colour is changed from white to green, in which case the IAPMR drops to 58%. This makes sense, because images projected on a green screen add an unnatural hue to the face image (see Figure 2(l)), incorporating a tilt skews the face (see Figure 2(n)), and these two factors combined make it more difficult to match the PA to the original (genuine) face (thereby reducing the system's vulnerability to this PA).

Overall, our analysis shows that IResNet100 is highly vulnerable to all PAs in the SOTERIA dataset, except for projector replay attacks on a green screen combined with tilting the CI. Changing the lighting in the recording environment, the curvature of printed face photos, the type of paper and printer used for creating print attacks, the screen technology for replay mobile attacks, and in most cases the tilting of the CI, do not significantly reduce the system's vulnerability (IAPMR). Since IResNet100 is one of the most popular SOTA FV models, our observations sug-

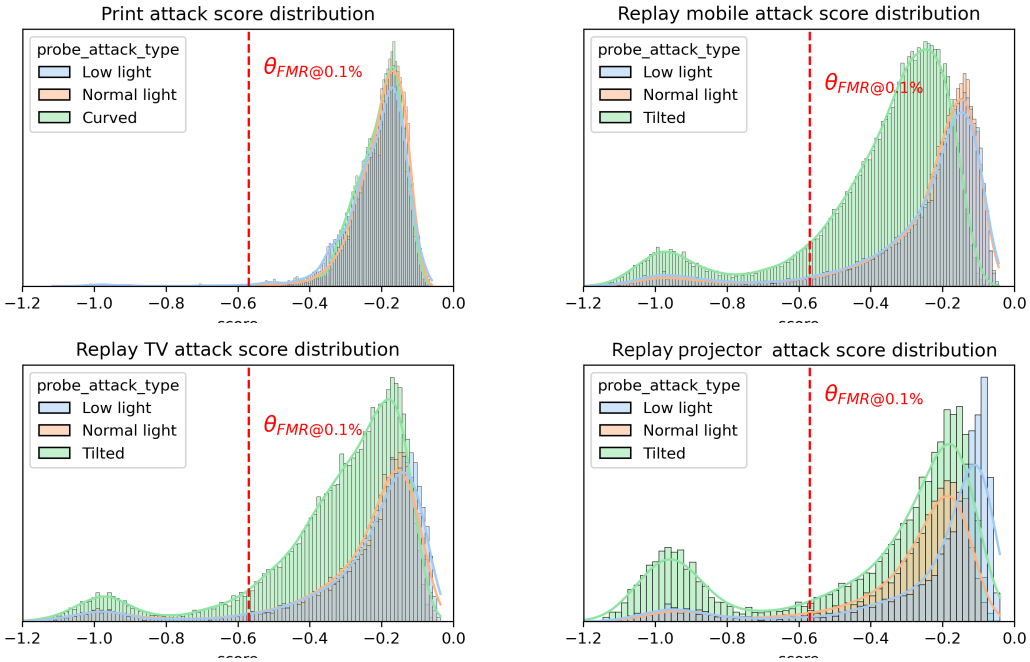


Figure 4. IResNet100 score distributions for 4 different PA types.

gest that the PAs in SOTERIA are sufficiently challenging to fool modern FV systems. This highlights the relevance of our dataset in the current face recognition landscape. Furthermore, as all PAs in SOTERIA can be easily created by the average person, our vulnerability analysis suggests a pressing need for PAD algorithms that are capable of thwarting such PAs.

5. Presentation Attack Detection

This section evaluates the PAD capabilities of an open-source PAD model, DeepPixBiS [3], trained and/or tested on the SOTERIA dataset. DeepPixBiS is a frame-level CNN-based framework for PAD that relies on pixel-wise binary supervision by using a loss function combining binary loss and pixel-wise binary loss. Section 5.1 presents a cross-dataset analysis of DeepPixBiS on publicly available PAD datasets, and Section 5.2 applies DeepPixBiS to PAs acquired in a real industrial scenario.

5.1. Cross-dataset Analysis

We trained DeepPixBis on three public, self-captured, and mobile-centric datasets: Replay-Mobile[6], OULU-NPU[5], and our new dataset (SOTERIA). For Replay-Mobile we used the default training protocol, for OULU-NPU we used protocol 4, and for SOTERIA we split the dataset into 30-20-20 subjects for train, dev and eval, respectively. For training, we used 10 frames equally spaced over the length of the video, and for evaluation we used 20. When training a PAD model, ideally we want it to generalize across different PAD datasets. This would provide confidence in the model's ability to detect PAs across different domains, and not only PAs captured in the same conditions as the PAs on which the model was trained. To evaluate the suitability of SOTERIA for training a generalizable PAD model, Table 4 compares the APCER of DeepPixBiS when trained on SOTERIA versus Replay-Mobile and OULU-NPU.

From Table 4 we see that, when DeepPixBiS is trained on SOTERIA, the cross-dataset APCER is comparable to the intra-dataset APCER for some PA types. For print attacks, the APCER of 4.1% obtained when evaluating DeepPixBiS on Replay-Mobile is better than the APCER of 11.8% on SOTERIA. When evaluating on OULU-NPU, however, the APCER of 20.7% is worse than that on SOTERIA. In comparison, when DeepPixBiS is trained on OULU-NPU, the cross-dataset APCER is always worse than the intra-dataset APCER: 64.8% on SOTERIA's print attacks and 23.9% on Replay-Mobile, compared to 8.9% on OULU-NPU itself. A similar trend can be observed when DeepPixBiS is trained on Replay-Mobile (albeit with lower APCERs). So, in terms of print attacks, the SOTERIA-trained model seems reasonably generalizable, at least more so than the Replay-Mobile and OULU-NPU models.

In terms of replay attacks, Table 4 shows that, when DeepPixBiS is trained on SOTERIA, the APCER of 3.4% for replay attacks on OULU-NPU is comparable to the 0.8 – 7.3% APCER for replay attacks on SOTERIA. However, the APCER of 30.6% on Replay-Mobile's matte screen (re-

	Replay-Mobile		OULU-NPU		SOTERIA			
PAI	print	matte screen	print	replay	print	replay-mobile	replay-tv	replay-projector
Replay-Mobile	0.0	0.0	14.9	27.7	12.9	52.8	25.3	45.6
OULU-NPU	23.9	22.7	8.9	5.1	64.8	10.2	41.2	32.9
SOTERIA	4.1	30.6	20.7	3.4	11.8	0.8	7.3	6.2

Table 4. APCER(%) @EER of DeepPixBiS when trained (rows) and tested (columns) on different datasets.

play) attacks is much higher than the APCER achieved on SOTERIA's replay attacks. This is understandable, because SOTERIA contains no PAs captured on matte screen devices (e.g., tablets). When DeepPixBiS is trained on OULU-NPU instead, the cross-dataset APCERs of 22.7% for matte screen attacks on Replay-Mobile and 10.2 – 41.2% on SOTERIA's replay attacks are all higher than the APCER of 5.1% for replay attacks on the OULU-NPU dataset itself. Similarly, when DeepPixBiS is trained on Replay-Mobile, the cross-dataset APCER for replay attacks is always much worse than the 0% intra-dataset APCER (27.7% on OULU-NPU and 25.3 – 52.8% on SOTERIA). So it seems that, even for replay PAs, the SOTERIA-trained DeepPixBiS model generalizes reasonably well, more so than the Replay-Mobile and OULU-NPU models.

The final observation from Table 4 is that, when DeepPixBiS is trained on Replay-Mobile or OULU-NPU, the PAD performance on SOTERIA's attacks suffers. In particular, the Replay-Mobile model achieves an intra-dataset APCER of 0%, but on SOTERIA the APCER varies from 12.9% to 52.8%. Similarly, on its own dataset the OULU-NPU model obtains APCER of 5.1 – 8.9%, whereas on SOTERIA the APCER is 10.2 – 64.8%. This may be attributed to the use of different devices in capturing the SOTERIA PAs, as well as to two novelties in SOTERIA that are not present in the other datasets (highlighting the need to include these PAs in PAD training): (i) the replay TV and projector attacks, and (ii) the inclusion of side-to-side head movements in all replay attacks. So, we may conclude that the Replay-Mobile and OULU-NPU datasets do not lend themselves well to training a PAD model (at least DeepPixBiS) that is generalizable across the set of PAs represented in SOTERIA.

These findings suggest that training on the SOTERIA dataset results in PAD models that are more generalizable than models trained on the Replay-Mobile or OULU-NPU datasets. However, to improve the overall performance of the PAD model, a better approach may be to train it on multiple datasets instead of a single one. This way, a larger number and greater variety of BF and PA samples would be present in the training data, so we could expect the trained PAD model to be more agnostic to the different datasets' domains and thus more generalizable. This will be studied in future works.

5.2. Application to Industrial Scenario

As a complementary analysis, we conducted experiments using attacks observed in an industrial application. Identity verification companies usually provide services to automatically authenticate an identity document and its owner. This includes a face verification step that is conducted by comparing a video "selfie" to the face picture in the document: this check is subjected to PAs, which are commonly observed. As the selfie is captured by the user "in the wild", recordings are noisy and diverse in terms of environment, capture quality, acquisition device, post-processing and compression. PA are also highly diverse in terms of type (see Distribution in Table 5) and quality, yet most PA are low-effort. We extracted and manually labeled 20000 sessions from an industrial production flow, among which approximately 1000 samples are PAs. This includes various PAs, one of which is a new PA unseen by all explored datasets: referred to as ID Doc., it corresponds to the presentation of an identity document instead of a selfie, which is usually glossy with security features possibly overlaying the face zone. As these data are confidential, this short analysis exclusively aims to provide insights into challenges that must be dealt with for PAD models to be industry-ready.

The three models trained for the cross-dataset analysis in Section 5.1 were applied to these data. The APCER per observed PA type for each model is presented in Table 5. We can see that the model trained using the SOTERIA dataset outperforms the two others, with lower APCER for all PA types, thus underlining the generalization capabilities enabled by our dataset. We also observe that, while PAs using monitors and prints as PAI tend to be encountered less frequently than the other PAs in practice, they also seem to be harder to detect, with up to 17.9% of print attacks and 16.7% of picture replays from monitors not being rejected.

Overall, the model trained using SOTERIA achieves 10.61% EER, and 55.81% BPCER @ APCER of 1%. In industry, a minimal APCER is targeted; however, the resulting BPCER achieved by our best model is impractical as more than half of BF presentations would be wrongfully rejected.

	Picture				Video	
PAI	Print	ID doc.	Mobile	Monitor	Mobile	Monitor
Distribution (%)	12.2	19.3	43.6	3.50	17.9	2.40
Replay-Mobile	45.5	57.2	39.2	25.0	33.9	20.8
OULU-NPU	28.5	7.7	21.9	38.9	21.7	29.2
SOTERIA	**17.9**	**4.10**	**8.2**	**16.7**	**12.2**	**12.5**

Table 5. APCER(%) @EER achieved by DeepPixBis models, trained on 3 datasets, on production samples.

6. Conclusion

We captured and presented the SOTERIA dataset, which provides over 8k BF samples from 70 consenting subjects, and 24k PA samples crafted and recorded using different devices and in diverse scenarios. Our cross-dataset and industry-based experiments show that SOTERIA enables PAD models to achieve decent generalization capabilities for both print and replay PAs, compared to training on other mobile-based PA datasets like Replay-Mobile and OULU-NPU. The richness of the dataset allows for further analyses that we plan to present in future work, including an analysis of the effect of devices (as PAI or CI) and recording environments on PAD capabilities, an investigation into the impact of motion challenge on PAD, and an evaluation of the fairness of PAD algorithms (e.g., gender bias). Areas of improvement include the addition of new PAs, in particular prints attacks, which are under-represented in the dataset.

Acknowledgments

We are deeply grateful to all the volunteers who provided their data and time to construct the SOTERIA dataset, the engineers from both IDnow and Idiap who developed the data capture apps, and everyone involved in creating and capturing the PAs.

Funding

SOTERIA has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101018342.

References

[1] Bob: A framework for signal processing and machine learning.

[2] W. R. Almeida, F. A. Andaló, R. Padilha, G. Bertocco, W. Dias, R. d. S. Torres, J. Wainer, and A. Rocha. Detecting face presentation attacks in mobile devices with a patch-based cnn and a sensor-aware loss function. *PloS one*, 15(9), 2020.

[3] S. M. Anjith George. Deep pixel-wise binary supervision for face presentation attack detection. In *ICB 2019*, 2019.

[4] A. Anjos, M. Günther, T. de Freitas Pereira, P. Korshunov, A. Mohammadi, and S. Marcel. Continuously reproducing toolchains in pattern recognition and machine learning experiments. In *International Conference on Machine Learning (ICML)*, Aug. 2017.

[5] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid. Oulu-npu: A mobile face presentation attack database with real-world variations. In *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*. IEEE, 2017.

[6] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel. The replay-mobile face presentation-attack database. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2016.

[7] M. Fang, F. Boutros, A. Kuijper, and N. Damer. Partial attack supervision and regional weighted inference for masked face presentation attack detection. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*. IEEE, 2021.

[8] M. Fang, M. Huber, and N. Damer. Synthaspoof: Developing face presentation attack detection based on privacy-friendly synthetic data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[9] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel. Biometric face presentation attack detection with multi-channel convolutional neural network. *IEEE Transactions on Information Forensics and Security*, 15, 2019.

[10] H.-P. Huang, D. Sun, Y. Liu, W.-S. Chu, T. Xiao, J. Yuan, H. Adam, and M.-H. Yang. Adaptive transformers for robust few-shot cross-domain face anti-spoofing. In *European Conference on Computer Vision*. Springer, 2022.

[11] A. Komaty, V. K. Hahn, C. Ecabert, and S. Marcel. Can personalised hygienic masks be used to attack face recognition systems? In *2023 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2023.

[12] A. Liu, Z. Tan, J. Wan, S. Escalera, G. Guo, and S. Z. Li. Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021.

[13] Z. Ming, M. Visani, M. M. Luqman, and J.-C. Burie. A survey on anti-spoofing methods for facial recognition with rgb cameras of generic consumer devices. *Journal of Imaging*, 6(12), 2020.

[14] Z. Ming, Z. Yu, M. Al-Ghadi, M. Visani, M. M. Luqman, and J.-C. Burie. Vitranspad: video transformer using convolution and self-attention for face presentation attack detection. In *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022.

[15] A. Mohammadi, S. Bhattacharjee, and S. Marcel. Improving cross-dataset performance of face presentation attack detection systems using face recognition datasets. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020.

[16] D. Pasmino, C. Aravena, J. Tapia, and C. Busch. Flickr-pad: New face high-resolution presentation attack detection database. *arXiv preprint arXiv:2304.13015*, 2023.

[17] N. Sergievskiy, R. Vlasov, and R. Trusov. Generalizable method for face anti-spoofing with semi-supervised learning. *arXiv preprint arXiv:2206.06510*, 2022.

[18] C.-Y. Wang, Y.-D. Lu, S.-T. Yang, and S.-H. Lai. Patchnet: A simple face anti-spoofing framework via fine-grained patch recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[19] D. Wang, J. Guo, Q. Shao, H. He, Z. Chen, C. Xiao, A. Liu, S. Escalera, H. J. Escalante, Z. Lei, et al. Wild face anti-spoofing challenge 2023: Benchmark and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[20] Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao. Deep learning for face anti-spoofing: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(5), 2022.

[21] Y. Zhang, Z. Yin, Y. Li, G. Yin, J. Yan, J. Shao, and Z. Liu. Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*. Springer, 2020.