Mobile Contactless Fingerprint Presentation Attack Detection: Generalizability and Explainability

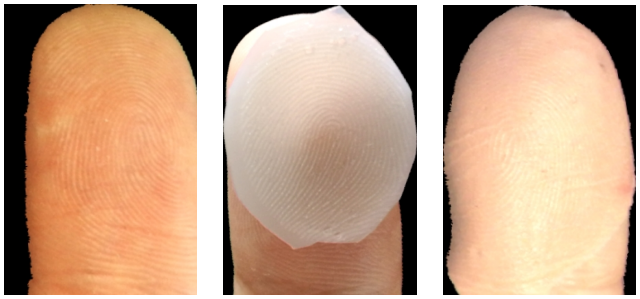
Jannis Priesnitz[1], Roberto Casula[2], Jascha Kolberg[1], Meiling Fang[3,4], Akhila Madhu[1], Christian Rathgeb[1], Gian Luca Marcialis[2], Naser Damer[3,5], Christoph Busch[1]

Abstract—Contactless fingerprint recognition is an emerging biometric technology that has several advantages over contact-based schemes, such as improved user acceptance and fewer hygienic concerns. Like for most other biometrics, Presentation Attack Detection (PAD) is crucial to preserving the trustworthiness of contactless fingerprint recognition methods. For many contactless biometric characteristics, Convolutional Neural Networks (CNNs) represent the state-of-the-art of PAD algorithms. For CNNs, the ability to accurately classify samples that are not included in the training is of particular interest, since these generalization capabilities indicate robustness in real-world scenarios.
In this work, we focus on the generalizability and explainability aspects of CNN-based contactless fingerprint PAD methods. Based on previously obtained findings, we selected four CNN-based methods for contactless fingerprint PAD: two PAD methods designed for other biometric characteristics, an algorithm for contact-based fingerprint PAD and a general-purpose ResNet18. For our evaluation, we use four databases and partition them using Leave-One-Out (LOO) protocols. Furthermore, the generalization capability to a newly captured database is tested. Moreover, we explore t-SNE plots as a means of explainability to interpret our results in more detail. The low D-EERs obtained from the LOO experiments (below 0.1% D-EER for every LOO group) indicate that the selected algorithms are well-suited for the particular application. However, with an D-EER of 4.14%, the generalization experiment still has room for improvement.

Index Terms—Contactless Fingerprint Recognition, Presentation Attack Detection, Generalizability

1 Introduction

In recent years, contactless fingerprint recognition has been introduced as a more convenient alternative to contact-based schemes [11], [12], [13]. Contactless fingerprint technologies enable the recognition of individuals without any contact between a capture device surface and a fingertip [11], [13]. In contrast to contact-based capturing schemes where the finger is pressed onto a planar surface, contactless recognition workflows do not require any contact between the subject and the capturing subsystem. This avoids distinct issues like low contrast caused by dirt, humidity on the capturing device, or latent fingerprints. Comparative usability studies between contactless and contact-based fingerprint recognition schemes indicate that contactless capturing schemes have better usability and higher user acceptance [14], [15], [16]. This is especially true in multi-user scenarios, where different individuals share one capture device. In these cases, the subjects might have fewer hygienic concerns using contactless fingerprint recognition [14], [15]. For contactless fingerprint recognition, a wide variety of capturing devices have been developed so far. This includes expensive stationary devices for capturing 3D samples, specialized hardware setups for capturing 2D samples and lightweight mobile solutions. However, it can be observed from the literature that most contactless fingerprint capturing devices are mobile handheld devices like smartphones [13], [14], [15].



(a) BP (b) low-quality PA (c) high-quality PA

Fig. 1: Fingerprint images captured using a contactless fingerprint capture device. (a) BP, (b) wood-glue PA placed on a fingertip, (c) color adjusted Body Double PA as overlay.

Like most biometric systems, contactless fingerprint recognition schemes are vulnerable to Presentation Attacks (PAs). Here, artificial replicas, *i.e.* Presentation Attack Instruments (PAIs), representing a fingerprint characteristic are used to spoof the system. These replicas can be made of various materials like

- [1] *da/sec – Biometrics and Security Research Group, Hochschule Darmstadt, Germany, E-mail:* {firstname.lastname}@h-da.de
- [2] *PRA Lab, University of Cagliari, Italy E-mail:* {firstname.lastname}@unica.it
- [3] *Fraunhofer Institute for Computer Graphics Research IGD, Germany E-mail:* {firstname.lastname}@igd.fraunhofer.de
- [4] *College of Information Engineering, Yangzhou University, Yangzhou, China E-mail:* {firstname.lastname}@yzu.edu.cn
- [5] *Department of Computer Science, TU Darmstadt, Darmstadt, Germany*

Manuscript received April 19, 2005; revised August 26, 2015.

TABLE 1: Overview of published works concerning contactless fingerprint PAD. BD: Body Double, BW: beeswax, DS: Dragon Skin, EL: Ecoflex Layover, GL: Gelatine, LL: Latex Layover, MG: Moldable Glue, PL: Playdoh Layover, PP: Printed Photo Paper, RP: replay attack, SC: Silicone, SF: Synthetic Fingerprint, SP: Silly Putty, WL: Woodglue Layover

Author	Year	Method	Databases	DB size (BP / PA)	PAI species	Detection Performance
Stein *et al*. [1]	2013	Reflection properties on video frames	in-house	N/A	SC, PP, GL	77% detected
Taneja *et al*. [2]	2016	LBP, DSIFT, LICUDI & SVM	PA: Fingerphoto Spoofing, BF: ISPFDv1	12,288 (8,192 / 4,096)	PP, RP	61.89% TAR @ 0.1% FAR (LBP)
Zaghetto *et al*. [3]	2017	GLCM, ILBP, PCA & ANN	in-house	400 (200 / 200)	BW, PL, WL, LL, SC	97.56% detected
Fujio *et al*. [4]	2018	AlexNet, LBP & SVM	PA: Fingerphoto Spoofing, BF: ISPFDv1	12,288 (8,192 / 4,096)	PP, RP	0.04% HTER (CNN)
Wasnik *et al*. [5]	2018	LBP, BSIF, HOG & SVM	in-house	900 (750 / 150)	RP	0% BPCER @ 5% APCER, D-EER = 0.49
Marasco and Vurity [6]	2022	Multi color-spaces, patch-based, multiple CNNs	PA: Fingerphoto Spoofing, BF: ISPFDv1	12,288 (8,192 / 4,096)	PP, RP	0.84% D-EER (fusion)
Adami *et al*. [7]	2023	ResNet with adapted loss functions	COLFISPOOF, LivDet2023	65,972	EL, PL, WL, SF, LL, PP	0.63% APCER @ 0.12% BPCER
Priesnitz *et al*. [8]	2023	SpoofBuster	PA: COLFISPOOF; BP: ISPFDv1 and v2, two in-house DBs	28,540 (1,069, 4,096, 16,175 / 7,200)	BD, EL, WL, SP, PP, DS, MG	avg. 1.08% APCER @ 1% BPCER, 0.86% D-EER (ISPFDv2)
Purnapatra *et al*. [9]	2023	DenseNet-121, NasNetMobile	LivDet2023	23,036 (8,604 / 14,432)	EL, PL, WL, SF, LL, PP	0% – 79.01% APCER @ 0.18% – 9.04% BPCER
Purnapatra *et al*. [10]	2023	8 LivDet candidates	LivDet2023	23,036 (8,604 / 14,432)	EL, PL, WL, SF, LL, PP	Winner: avg. 11.35% APCER @ 0.62% BPCER

gelatin, silicone, different glues, or latex [17], [18]. Apart from materials that have proven to attack contact-based systems successfully, new unknown PAIs represent a threat, especially for contactless setups. Especially printout attacks have a high attack potential since the color of the PAI can be precisely adjusted to real fingerprints. Moreover, attacks where the capturing device is spoofed by another display device presenting a fingerprint are feasible [2]. Figure 1 shows a contactless fingerprint also referred to as Bona fide Presentation (BP) and two PA artifacts of different visual resemblance.

PAD mechanisms aim to reliably detect PAs and distinguish them from BPs. For contact-based fingerprint biometrics, PAD mechanisms can be directly integrated into the capture device, where *e.g*. a finger's impedance is measured while touching the surface [19]. However, these countermeasures are not implementable in contactless biometric systems. Here, PAD algorithms that operate exclusively on the captured images are required. However, it should be noted that the signals obtained from a mobile contactless fingerprint capturing device are most commonly color images. This property can also be exploited in PAD algorithms.

Nowadays, CNNs represent the state-of-the-art in the area of general purpose image classification algorithms and are also widely used for PAD on contactless biometrics. In comparison to image classification using hand-crafted features, CNNs generally require less pre-processing and generalize better to unseen data. This has been demonstrated in various competitions that benchmarked general object classification [20], [21] and special biometric PAD algorithms [22], [23], [24].

In our previous work [25], we presented a comprehensive benchmark of contactless fingerprint PAD (COLFIPAD), which includes a total number of 135 experiments. Nine CNNs which were originally designed for general image classification tasks or PAD for other biometrics, *e.g*. contact-based fingerprint, finger vein, face, and iris, are considered for that work. We conducted our experiments on the COLFISPOOF [26] database together with three bona fide databases. For our evaluations, we used several LOO protocols in order to get a comprehensive assessment of the PAD performance in various scenarios. Our obtained results indicate that state-of-the-art PAD algorithms can accurately detect PAs on contactless fingerprints.

In this work, we put a special focus on the generalizability and explainability aspects of CNN-based contactless fingerprint PAD. The main goal is to explore to what extent CNNs are able to generalize to new PAI species, environmental scenarios and different capturing workflows. Hence, we select the three best-performing methods from our COLFIPAD benchmark [25]. Furthermore, we include the SpoofBuster algorithm in our evaluations [8]. The SpoofBuster was originally designed for contact-based fingerprint PAD and was shown to be suitable for contactless fingerprint PAD.

Our contributions are as follows:

- We summarize and contextualize the results obtained from the COLFIPAD and SpoofBuster experiments.
- A score-level fusion is performed to evaluate the robustness benefits of the combinations of various algorithms.
- Additionally, we evaluate the cross-database generalization capabilities of our methods using a newly captured PAD database that includes color-adjusted PAI overlays from real entities.
- Finally, we use t-SNE graphs in order to explain our obtained findings in detail.

Our results prove that the employed PAD algorithms can accurately detect PAs on contactless fingerprints and generalize to new PAIs and capturing environments. In particular, for the baseline and all LOO protocols, the best-performing CNNs achieve a Detection Equal Error Rate (D-EER) below 0.1%. Furthermore, fusion strategies can contribute to more robust and accurate detection results. Especially in a cross-database experiment, the fusion of two algorithms reduces the D-EER from 5.99% to 4.14%. However, our evaluations also showcase limitations in terms of generalizability to new capturing workflows. Here, the best result is only at 15.04% D-EER. Our explainability evaluation using t-SNE plots indicates that there are substantial differences between the considered PAD methods. To summarize, the obtained findings indicate that contactless fingerprint PAD employing algorithms derived from other characteristics results in low error rates and generalizes well to unseen PAI species. Nonetheless, we have also demonstrated that the generalization capabilities to new capturing devices are constrained.

The rest of the paper is structured as follows: Section 2 summarizes the related work for contactless fingerprint PAD.

Furthermore, Section 3 presents the considered PAD methods, whereas Section 4 describes the experimental setup. The results are discussed in Section 5. Finally, Section 6 concludes this paper.

2 Related Work

Contact-based fingerprint PAD is a well-studied research area. Comprehensive overviews discuss relevant aspects in this field [18], [23], [27]. Recently, the development of software-based PAD mechanisms evolved from hand-crafted features to deep learning methods like CNNs. In contrast to contact-based fingerprint PAD, contactless fingerprint PAD mechanisms have not yet been studied comprehensively. Table 1 provides an overview on contactless fingerprint PAD mechanisms, PAI species, available databases and performance details.

2.1 PAI species

In the literature, it has been shown that, various PAI species are able to spoof contactless fingerprint recognition systems. Several works consider PAI species from the contact-based domain for their experiments. This group includes a wide variety of PAIs. All have in common that a soft, deformable material is used that simulates a ridgeline characteristic when it is pressed onto a planar surface [18]. This includes rather persistent materials like silicone, glues, latex and gelatine but also semi-persistent materials like wax, playdoh and dragon skin. Here, the latter ones tend to lose the ridge-line characteristic when pressed onto a surface. It was shown that all mentioned PAI species are able to attack contactless recognition schemes [3], [26]. Most notably, the persistence of the material plays a subordinate role for contactless fingerprint recognition. However, in contrast to contact-based schemes, the color of the PAI is of increased importance since most contactless devices capture color images, as shown by Kolberg *et al*. [26] who also published a database that includes various PAI species. Studies including several of the aforementioned PAI species were conducted by Zaghetto *et al*. [3] Adami *et al*. [7], Priesnitz *et al*. [8] and Purnapatra *et al*. [9], [10]. The results indicate that PAIs from the contact-based domain are a minor challenge for state-of-the-art PAD algorithms.

The second group represents PAI species, which are not applicable to contact-base devices. These attacks contain a ridge-line pattern but are not assembled using soft material. Two attacks specific to contactless fingerprint recognition have so far proven feasible: printout attacks and replay attacks [2], [5]. Printout attacks are assembled by printing the fingerprint on paper, whereas replay attacks use another display device, *e.g*. a smartphone, to present the PA. Both attacks have in common that they are easy to assemble and adjustable to the capture device. The publicly available Fingerphoto Spoofing database [2] includes both PAI species. Moreover, the COLFISPOOF database also contains printout attacks. Evaluations on these databases indicate that replay attacks are fairly easy to detect, whereas printout attacks pose a major challenge, especially if they are adjusted to skin color.

A third group represents digital PAs, also referred to as injection attacks [28]. Here, the PA is a digital signal containing a fingerprint characteristic. Chugh *et al*. [29] demonstrate that digital PAIs can be prepared and used to launch attacks on fingerprint recognition systems. In this context, Malhotra *et al*. [30] showcased that fingerprints obtained from social media can be used to for digital attacks and presented a method which avoids this.

2.2 PAD Algorithms

The evolution of PAD algorithms started with the analysis of image properties. Stein *et al*. [1] analyzed the reflection properties of different PAI species. Several frames of a video-based capturing attempt were considered to analyze if the amount of white reflecting pixels in the fingerprint core is above a static threshold. However, these methods are known to generalize poorly to new application scenarios and PAI species.

Further development concentrates on the extraction of discriminative features such as LBP or SIFT. The classification task is typically done by SVMs or simple neuronal networks, as shown by Taneja *et al*. [2] and Wasnik *et al*. [5]. Zaghetto *et al*. [3] suggested a PAD mechanism that extracts texture features from pre-processed images. On the extracted features, the authors conduct a Principal Component Analysis (PCA), which is used for the classification together with a feedforward Artificial Neural Network (ANN). An advantage of this approach is that there are no preconditions regarding the image size, since the used image feature extractors work on images of arbitrary size. Feature-based methods improve the robustness of the PAD systems and reduce the error rate. On the contrary, these methods tend to generalize fairly to new PAI species, as evaluated by Fujio *et al*. [4].

The more recent research was focused on CNNs for PAD purposes. Here, CNNs for general purpose object detection are fine-tuned for the special PAD task, like proposed by Fujio *et al*. [4] and Marasco and Vurity [31]. The proposals process the fingerprint samples in various ways in order to maximize detection accuracy. The most common sample pre-processing is a center cropping, whereas more advanced proposals consider *e.g*. multiple color channels [6] and minutiae-based patch extraction [6], [25]. Moreover, adaptations of loss function have been tested [7].

In 2023, the first LivDet competition on contactless fingerprint PAD was conducted by Purnapatra *et al*. [10]. Five competitors submitted eight algorithms to the competition. Most notably, all submitted algorithms are based on CNNs.

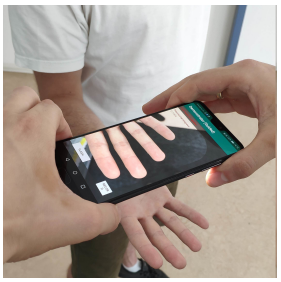
It should be noted that several works propose contactless fingerprint PAD schemes on ShortWave InfraRed (SWIR) [32], [33], Optical Coherence Tomography (OCT) [34], [35] or Laser Speckle Contrast Imaging (LSCI) [33] images. These mechanisms require specialized hardware and highly constrained capturing scenarios. The PAD algorithms suggested in these works also rely on deep learning methods like CNNs and convolutional autoencoders.

In summary, we can observe that recently, contactless fingerprint PAD gained a lot of attention. Various machine learning methods are considered to solve the task of contactless fingerprint PAD. The first approaches mainly relied on texture descriptors in combination with SVMs, whereas nowadays, CNNs represent the majority of new proposals. As shown in Table 1, the PAD research is boosted by publicly available databases like the IIITD Fingerphoto Spoofing database, the COLFISPOOF database and more recently, the database included in the LivDet2023 competition [10]. However, new proposals are difficult to compare against baseline methods because the databases used, pre-processing, evaluation protocols and performance metrics may vary.

3 Contactless Fingerprint PAD Methods

We selected four CNN-based PAD algorithms: The three best-performing methods from the COLFIPAD benchmark [25] and a fourth algorithm named SpoofBuster which we adapted in a

(a) Constrained scenario



(b) Attended scenario

Fig. 2: Two contactless fingerprint capturing scenarios: (a) Constrained scenario where a box-like setup avoids external illumination and (b) attended scenario where an operator is handling the capturing device.

previous work [8]. In this section, we describe generalization requirements and the considered methods in detail.

3.1 Generalization Requirements

A key requirement for robust PAD mechanisms in real-world setups is the ability to generalize to new scenarios. In this context, generalization refers to the correct classification of samples that deviate from the initial scenario that the algorithm was designed and trained for. Attackers might intentionally or unintentionally circumvent PAD algorithms by using distinct capturing setups. Generalization should be given in three main aspects:

- **Subjects:** The partitioning into training, validation and testing of the considered database should be subject-disjoint. This means that all samples of a bona fide subject should only occur in either the training, the validation, or the test partition due to their natural similarity. If this is not the case, a BP could be correctly classified based on the specific fingerprint characteristic and not based on the differences to PAs.
- **PAIs and PAI species:** As discussed in Section 2, a wide variety of PAI species are known to successfully attack contactless fingerprint recognition schemes. Only one PA sample might successfully attack a recognition system, but many samples are required to train a CNN. Hence, the number of samples per PAI species has to be high enough for a reliable training process. Furthermore, the variety of the PAI species presented to the CNN during the training should be high enough. This increases the CNNs generalization capabilities to unseen data. Furthermore, the PAIs should also be assembled thoroughly to train the classifier on high-quality data. Figure 4 illustrates the materials used for the preparation PAIs which are included in the COLFISPOOF database [26]. As mentioned, attackers might challenge a recognition workflow with new PAIs that are not included in the training set. For this reason, it is crucial to assess the algorithms' generalization capabilities to new PAI species. *E.g.* an algorithm that was trained on Latex and Silicone materials should also be able to detect printout attacks.
- **Capturing device and environments:** For mobile contactless fingerprint recognition, off-the-shelf devices such as smartphones are used in many scenarios. For this reason, camera setups and hence, captured fingerprint samples, may vary between different capturing devices. PAD algorithms should be robust against these deviations in order to be deployed in large-scale applications. In addition to the capturing device, the capture environment drastically influences the biometric sample and hence the decision obtained from the PAD algorithm. *E.g.* different illumination situations can lead to increased image noise or deviated skin colors. PAD algorithms should be robust against these deviations in the capturing environment. Figure 2 illustrates different capturing scenarios included in our test database.

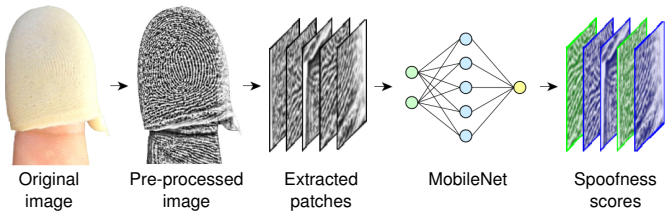


Fig. 3: Overview of the proposed SpoofBuster workflow including pre-processing, patch extraction and classification of the individual patches.

To summarize the generalization requirements, new PAD proposals should aim to include a subject-disjoint partitioning of the database and a strategy to test the proposals on PAI species which are unseen in the training set, *e.g.* LOO protocols. Furthermore, cross-database evaluations should be conducted using different capturing devices and environmental scenarios. It should be noted, that generalization to new environmental scenarios is a requirement that is not present for contact-based fingerprint recognition and additionally, generalization between capturing devices might also deviate less for contact-based methods.

3.2 COLFIPAD Candidates

We select the three best-performing CNNs for PAD from our COLFIPAD benchmark:

- A-PBS [36] is an attention-based deep pixel-wise binary supervision method for iris PAD based on DenseNet. First, the pixel-wise supervision recognizes fine-grained pixel or patch-level cues. Then, the attention mechanism guides the network to automatically find regions that most contribute to an accurate PAD decision.
- LMFD [37] represents a dual-stream convolution neural network for face PAD. Here, one stream learns features in the frequency domain, which are less influenced by sensors and illumination variations. The other stream leverages the RGB images. A hierarchical attention module joins the information from the two streams at different layers of the CNN.
- ResNet18 [38] is a 72-layered architecture with 18 deep layers. This network aims to enable large numbers of convolutional layers to function efficiently by using shortcut connections.

For ResNet18, we exchanged the last fully connected layers to support binary classifications. Apart from this, we adopted all the settings from the original proposals. All CNNs use pre-trained models that are fine-tuned using the considered databases.

3.3 SpoofBuster

As a fourth PAD algorithm, we consider the SpoofBuster originally proposed by Chugh *et al*. [39]. The SpoofBuster was originally designed for contact-based fingerprint PAD and is patch-based, whereas the patch locations are defined by minutiae positions, which are extracted using the MINDTCT method [40]. The method uses a MobileNetv1 with a modified last fully connected layer to support binary classifications. The CNN gets several image patches of one sample and outputs a Presentation Attack (PA) score in the range [0.0, 1.0] for every patch. All PA scores are averaged to the final PAD result for the tested sample. The original proposal was adapted and tested on contactless fingerprints in an earlier work [8]. In summary, the following main adaptations are applied to the original SpoofBuster method in order to make it most suitable for contactless fingerprint PAD:

- **Patch size:** We use an increased patch size of 112×112 pixels instead of 96×96 pixels as in the original approach from Chugh [39]. This has two reasons: firstly, the acquired and normalized contactless fingerprints are approx. 10% larger in our setup compared to contact-based ones, so the patch size needs to be increased. Secondly, due to the lack of rotation, we avoid protruding edges and, for this reason, it is possible to increase the patch size. Preliminary experiments using various patch sizes confirmed that the selected one works best.
- **Patch angle and alignment:** We do not implement a patch alignment based on the minutiae angle, but extract unaligned patches. The main reason for this is that most extracted patches contain parts of the fingerprint's core area that does not have a distinct orientation. Hence, an alignment would be of low benefit since only the center area would be aligned, whereas the border areas would remain unaligned. Preliminary comparative experiments with and without alignment show no significant differences for our experimental setup.
- **Patch number reduction:** To reduce the number of patches, we consider a combination of minutiae quality and background thresholding. The minutiae quality threshold excludes minutiae with a quality score below 0.25 (in a range [0, 1]) whereas the background threshold excludes patches with more than 10% white background pixels.

It should be noted that various adaptations have been tested in advance. The discussed ones turned out to perform best in the presented experimental setup. All further settings, such as the usage of MINDTCT and the MobileNetv1 training parameters, remain the same as in the original SpoofBuster algorithm. In summary, we replaced the last MobileNetv1 layer with a 2-unit softmax layer, set the learning rate to 0.0001 and trained 50 epochs with an early stop if not improvement was seen for 10 epochs.

4 Experimental Setup

This section describes the databases, pre-processing and evaluation protocols we used for the evaluation of the selected CNNs.

4.1 Considered Databases

For our work, we consider one database containing contactless fingerprint PAs in combination with three databases that contain bona fide contactless fingerprints. Furthermore, for our cross-database evaluation, a new database was acquired. Table 2 summarizes the number of samples for each database, whereas this section describes the acquisition processes:

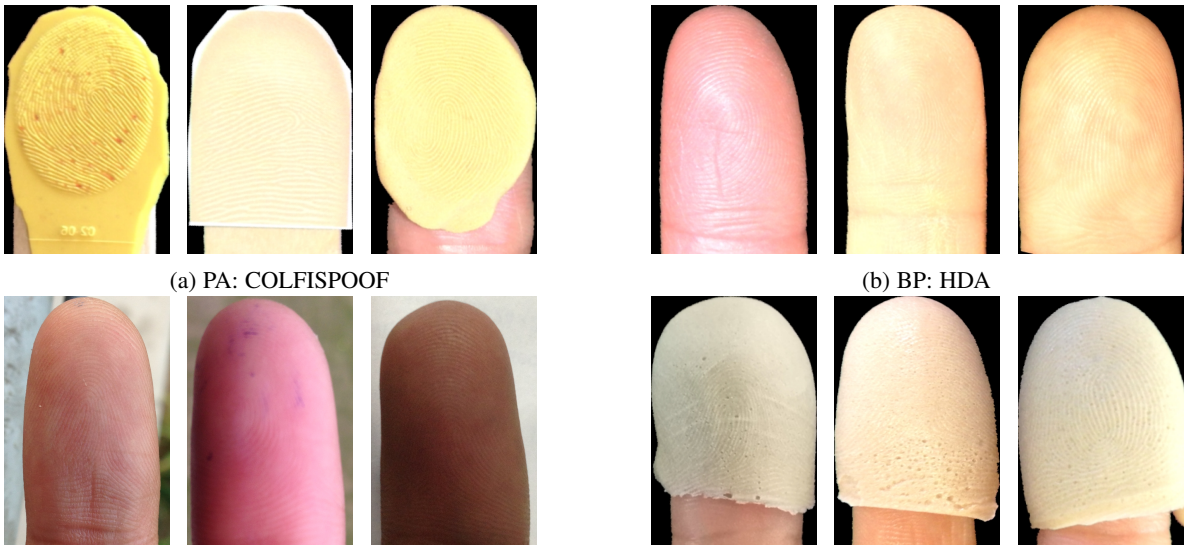
(a) PAI Materials (b) PAI overlay preparation

Fig. 4: Illustrations of the PAI preparation process: (a) materials which have proven to successfully attack contactless fingerprint recognition and (b) the preparation process of PAI overlays which are adjusted to skin color.

TABLE 2: Number of BPs and PAs for each database. It should be noted that the UniCa-HDA database is only used for testing.

Database	BF Samples	PA Samples
COLFISPOOF	–	7,200
HDA	1,069	–
ISPFDv1	4,029	–
ISPFDv2	16,175	–
UniCa-HDA	2,040	1,512

- COLFISPOOF [26]: COLFISPOOF is acquired using a contactless fingerprint recognition system utilizing two different smartphones as capture devices. The database comprises 7,200 samples of 72 different PAI species. It includes various PAI species that are known to effectively attack contact-based capture devices like playdoh, dragon skin, sillyputty (each with various colors), transparent overlays made of gelafix and different glues. It also includes printout attacks of different colors, which effectively attack contactless capturing devices. The PAIs contain only synthetic ridge patterns and no BPs are included in the database, which is why the database is publicly available. It should be noted that all PAs are captured and pre-processed using an automated contactless fingerprint recognition workflow. The detailed description of this can be found in [8], [14], [26] Figure 5a presents some example images of this database.
- HDA [14]: The HDA database consists of contactless samples captured in two different setups: a constrained box-setup and an unconstrained tripod-setup. For the capturing, the authors used two different smartphones. An application automatically captured the four inner-hand fingers and processed them into fingerprint samples. The authors captured and processed a database of 29 subjects in two rounds, which resulted in 1,360 individual fingerprints. Example images of the database are presented in Figure 5b.
- ISPFDv1 [41]: The IIITD SmartPhone Fingerphoto Database v1 consists of contactless fingerprint images collected with an Apple iPhone 5 smartphone. For the capturing, the smartphone's inbuilt photo application was used and hence the samples included in the database are not further pre-processed or quality assured. The authors



(a) PA: COLFISPOOF (b) BP: HDA

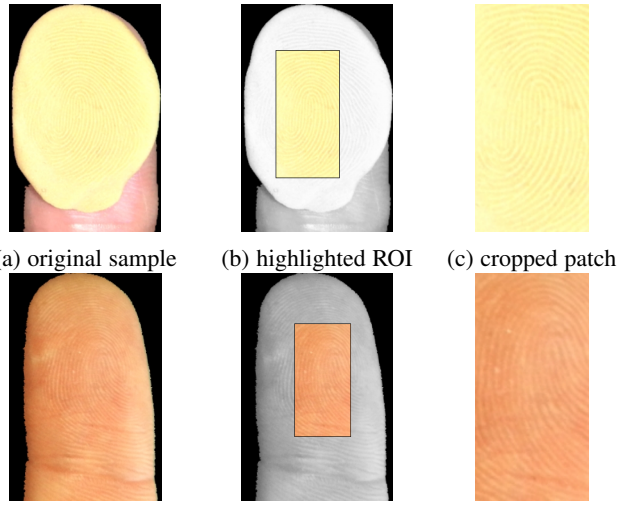
(c) BP: ISPFDv1 / ISPFDv2 (d) AP: UniCa-HDA

Fig. 5: Example finger images of the used datasets. It should be noted that the ISPFDv2 is visually similar to the ISPFDv1.

captured four different environmental scenarios, including indoor and outdoor images with natural and white background. Every scenario contains 1,024 finger photos from 64 subjects, which results in 4,096 contactless samples. Figure 5c depicts example images of the database.

- ISPFDv2 [42]: The IIITD Smartphone Finger-Selfie Database v2 consists of 16,175 self-captured and unprocessed fingerphotos obtained from 304 subjects. The fingerphotos are taken using two smartphones. Like for the ISPFDv1 database the inbuilt photo application was used for the capturing. The database contains three scenarios: Indoor with a constrained and unconstrained background and outdoor.
- UniCa-HDA: In addition, we conduct a test on the newly captured UniCa-HDA database, comprising BPs and PAs. Here, the same capturing setup as in the HDA database was used, but a different environmental scenario was chosen. Along with the newly acquired BPs, we prepared and captured PAs of high quality. For this, the subject's fingers are covered with a thin overlay of Body Double, which is then turned inside out after the material has solidified. To make the PAI as realistic as possible, we added additional color to the PA material. This process makes the PAI appear as similar as possible to the subject's real skin color, *c.f.* Figure 5 (d). The overlays are then captured in the same setup as the BP fingerprints. Figure 4 illustrates the PAI preparation process. It should be noted that the obtained PAIs are a mirrored copy of the original source.

For our experiments, we combine COLFISPOOF with each of the bona fide databases. Furthermore, we evaluate all sub-databases of each database together to maximize the number of samples for training and evaluation and to provide a realistic scenario. As introduced before, it should be noted that in mobile real-world application scenarios, different environmental situations can occur during capturing attempts. For this reason, a training, and evaluation database that includes different scenarios should be chosen in order to achieve a robust PAD mechanism that generalizes across different capturing environments.



(a) original sample (b) highlighted ROI (c) cropped patch

(d) original sample (e) highlighted ROI (f) cropped patch

Fig. 6: Illustration of the patch generation process for the COLFIPAD candidates: (a – c) PA sample (COLFISPOOF) and (d – f) a BP sample (HDA Database).

4.2 Sample Pre-processing

We apply two different pre-processing methods to the database in order to meet the requirements of the considered CNNs and conduct a fair evaluation.

COLFIPAD Candidates

The databases have to be pre-processed to be suitable for the training and evaluation procedure. Also, the background and finger / PAI border region are cropped, since some PAIs included in the COLFISPOOF database are not looking realistic in the outer areas (*c.f.* Figure 5a). This process avoids that the PAD methods classify samples based on their border region and might fail on more thoroughly prepared PAIs. We adopt the pre-processing pipeline

TABLE 3: Number of PAIs and bona fide samples for each partition for the evaluation protocols.

Database	Protocol	Train	Validation	Test
COLFISPOOF	baseline	2,160	1,440	3,600
	LOO printout	4,200	1,800	1,200
	LOO transparent	4,690	2,010	500
	LOO default color	2,450	1,050	3,700
	LOO colored silicone	3,780	1,620	1,800
HDA	–	322	547	200
ISPFDv1	–	1,207	766	2,056
ISPFDv2	–	4,664	3,287	8,224
UniCa-HDA	Cross-database	*testing only*		3,552

suggested in the COLFISPOOF paper [26] for our experiments and apply it to both, COLFISPOOF and the bona fide databases.

Since both ISPFD databases consist of fingerphotos, the finger area is segmented from the background. We use a semantic segmentation method based on DeepLabv3+ [43] for this task [44]. Moreover, all fingerprint images are rotated to an upright position and normalized to a ridge-line distance of approximately 9 pixels, which corresponds to 500 ppi live-scanned samples. The normalization avoids that the CNNs learning distinct characteristics like different ridge frequencies.

Further, we crop a patch of 100×200 pixels out of the center area of the image. This pre-processing ensures that only the Region of Interest (ROI) from within the fingertip is considered, and hence the algorithms do not learn to detect certain artifacts at the border area of a PAI or fingerprint image. Also, the fixed size ensures that the input data is equally suited for all considered algorithms. Figure 6 shows example images of after this first pre-processing step. It should be noted that the pre-processing workflow is publicly available [1].

SpoofBuster

For the SpoofBuster algorithm, the samples are pre-processed using the contactless fingerprint pre-processing proposed in [14]. Since the ISPFD database contains unsegmented and un-rotated finger photos, the fingerprint region of interest is segmented by the same method as introduced before. The segmented finger image is then rotated to an upright position. All other databases already provide segmented and rotated fingerprint images, so this step is omitted.

The segmented data is then converted to grayscale, and a Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to emphasize the ridge-line characteristics. The CLAHE algorithm is iteratively applied with a decreasing size of tile grids, starting at a window-size of 64 pixels and ending at a grid-size of 8. First, this process equalizes the brightness throughout the fingerprint region and second, it emphasizes the ridge pattern. Next, the fingerprint samples are normalized to a fixed ridge-line frequency of approximately 9 pixels, which aligns to approximately 500 ppi live-scanned fingerprints. It should be noted that in general, no samples are manually pre-processed or discarded from the experiment due to erratic pre-processing.

4.3 Score Level Fusion

The fusion of scores of several biometric systems is a promising approach to improve the overall system's accuracy. This also applies to PAD methods, where typically features or results of various algorithms are fused [45], [46]. In this work, we conduct a normalized score level fusion across the considered algorithms , like discussed in [47] and standardized in [48]. This is required for two reasons: First, the considered algorithms do not operate in the same interval and second, not all algorithms use the score range to full extent. This process consists of two steps: First, the scores of the algorithms to fuse are normalized to a closed interval $[0, 1]$. Here, we adjust all scores so that their average is 0.5 and then multiply the all scores with a factor so that the full range between 0 and 1 is used. This ensures that all algorithms are equally weighted during the fusion. It should be noted, that score fusion in general increases the computational complexity and hence, individual algorithms should be favored if possible.

4.4 Evaluation Protocols and Metrics

For our research, we consider three stages of generalizability evaluations: baseline protocols, LOO protocols and cross-database evaluations. The baseline protocol, includes all PAI species in the training and validation sets. It randomly splits the samples for each PAI species into training (30%), validation (20%), and test (50%) partitions. This partitioning ensures that as many samples as possible are included in the test partition while providing enough samples for training the CNNs. This setup provides the most realistic results. It should be noted that it is assumed that larger training partitions lead to more robust results. The non-overlapping partitions ensure that PAD algorithms are tested on samples that are not considered during training and validation, and thus guarantee a fair evaluation.

The more advanced LOO protocols to analyze the PAD performance in the presence of unknown attacks. They do not split samples of each PAI, but rather split PAIs into training, validation, and testing. With this method, we can evaluate the generalization capabilities to PAIs that have not been seen during the training. Since the COLFISPOOF includes 72 PAI species, similar materials but different colors are considered as one LOO group. In total, four LOO protocols are created: printouts, transparent overlays, default color materials and colored silicone. It should be noted that due to the availability of different colors, the number of samples for every LOO group is different. The exact number of samples can be seen in Table 3 For more detailed information, the reader is referred to the original COLFISPOOF paper [26].

The cross-database evaluation represents the most challenging protocol, which benchmarks the generalization capabilities to new high-quality PAs and a different capturing scenario. Here, we train and validate on the entire COLFISPOOF database in combination with one of the bona fide databases (60% train, 40% validation) and test on the UniCa-HDA database. With this experiment, the generalization capabilities are analyzed in two ways: first, the generalization to a new type of PA and second, the generalization from PAs contained in the COLFISPOOF database to those from a newly captured database. In particular, this is interesting, since the PAs in the COLFISPOOF database were generated based on synthetic ridge line patterns, while in the newly captured database, PAs are generated from real fingerprints. This experiment is referred to as cross-database experiment.

We use the in ISO/IEC 30107-3 [49] standardized APCER vs. BPCER metric and the D-EER metric to report the results of our experiments. To make our work comparable to others, we fix the BPCER at an operation point of 1% and report the corresponding

1. https://github.com/dasec/COLFISPOOF

TABLE 4: Overview of APCERs for a fixed BPCER of 1% and D-EERs obtained from the baseline and LOO experiments.

	A-PBS	LMFD	ResNet18	SpoofBuster	HDA		ISPFDv1		ISPFDv2	
					APCER	D-EER	APCER	D-EER	APCER	D-EER
Baseline	X				0.36	0.56	**0.0**	0.11	**0.0**	0.02
		X			0.22	1.2	0.03	0.36	**0.0**	0.08
			X		0.89	0.92	0.08	0.23	**0.0**	0.22
				X	5.28	1.86	0.03	0.64	0.83	0.92
	X	X			0.33	0.56	**0.0**	0.09	**0.0**	0.02
	X		X		0.47	0.7	**0.0**	0.11	**0.0**	0.02
	X			X	0.19	0.3	**0.0**	0.04	**0.0**	0.02
		X	X		0.42	0.71	0.03	0.19	**0.0**	0.08
		X		X	0.17	**0.26**	0.03	0.06	**0.0**	0.04
			X	X	**0.0**	0.31	**0.0**	0.04	**0.0**	0.06
	X	X	X		0.39	0.54	**0.0**	0.15	**0.0**	0.02
	X	X		X	0.08	0.31	**0.0**	0.04	**0.0**	**0.0**
	X		X	X	0.03	0.31	**0.0**	**0.0**	**0.0**	**0.0**
		X	X	X	0.11	0.31	**0.0**	0.04	**0.0**	**0.0**
LOO colored silicone	X				0.06	0.2	**0.0**	0.06	**0.0**	0.01
		X			0.06	1.03	0.03	0.36	**0.0**	0.76
			X		1.33	1.25	0.08	0.29	0.28	0.5
				X	0.17	0.32	0.08	0.68	**0.0**	0.39
	X	X			0.11	0.2	**0.0**	0.06	**0.0**	0.27
	X		X		0.06	0.38	**0.0**	0.09	**0.0**	0.1
	X			X	0.0	**0.0**	**0.0**	0.06	**0.0**	**0.0**
		X	X		1.39	1.28	0.03	0.23	0.17	0.44
		X		X	**0.0**	0.24	0.03	0.04	**0.0**	0.22
			X	X	**0.0**	0.06	**0.0**	0.09	**0.0**	0.06
	X	X	X		0.06	0.38	**0.0**	0.11	**0.0**	0.22
	X	X		X	**0.0**	**0.0**	**0.0**	0.04	**0.0**	0.05
	X		X	X	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	0.01
		X	X	X	**0.0**	0.03	**0.0**	0.04	**0.0**	0.17
LOO default color	X				2.27	1.44	0.43	0.84	**0.0**	0.05
		X			1.24	2.17	**0.0**	2.04	**0.0**	1.21
			X		3.08	1.81	0.43	0.69	0.3	0.6
				X	2.41	1.85	**0.0**	0.54	**0.0**	0.43
	X	X			2.11	1.42	0.92	0.92	**0.0**	0.22
	X		X		1.32	1.1	0.19	0.49	**0.0**	0.02
	X			X	0.05	0.3	**0.0**	0.11	**0.0**	**0.0**
		X	X		1.54	1.27	0.76	0.87	0.08	0.43
		X		X	1.11	1.17	**0.0**	0.38	**0.0**	0.24
			X	X	0.19	0.3	**0.0**	**0.0**	**0.0**	0.05
	X	X	X		1.11	1.1	0.22	0.84	**0.0**	0.22
	X	X		X	0.05	0.36	**0.0**	0.33	**0.0**	0.02
	X		X	X	**0.0**	**0.29**	**0.0**	0.11	**0.0**	**0.0**
		X	X	X	0.27	0.37	**0.0**	0.3	**0.0**	0.16
LOO printout	X				2.5	1.83	0.25	0.58	**0.0**	**0.08**
		X			1.92	1.66	0.08	0.86	**0.0**	0.14
			X		5.25	1.66	0.67	0.68	0.33	0.42
				X	14.5	2.49	0.33	0.68	3.92	2.42
	X	X			2.75	1.44	0.17	0.56	**0.0**	0.08
	X		X		2.08	1.27	0.17	0.58	**0.0**	0.1
	X			X	**0.0**	**0.28**	0.33	0.43	0.25	0.42
		X	X		1.67	1.27	0.33	0.5	**0.0**	0.26
		X		X	1.0	0.97	**0.0**	**0.0**	**0.0**	0.17
			X	X	**0.0**	0.32	0.25	0.58	0.33	0.34
	X	X	X		1.17	1.09	0.17	0.56	**0.0**	**0.08**
	X	X		X	0.08	0.36	**0.0**	0.4	**0.0**	**0.08**
	X		X	X	0.08	0.32	0.25	0.24	0.25	0.24
		X	X	X	0.08	0.64	0.17	0.24	**0.0**	0.17
LOO transparent	X				18.8	6.4	**0.0**	0.18	**0.0**	0.05
		X			18.6	9.15	**0.0**	0.44	**0.0**	0.58
			X		26.4	7.74	0.2	0.76	1.2	1.2
				X	15.6	3.75	**0.0**	0.64	0.4	0.6
	X	X			20.8	6.02	**0.0**	0.05	**0.0**	0.18
	X		X		17.8	5.44	**0.0**	0.4	**0.0**	0.04
	X			X	1.6	1.47	**0.0**	**0.0**	**0.0**	**0.0**
		X	X		25.2	7.55	**0.0**	0.37	0.2	0.6
		X		X	4.8	2.18	**0.0**	**0.0**	**0.0**	0.36
			X	X	3.0	2.18	**0.0**	**0.0**	**0.0**	0.16
	X	X	X		21.6	5.82	**0.0**	0.22	**0.0**	0.18
	X	X		X	**0.6**	**0.97**	**0.0**	**0.0**	**0.0**	0.03
	X		X	X	2.8	1.57	**0.0**	**0.0**	**0.0**	**0.0**
		X	X	X	6.4	1.47	**0.0**	**0.0**	**0.0**	0.2

APCER. The D-EER indicates the lowest detection error made by an algorithm without considering any specific operation point.

5 RESULTS

In this section, we present the results of the experiments conducted and discuss our findings in more detail. Table 4 gives an overview on APCERs at a BPCER of 1% and D-EERs, whereas the Figures 7 – 10 show the corresponding DET plots. First, we discuss the results obtained from the individual algorithms, the second part evaluates the results of our score level fusions and the third part presents the obtained cross-database results.

5.1 Baseline and LOO Protocols

The results of our previous COLFIPAD benchmark [25] show that A-PBS, LMFD and ResNet18 perform best, whereas the other tested algorithms show less accurate results. Furthermore, it can be summarized that the experiments on the HDA database show inferior results compared to the experiments conducted on the ISPFD databases. This is due to the low number of available training data. The three best-performers also show, in general, good generalizability to new PAIs. Here, the results of the four defined LOO-protocols showcase that algorithms are able to generalize well across various PAI species.

From the PAD scores in Table 4, we can observe that all algorithms have at least a few misclassified BPs. For both ISPFD databases, we observe that some samples are misclassified by every method. These samples suffer from erratic capture or pre-processing. Namely, few images contain compression artifacts, whereas others were not segmented accurately. Figure 9a – c show examples of these insufficiencies. It should be noted that these samples should not occur in a real-world process where capturing and pre-processing are calibrated to each other.

Further, we discuss the detailed results of the considered evaluation protocols by analyzing individual misclassified samples. For the baseline protocol, A-PBS and LMFD show the highest possible detection rate. All PAIs are correctly classified, whereas only the aforementioned erratic BPs are wrongly classified as PAIs. ResNet18 tends to struggle to separate BPs from PAs in the baseline experiments. Here, the average PAD scores are closer together compared to A-PBS and LMFD.

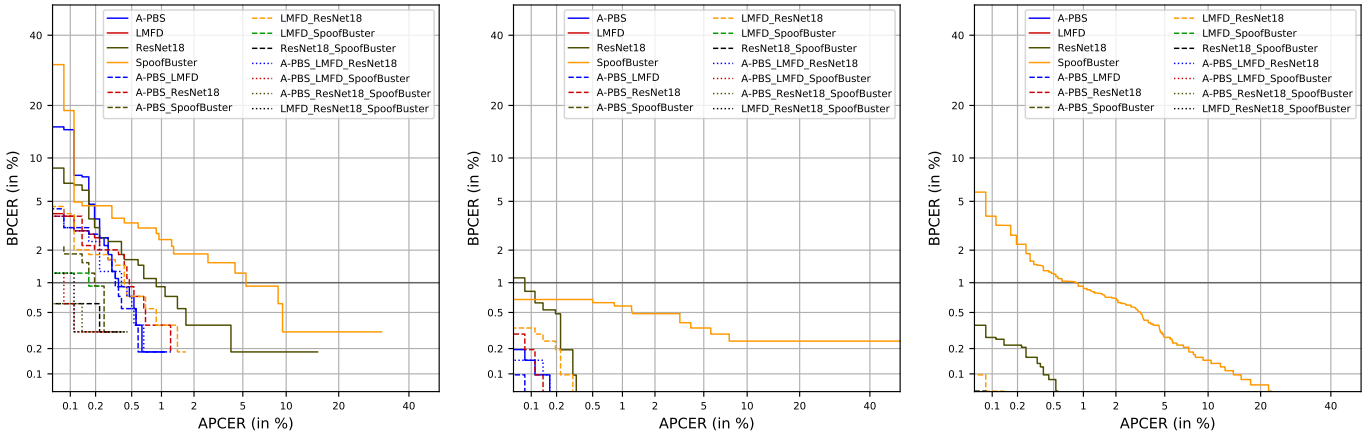
A general observation on all LOO experiments with the A-PBS algorithm is that the vast majority of PAs are correctly classified with high confidence, whereas the algorithm tends to struggle with BPs. LMFD and ResNet18 do not have this disbalance. Furthermore, a weakness of A-PBS can be identified in the LOO group default color. Here, especially dark purple playdoh (*c.f.* Figure 9d) is prone to misclassification. This PAI is also challenging for LMFD and ResNet18. LMFD also tends to misclassify pink sillyputty whereas ResNet18 also has trouble with playdoh of other colors, *e.g.* teal and pink.

Another finding is that A-PBS and ResNet18 do not accurately detect printout attacks using SynCoLFinGer PAIs (*c.f.* Figure 9e) whereas LMFD is more stable against these types of attacks. For the LOO group transparent overlays, no general trend can be identified. Both, A-PBS and LMFD robustly identify all PAs but struggle to correctly classify BPs. In this experiment, ResNet18 shows inferior performance. From this, it can be concluded that the algorithms are failing to generalize well to PAIs that have a similar color, like BPs (printout SynCoLFinGer, transparent overlays) if only different colored PAIs are included in the training set.

For the LOO group colored silicone, A-PBS works very robustly, whereas LMFD fails to classify several materials (*e.g.* darkred drangonskin, orange dark brown dragonskin) especially on the ISPFDv2 database. Again, ResNet18 shows inferior detection performance and misclassifies various PAIs.

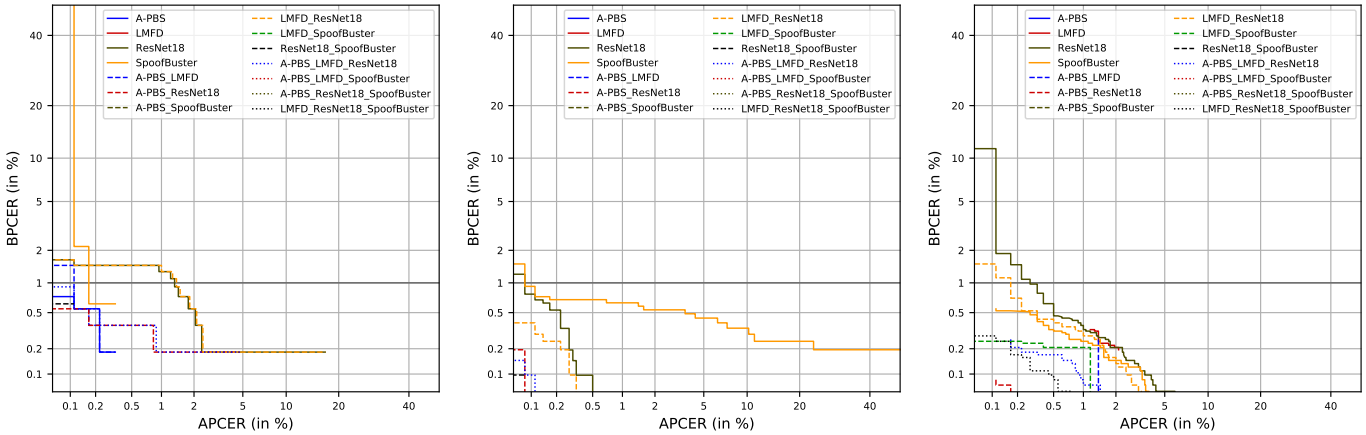
Unless discussed exceptions, no significant difference in terms of PAD scores can be observed between the two ISPFD databases. However, it can be seen that the experiments conducted on the ISPFDv2 database show better results, which could be caused by a more homogenous or larger database.

Compared to the COLFIPAD results, the SpoofBuster method shows inferior results for the baseline protocol, but performs on



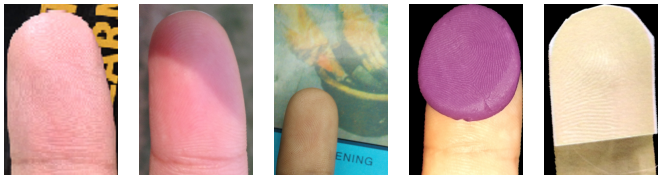
(a) HDA (b) ISPFDv1 (c) ISPFDv2

Fig. 7: DET curves obtained on the considered databases using the baseline evaluation protocol.



(a) HDA (b) ISPFDv1 (c) ISPFDv2

Fig. 8: DET curves obtained on the considered databases using the LOO protocol colored silicone.



(a) JPG artifacts (b) environment (c) segmentation (d) playdoh purple dark (e) printout

Fig. 9: Examples of misclassified samples: (a – c) BPs, (d – e) PAs.

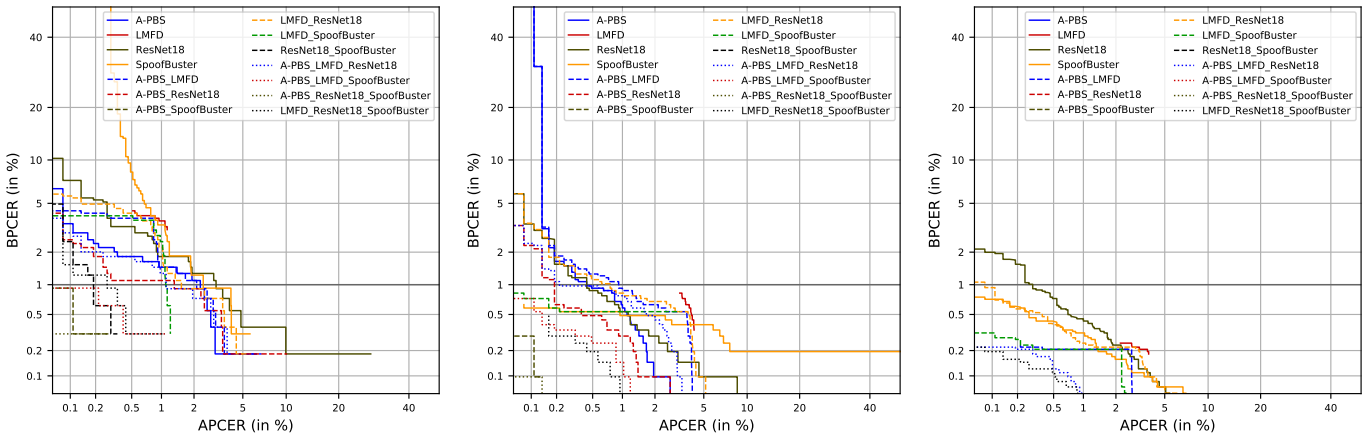
average equally well for the LOO protocols. The better generalization capabilities are explained by the fact that patches of grayscale images are used and thus no color bias is learned. One exception is the LOO protocol printout. Here, the detailed structure of the print and the paper might get lost during the pre-processing, which makes the detection more challenging. Despite the patch-based process and hence various training patches per sample, it is not observable that SpoofBuster is more robust against a low number of training samples, as is the case for the HDA database.

The findings obtained from our previous works are summarized as follows:

- **A-PBS** overall performs best and precisely detects PAIs in almost all scenarios, but has trouble robustly classifying all BPs.
- **LMFD** accurately separates BPs from PAs in many scenarios, and is particularly good at classifying BPs. However, it fails to detect PAIs in some rare cases.
- **ResNet18** as a general-purpose CNN is a viable alternative to the aforementioned algorithms. It shows only slightly decreased detection performance and generalizes well to unseen PAIs.
- **SpoofBuster** represents a vital alternative to the aforementioned methods. The detection performance is worse for the baseline and loo printout protocol, but on par with the COLFIPAD algorithms for the other loo protocols.

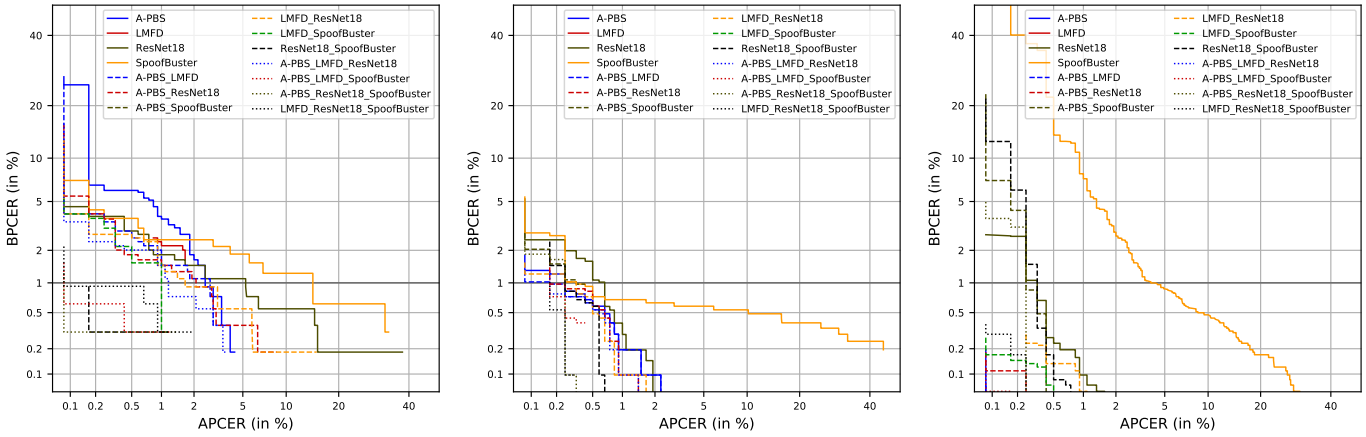
5.2 Score Fusion

We conduct the score fusion of two and three algorithms for all conducted protocols. From the results, we can observe that the fusion contributes to a more robust detection of PAIs, especially if a small number of training data samples are available. For the large



(a) HDA (b) ISPFDv1 (c) ISPFDv2

Fig. 10: DET curves obtained on the considered databases using the LOO protocol default color.



(a) HDA (b) ISPFDv1 (c) ISPFDv2

Fig. 11: DET curves obtained on the considered databases using the LOO protocol printout.

ISPFDv2 database, the APCER a BPCER of 1% is already 0% for A-PBS. However, we see that the D-EER is further reduced due to fusion. This trend amplifies if less training data is available, such as for the ISPFDv1 and HDA databases. Especially the improvement for the HDA database and the challenging LOO protocols printout and transparent is more significant. Most notably, a fusion of SpoofBuster with the color-based CNNs leads to the best improvement, as Table 4 indicates. Here, different approaches lead to more robust detection results across all protocols. Furthermore, in most cases, the fusion of the three best-performing algorithms outperforms the fusion of only two.

The results obtained from the ISPFDv2 database also indicate that a fusion might be more robust against error-prone samples such as the ones discussed above and presented in Figure 9. Here, none of the algorithms were able to classify these challenging samples entirely correct. However, as the fusion results show, the fusion method is able to perfectly separate all samples (*c.f.* baseline protocol).

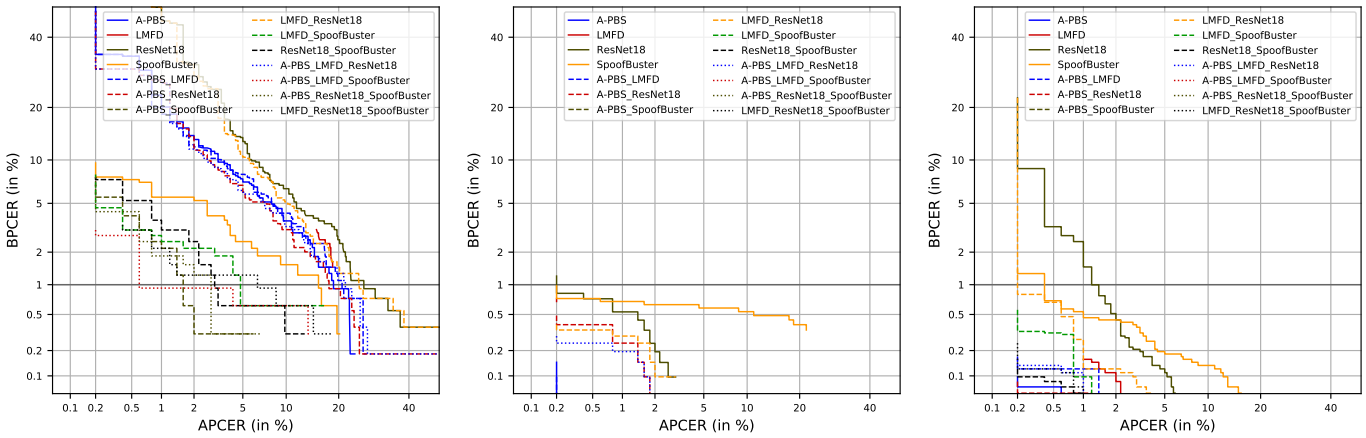
5.3 Cross Database Generalization

In addition to the LOO experiments, we analyze the generalization capabilities of the algorithms and their fusions to new scenarios in a cross-database scenario. As introduced in Section 4, we train the algorithms on COLFISPOOF in combination with one of the bona fide databases and test them on the new UniCa-HDA database. Figure 13 presents the obtained DET plots, whereas Table 5 shows the APCER for a BPCER of 5% and the D-EERs.

The obtained results show good generalization capabilities for the HDA database but also highlight limitations for both

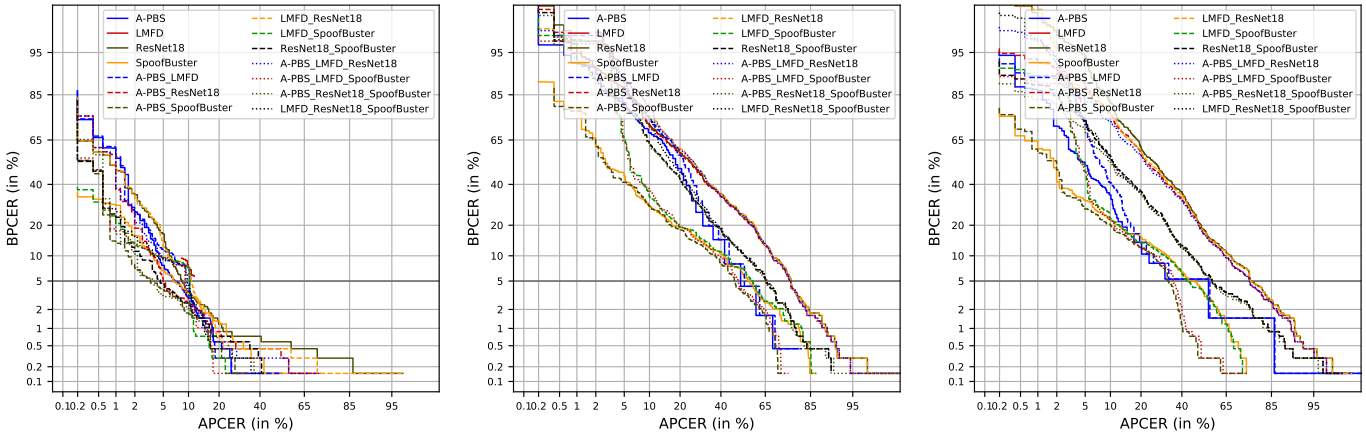
TABLE 5: Overview of APCERs for a fixed BPCER of 5% and D-EERs obtained from the baseline and LOO experiments.

	A-PBS	LMFD	ResNet18	SpoofBuster	HDA		ISPFDv1		ISPFDv2	
					APCER	D-EER	APCER	D-EER	APCER	D-EER
	X				6.75	5.99	**42.06**	26.28	**30.95**	15.82
		X			11.51	9.12	76.45	48.92	74.72	43.44
			X		8.33	7.18	77.38	37.08	76.79	37.50
				X	6.75	5.99	56.15	19.42	45.63	16.65
	X	X			10.32	7.94	44.64	29.68	32.94	17.01
	X		X		4.96	4.98	77.18	36.9	76.39	35.72
Cross	X			X	**3.37**	4.22	47.62	**18.66**	32.34	**15.04**
database			X		10.71	8.36	77.38	37.33	76.79	36.90
		X		X	10.32	8.36	57.34	19.68	42.66	16.30
			X	X	4.56	4.73	65.48	28.80	57.34	26.43
	X	X	X		9.33	7.94	77.18	37.08	76.39	35.80
	X	X		X	9.13	7.94	49.6	19.68	35.12	15.12
	X		X	X	3.77	**4.14**	65.28	28.80	57.14	25.76
		X	X	X	9.72	8.11	65.87	29.74	57.34	26.01



(a) HDA (b) ISPFDv1 (c) ISPFDv2

Fig. 12: DET curves obtained on the considered databases using the LOO protocol transparent overlay.



(a) HDA (b) ISPFDv1 (c) ISPFDv2

Fig. 13: DET curves obtained on the considered databases in the cross-database experiment.

ISPFD databases. For the HDA database, we see that A-PBS and SpoofBuster perform best with a similar D-EER or 5.99% whereas LMFD and ResNet18 show slightly inferior results. Taking the fusion result into account, a combination of A-PBS, ResNet18 and SpoofBuster reduces the D-EER to 4.14%.

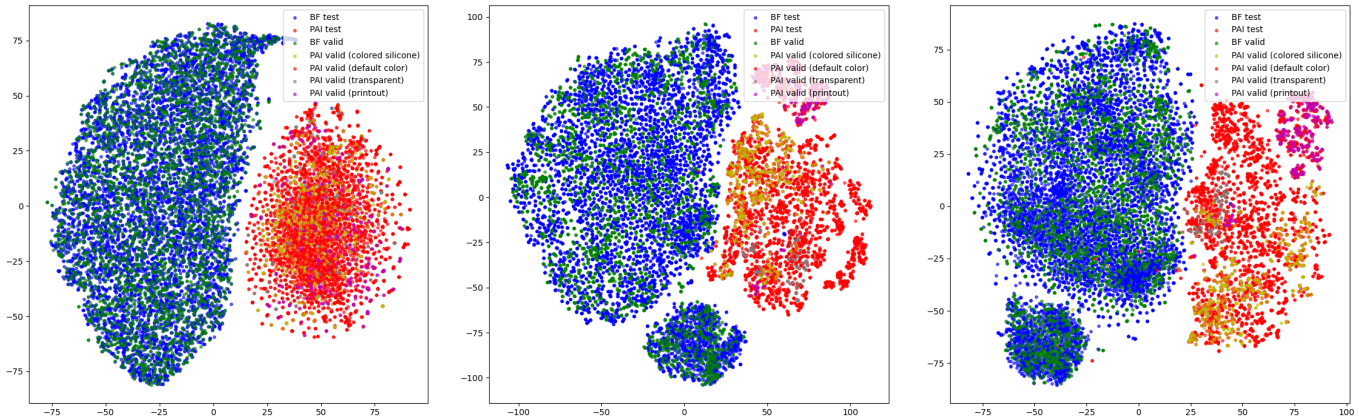
In contrast to the aforementioned scenario, none of the CNNs is able to generalize well from ISPFDv1 or ISPFDv2 to the UniCa-HDA test database. Here, the majority of BPs are classified as PAs which results in very high APCERs. Consequently, fusion strategies are not very beneficial in this case.

5.4 Explainability Results

In addition to our DET curves, we present t-SNE plots for two representative experiments. All presented plots include the validation and the test set in order to assess how the trained models deviate from the test data.

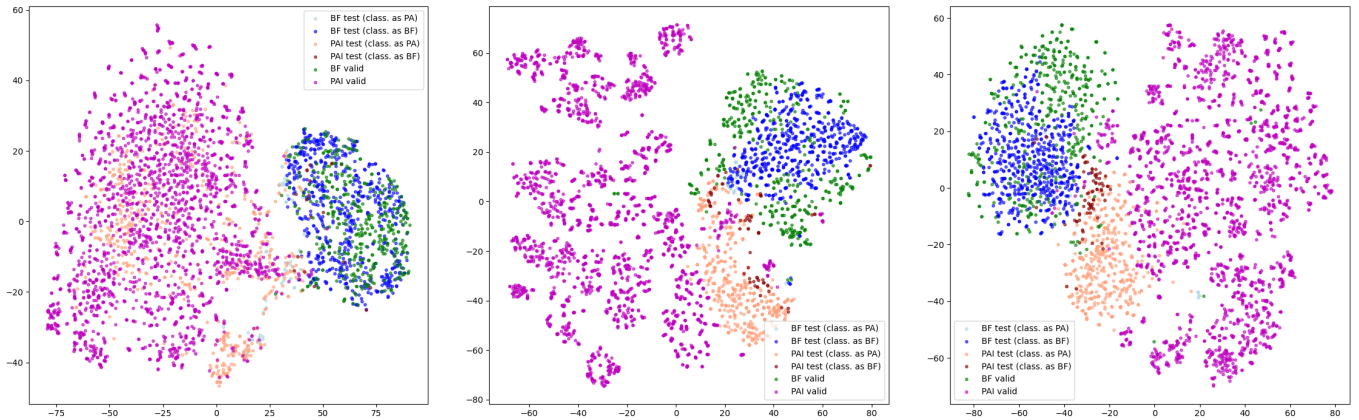
Figure 14 shows the obtained t-SNE plots obtained from the ISPFDv2 baseline experiment. For this plot, we further separate the validations set according to PAI species, like for the LOO protocols. We can observe that all plots indicate a clear separation between BPs and PAs. However, there are certain differences between A-PBS and both other algorithms. A-PBS has two disjoint areas, whereas the PAI species are not clustered together. For LMFD and ResNet18 the plots are different. Here, the BPs are separated into two clusters: a big one a rather sparse accumulation and a little cluster with a dense accumulation. These clusters might be based on the two sub-databases included in the ISPFDv2 database, the indoor and outdoor subsets. Furthermore, we also see a clustering of the different PAI species in the validation set. From these findings, we can summarize that A-PBS learns the distinct feature to separate BPs from PAs much better than the other CNNs.

Figure 15 presents t-SNE plots of the cross-database experiment, including the HDA and COLFISPOOF for training and the UniCa-HDA database for testing. The plots indicate a huge difference between the three CNNs. For A-PBS, the validation set and the test set are heavily superimposed for both the BPs and the PAs. The misclassified samples are located merely located at the transition between the BP and PA validation sets. This indicates that A-PDS generalizes reasonably well from the COLFISPOOF database to the UniCA-HDA database and to new capturing scenarios. For both other CNNs, the t-SNE plots indicate subordinate generalization capabilities. We see the PAI test data is clearly separated from the training data and much closer to the BPs.



(a) A-PBS (b) LMFD (c) ResNet18

Fig. 14: t-SNE plots obtained from the ISPFDv2 baseline experiments.



(a) A-PBS (b) LMFD (c) ResNet18

Fig. 15: t-SNE plots obtained from the UniCa-HDA cross database experiments.

Furthermore, the BP in the test partition is not fully superimposed by the validation data but is located around it. This indicates that both algorithms are not able to generalize well from the COLFISPOOF database to the UniCa-HDA database. However, it should be noted that despite this, both CNNs can classify a large share of the test samples correctly.

5.5 Summary and Discussion

Our experimental results show the great potential of CNN-based PAD for contactless fingerprint recognition. CNNs detect PAs they are trained on with very low error rates. Furthermore, they are able to generalize to new materials and capturing environments to a high extent.

As motivated in Section 3, a PAD mechanism might be exposed to a wide variety of possible PAs, different environmental scenarios and capturing workflows. Our results also showcase challenges and limitations. General-purpose CNNs like the tested ResNet18 perform considerably well for PAs they are trained on and generalize to a certain extent to new PAs. However, they might have severe limitations when it comes to cross-database generalization scenarios. Also, specialized PAD CNNs like the tested LMFD algorithm show limitations when it comes to cross-database generalizability.

Omitting distinct information, like converting samples from color to grayscale, does not have to be detrimental, as our SpoofBuster experiments demonstrate. This algorithm shows inferior results in our baseline and some LOO scenarios, but performs considerably well for the more challenging cross-database experiments. It is assumed that the limited color information, together with the more elaborated method, leads to good generalizability.

Score fusion leads to a more robust PAD mechanism. In all experiments conducted, fusions result in the lowest error rates. However, the algorithms considered for fusion should be selected carefully to improve the detection performance. Especially for the cross-database scenarios, a fusion of different approaches, like in our case, A-PBS and SpoofBuster shows the best results.

Generalization is a strong requirement in order to implement PAD in large-scale systems and unconstrained environmental scenarios. We have shown that the considered algorithms generalize reasonably well to new capturing scenarios and unseen PAI species. However, we also clearly identified limitations when new capturing workflows are used to acquire the samples. Here, the capturing workflow including segmentation, normalization and quality control, might have a major impact on the samples presented to the PAD algorithm.

Explainability methods like the presented t-SNE plots are

beneficial to assess the generalization capabilities of a method. We showcased that there are severe differences in the t-SNE representations of A-PBS, LMFD and ResNet18, which could also indicate general generalization capabilities. In general, explainability tools should be considered to acquire a more profound understanding of the learned features and potential biases. Here, our approach of including the validation set and analyzing the deviation from the test set is suggested for a significant analysis.

6 Conclusion

For most biometric systems, PAD is crucial to preserving operational security. Especially, capturing processes that operate without any contact between the capturing device and the biometric trait offer severe challenges for robust and efficient PAD.

In this paper, we address the research area of contactless fingerprint PAD from an operational point of view. First, we discuss distinct challenges and promising approaches for contactless fingerprint PAD. In the second part, we select four CNN-based contactless fingerprint PAD methods and conduct a comprehensive and comparative evaluation. Here, we considered the COLFISPOOF database in combination with an HDA and the publicly available ISPFDv1 and ISPFDv2 databases to evaluate the PAD performance in baseline and more advanced LOO experiments. Furthermore, we evaluate the cross-database generalization capabilities to unseen PAI species and new capturing scenarios using a newly captured UniCa-HDA database. For all experiments, we discuss the results for the individual algorithms and algorithm fusion based on score level. During an explainability evaluation, we also illustrate the capabilities of t-SNE plots for a more profound understanding of the obtained results.

The obtained results show a good performance for many experiments but also highlight the limitations of the considered methods. We conclude that the algorithms should be precisely designed and adjusted to the application and environmental scenario. The tested algorithms are able to generalize to new PAI species and environmental situations, but they cannot generalize to new capturing devices. Moreover, we conclude that the score fusion is a vital tool to establish a more robust PAD mechanism. Especially, a fusion of different approaches could lead to more robust detection results.

Our work opens up various new possible research directions. *E.g*. a fusion of databases captured using different devices and environmental setups could be explored to study if the obtained model leads to better generalization capabilities. Instead of carrying out an algorithm fusion strategy, a dual-channel CNN *e.g*. , operating on enhanced grayscale and color images, could solve the detection task more, efficient and accurate. Furthermore, additional databases focusing on various capturing workflows, environments, PAIs and a wide variety of subjects would lead to a more profound assessment of the algorithm's generalization capability.

Acknowledgments

This research work has been funded by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[1] C. Stein, V. Bouatou, and C. Busch, "Video-based fingerphoto recognition with anti-spoofing techniques with smartphone cameras," in *International Conference of the Biometric Special Interest Group (BIOSIG)*, 2013, pp. 1–12.

[2] A. Taneja, A. Tayal, A. Malhorta, A. Sankaran, M. Vatsa, and R. Singh, "Fingerphoto spoofing in mobile devices: a preliminary study," in *IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2016, pp. 1–7.

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

[4] M. Fujio, Y. Kaga, T. Murakami, T. Ohki, and K. Takahashi, "Face/fingerphoto spoof detection under noisy conditions by using deep convolutional neural network," in *BIOSIGNALS*, 2018, pp. 54–62.

[5] P. Wasnik, R. Raghavendra, K. Raja, and C. Busch, "Presentation attack detection for smartphone based fingerphoto recognition using second order local structures," in *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 2018, pp. 241–246.

[6] E. Marasco and A. Vurity, "Late deep fusion of color spaces to enhance finger photo presentation attack detection in smartphones," *Applied Sciences*, vol. 12, no. 22, 2022.

[7] B. Adami, S. Tehranipoor, N. Nasrabadi, and N. Karimian, "A Universal Anti-Spoofing Approach for Contactless Fingerprint Biometric Systems," Oct. 2023, arXiv:2310.15044 [cs].

[8] J. Priesnitz, R. Casula, C. Rathgeb, G. L. Marcialis, and C. Busch, "Towards contactless fingerprint presentation attack detection using algorithms from the contact-based domain," in *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. IEEE, September 2023.

[9] S. Purnapatra, C. Miller-Lynch, S. Miner, Y. Liu, K. Bahmani, S. Dey, and S. Schuckers, "Presentation attack detection with advanced cnn models for noncontact-based fingerprint systems," in *2023 11th International Workshop on Biometrics and Forensics (IWBF)*. IEEE, 2023, pp. 1–6.

[10] S. Purnapatra, H. Rezaie, B. Jawade, Y. Liu, Y. Pan, L. Brosell, M. R. Sumi, L. Igene, A. Dimarco, S. Setlur, S. Dey, S. Schuckers, M. Huber, J. N. Kolf, M. Fang, N. Damer, B. Adami, R. Chitic, K. Seelert, V. Mistry, R. Parthe, and U. Kacar, "Liveness Detection Competition – Noncontact-based Fingerprint Algorithms and Systems (LivDet-2023 Noncontact Fingerprint)," Oct. 2023, arXiv:2310.00659 [cs].

[11] J. Priesnitz, C. Rathgeb, N. Buchmann, C. Busch, and M. Margraf, "An overview of touchless 2D fingerprint recognition," *EURASIP Journal on Image and Video Processing*, vol. 2021, no. 1, pp. 1–28, 2021.

[12] A. M. M. Chowdhury and M. H. Imtiaz, "Contactless fingerprint recognition using deep learning – a systematic review," *Journal of Cybersecurity and Privacy*, vol. 2, no. 3, pp. 714–730, 2022.

[13] X. Yin, Y. Zhu, and J. Hu, "A survey on 2D and 3D contactless fingerprint biometrics: A taxonomy, review, and future directions," *IEEE Open Journal of the Computer Society*, vol. 2, pp. 370–381, 2021.

[14] J. Priesnitz, R. Huesmann, C. Rathgeb, N. Buchmann, and C. Busch, "Mobile Contactless Fingerprint Recognition: Implementation, Performance and Usability Aspects," *Sensors*, vol. 22, no. 3, 2022.

[15] C. Kauba, D. Söllinger, S. Kirchgasser, A. Weissenfeld, G. Fernández Domínguez, B. Strobl, and A. Uhl, "Towards using police officers' business smartphones for contactless fingerprint acquisition and enabling fingerprint comparison against contact-based datasets," *Sensors*, vol. 21, no. 7, p. 2248, 2021.

[16] S. M. Furman, B. C. Stanton, M. F. Theofanos, J. M. Libert, and J. D. Grantham, "Contactless fingerprint devices usability test," National Institute of Standards and Technology, Gaithersburg, MD, Tech. Rep. NIST IR 8171, Mar. 2017.

[17] 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.

[18] J. Galbally, J. Fierrez, R. Cappelli, and G. L. Marcialis, "Introduction to presentation attack detection in fingerprint biometrics," in *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*. Springer, 2023, pp. 3–15.

[19] J. Kolberg, D. Gläsner, R. Breithaupt, M. Gomez-Barrero, J. Reinhold, A. von Twickel, and C. Busch, "On the effectiveness of impedance-based fingerprint presentation attack detection," *Sensors*, vol. 21, no. 17, 2021.

[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, pp. 211–252, 2015.

[21] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 740–755.

[22] D. Yambay, B. Becker, N. Kohli, D. Yadav, A. Czajka, K. W. Bowyer, S. Schuckers, R. Singh, M. Vatsa, A. Noore *et al.*, "LivDet iris 2017—iris liveness detection competition 2017," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2017, pp. 733–741.

[23] M. Micheletto, G. Orrù, R. Casula, D. Yambay, G. L. Marcialis, and S. Schuckers, "Review of the fingerprint liveness detection (livdet) competition series: from 2009 to 2021," *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*, pp. 57–76, 2023.

[24] A. Boyd, J. Speth, L. Parzianello, K. W. Bowyer, and A. Czajka, "Comprehensive study in open-set iris presentation attack detection," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 3238–3250, 2023.

[25] J. Priesnitz, J. Kolberg, M. Fang, A. Madhu, C. Rathgeb, N. Damer, and C. Busch, "COLFIPAD: A presentation attack detection benchmark for contactless fingerprint recognition," in *2023 IEEE International Joint Conference on Biometrics (IJCB)*, 2023.

[26] J. Kolberg, J. Priesnitz, C. Rathgeb, and C. Busch, "COLFISPOOF: A new database for contactless fingerprint presentation attack detection research," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 653–661.

[27] K. Karampidis, M. Rousouliotis, E. Linardos, and E. Kavallieratou, "A comprehensive survey of fingerprint presentation attack detection," *Journal of Surveillance, Security and Safety*, vol. 2, no. 4, pp. 117–161, 2021.

[28] A. Vrachnos, E. Papadaki, P. Lagou, N. Soumelidis, and E. Papamichail, *Remote ID Proofing good Practices*, European Union Agency for Cybersecurity, 2024.

[29] T. Chugh and A. K. Jain, "Fingerprint spoof detector generalization," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 42–55, 2020.

[30] A. Malhotra, S. Chhabra, M. Vatsa, and R. Singh, "On privacy preserving anonymization of finger-selfies," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 26–27.

[31] E. Marasco and A. Vurity, "Fingerphoto presentation attack detection: Generalization in smartphones," in *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 4518–4523.

[32] R. Tolosana, M. Gomez-Barrero, J. Kolberg, A. Morales, C. Busch, and J. Ortega-Garcia, "Towards fingerprint presentation attack detection based on convolutional neural networks and short wave infrared imaging," in *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2018, pp. 1–5.

[33] M. E. Hussein, L. Spinoulas, F. Xiong, and W. Abd-Almageed, "Fingerprint presentation attack detection using a novel multi-spectral capture device and patch-based convolutional neural networks," in *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2018, pp. 1–8.

[34] F. Liu, H. Liu, W. Zhang, G. Liu, and L. Shen, "One-class fingerprint presentation attack detection using auto-encoder network," *IEEE Transactions on Image Processing*, vol. 30, pp. 2394–2407, 2021.

[35] A. Kirfel, T. Scheer, N. Jung, and C. Busch, "Robust identification and segmentation of the outer skin layers in volumetric fingerprint data," *Sensors*, vol. 22, no. 21, p. 8229, 2022.

[36] M. Fang, N. Damer, F. Boutros, F. Kirchbuchner, and A. Kuijper, "Iris presentation attack detection by attention-based and deep pixel-wise binary supervision network," in *IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2021, pp. 1–8.

[37] M. Fang, N. Damer, F. Kirchbuchner, and A. Kuijper, "Learnable multi-level frequency decomposition and hierarchical attention mechanism for generalized face presentation attack detection," in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 3722–3731.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[39] T. Chugh and A. K. Jain, "Fingerprint presentation attack detection: Generalization and efficiency," in *2019 International Conference on Biometrics (ICB)*, 2019, pp. 1–8.

[40] C. I. Watson, M. D. Garris, E. Tabassi, C. L. Wilson, R. M. McCabe, S. Janet, and K. Ko, "User's guide to NIST biometric image software (nbis)," 2007.

[41] A. Sankaran, A. Malhotra, A. Mittal, M. Vatsa, and R. Singh, "On smartphone camera based fingerphoto authentication," in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2015, pp. 1–7.

[42] A. Malhotra, A. Sankaran, M. Vatsa, and R. Singh, "On matching finger-selfies using deep scattering networks," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 4, pp. 350–362, 2020.

[43] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[44] J. Priesnitz, C. Rathgeb, N. Buchmann, and C. Busch, "Deep Learning-Based Semantic Segmentation for Touchless Fingerprint Recognition," in *Proc. Intl. Conf. Pattern Recognition (ICPR) (Workshops)*, 2021.

[45] N. Poh and S. Bengio, "Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication," *Pattern Recognition*, vol. 39, no. 2, pp. 223–233, 2006.

[46] D. Yadav, N. Kohli, A. Agarwal, M. Vatsa, R. Singh, and A. Noore, "Fusion of handcrafted and deep learning features for large-scale multiple iris presentation attack detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 572–579.

[47] A. Jain, B. Klare, and A. Ross, "Guidelines for best practices in biometrics research," in *2015 International Conference on Biometrics (ICB)*. IEEE, 2015, pp. 541–545.

[48] ISO/IEC JTC1 SC37 Biometrics, "International standards ISO/IEC TR 24722, multimodal and other multibiometric fusion," International Organization for Standardisation, Tech. Rep., 2015.

[49] ——, *ISO/IEC 30107-3. Information Technology - Biometric presentation attack detection - Part 3: Testing and Reporting*, International Organization for Standardization, 2023.