Anomaly Detection With Convolutional Autoencoders for Fingerprint Presentation Attack Detection

Jascha Kolberg[iD], *Graduate Student Member, IEEE*, Marcel Grimmer, Marta Gomez-Barrero[iD], *Member, IEEE*, and Christoph Busch[iD]

Abstract—In recent years, the popularity of fingerprint-based biometric authentication systems significantly increased. However, together with many advantages, biometric systems are still vulnerable to presentation attacks (PAs). In particular, this applies for unsupervised applications, where new attacks unknown to the system operator may occur. Therefore, presentation attack detection (PAD) methods are used to determine whether samples stem from a bona fide subject or from a presentation attack instrument (PAI). In this context, most works are dedicated to solve PAD as a two-class classification problem, which includes training a model on both bona fide and PA samples. In spite of the good detection rates reported, these methods still face difficulties detecting PAIs from unknown materials. To address this issue, we propose a new PAD technique based on autoencoders (AEs) trained only on bona fide samples (i.e., one-class), which are captured in the short wave infrared domain. On the experimental evaluation over a database of 19,711 bona fide and 4,339 PA images including 45 different PAI species, a detection equal error rate (D-EER) of 2.00% was achieved. Additionally, our best performing AE model is compared to further one-class classifiers (support vector machine, Gaussian mixture model). The results show the effectiveness of the AE model as it significantly outperforms the previously proposed methods.

Index Terms—Fingerprint recognition, presentation attack detection, one-class classifier, autoencoder, anomaly detection.

I. INTRODUCTION

NOWADAYS, we encounter biometric recognition systems in many places of our daily life. Applications range from high security border control to user convenient smartphone unlocking. Especially fingerprint recognition systems are long established and widely used [1].

However, biometric systems can be affected by external attacks as the capture device is exposed to the public. Those presentation attacks (PAs) are defined within ISO/IEC 30107-1 [2] as a "presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system". During execution, a presentation attack instrument (PAI), e.g., a fingerprint overlay, can be used to either impersonate someone else (i.e., impostor) or to avoid being recognised (i.e., identity concealer). Summarising, the artefact that is used for a presentation attack is called PAI while different material combinations or recipies result in different PAI species. As a consequence, biometric systems require automated presentation attack detection (PAD) modules in order to distinguish bona fide presentations from attack presentations [3].

Since the periodic LivDet competitions started in 2009 for fingerprint [4] and in 2013 for iris [5], PAD in general has attracted a lot of research. In parallel to those research efforts, more and more different materials are found or combined to create new species [6]. On the one hand, older PAD methods might not detect new PAI species. On the other hand, it becomes much more challenging to collect diverse datasets in order to develop and evaluate (new) PAD approaches. Being a binary classification problem (bona fide vs. PA), common PAD approaches are trained on both classes and hence perform only as good as the chosen training set. In this scenario, unknown attacks [7] present only in the test set can significantly trouble the classifier, as it requires good generalisation properties that are hard to achieve. In order to avoid re-training the classifier each time a new PAI species is created, one-class classifiers can be used [8]. These models are solely trained on bona fide samples to detect anomalies in unseen data. They are especially designed to generalise much better than multi-class classifiers since all PAs are unknown to them.

In this context, we propose to involve convolutional autoencoders for unknown fingerprint PAD. We test different

architecture designs and show how the negative effect of outliers in the training set can be reduced in comparison to two-class classifiers. Finally, we benchmark the autoencoder against additional one-class classifiers to prove the soundness of our approach. The evaluation is carried out on data captured in the short wave infrared domain with over 24,000 samples, including 45 different PAI species. It should be noted, that the discussed design decisions should be generally applicable for other input data as well.

The remaining article is structured as follows: Section II summarises related work on fingerprint and one-class PAD. Our capture device is described in Section III and Section IV contains the autoencoder design and our proposed PAD method. In Section V we evaluate the experiments before Section VI concludes our findings.

II. RELATED WORK

This section reviews state-of-the-art approaches related to the contribution of this work. In the context of PAD, two different solutions exists: *i) software-based*, where a deeper analysis of the existing data for authentication is carried out, and *ii) hardware-based*, where new sensors are developed to capture additional data for PAD. Due to the high number of publications for fingerprint PAD within the last decade, we focus on hardware-based approaches in the first subsection and refer the reader to [9], [10] for comprehensive surveys. On the other hand, most classifiers are trained on both classes, hence in the second subsection we present an overview of one-class PAD for other modalities as well. In order to evaluate the vulnerabilities of biometric systems to PAs, the following metrics are defined within the ISO/IEC 30107-3 standard on biometric presentation attack detection - part 3: testing and reporting [11]:

Attack Presentation Classification Error Rate (APCER): "proportion of attack presentations using the same PAI species incorrectly classified as bona fide presentations".

Bona Fide Presentation Classification Error Rate (BPCER): "proportion of bona fide presentations incorrectly classified as attack presentations".

A. Hardware-Based Fingerprint PAD

Similar to other pattern recognition tasks, PAD benefits from information captured by additional sensors. This information is then analysed with dedicated software. To that end, an overview of hardware-based state-of-the-art fingerprint PAD methods is presented in Table I.

One of the most reliable methods for fingerprint PAD is based on optical coherence tomography (OCT) [30] sensors, which capture a 3D model of the fingertip up to two millimeter underneath the skin. In addition to PAD, this scan can be used to recover worn-out fingerprints, since it includes the inner fingerprint as well. Hence, it also reveals overlaying PAIs as well as full fake fingers. Using OCT scanners, Darlow *et al.* [15] detected double bright peaks in gelatin overlays and analysed the autocorrelation for gelatin full fingers. Their setup achieves a 100% detection accuracy on a database with 568 samples. Also Liu *et al.* [21] analyse the peaks of

OCT scans. They discover that 1D depth scans of bona fide samples contain exactly two peaks while one appears prior the maximum peak. Thus, they apply a threshold to successfully distinguish between bona fides and PAs. Training a convolutional neural network (CNN) on overlapping patches of a depth B-scan, Chugh and Jain [23] report a detection accuracy close to 100%. However, the utilised capture device does not acquire the fingerprint for biometric recognition purpose. An extensive review on OCT for fingerprint PAD is published by Moolla *et al.* [31]. It should be noted that the high costs of OCT scanners are an explicit disadvantage in contrast to other methods.

Another approach utilises different illumination sources to collect additional PAD data. Rowe *et al.* [12] developed the first multi-spectral fingerprint capture device in 2008. Their sensor captures the fingerprint in white, blue, green, and red illumination with a twofold goal: *i)* improving the recognition process, and *ii)* detection of PAIs. The authors prove the suitability of their design on a massive dataset of nearly 45,000 samples comprising 60% PAs. In a similar approach, Hengfoss *et al.* [13] analysed the reflections for all wavelengths between 400 nm and 1650 nm on the blanching effect (i.e., the finger is pressed against a surface such that the blood is squeezed out). They observe that these dynamic effects only occur for bona fide presentations and neither for PAIs nor for cadaver fingers. Additionally, they measured the pulse of the finger but conclude that it takes much longer and is less suited for PAD. Further optical methods for pulse, pressure, and skin reflections are presented by Drahansky *et al.* [14]. Their experiments show that skin reflections in the evaluated wavelengths of 470 nm, 550 nm, and 700 nm outperform the other two methods. In another approach, Kolberg *et al.* [26] visualise vein patterns by placing 940 nm LEDs above the finger. Using Gaussian pyramids, they are able to detect fingerprint PAIs since they usually do not include a vein pattern. However, for thin and transparent overlay attacks the bona fide veins still remain visible, which limits detection capabilites for overlay PAIs.

More recent publications focus on the short wave infrared (SWIR) spectrum between 900 nm and 1700 nm, which is not visible for the human eye but can be captured by adequate cameras. Gomez-Barrero *et al.* [16] utilise the spectral signature between different wavelengths for fingerprint PAD. Working with a rather small database, they show that most materials reflect the illumination in a different way than human skin. A subsequent study [17] further improves PAD performance on those 60 samples with the use of a CNN. Moreover, by fine-tuning two pre-trained CNNs and training a small residual network from scratch, Tolosana *et al.* [27] showed that deep learning approaches perform much better than spectral signatures for bigger datasets. Additionally, the results reveal that the small residual network trained from scratch outperforms the fine-tuned VGG19 and MobileNet CNNs, for user-convenient scenarios requiring a low BPCER. Another extensive benchmark [28] tests two additional CNNs and adds an advanced pre-processing layer to them. This layer is trained on the given dataset to pre-process a 4-channel SWIR image for usage in 3-channel CNNs, which

TABLE I
HARDWARE-BASED FINGERPRINT PAD METHODS WITH THEIR MOST RELEVANT METHODOLOGIES AS PERFORMANCE AND THE NUMBER OF PAI SPECIES, PA SAMPLES, AND BONA FIDE SAMPLES

Year	Ref.	Description	Performance	#PAI species	#PA samples	#BF samples
2008	[12]	Multi-spectral wavelet transform 430 nm, 530 nm, 630 nm + white light	APCER = 0.9% BPCER = 0.5%	49	27,486	17,454
2011	[13]	Multi-spectral blanching effect, pulse	APCER = 0% BPCER = 0%	4	7-15	11-28
2013	[14]	Optical methods pulse, pressure, skin reflections	APCER = 10% BPCER < 2%	N/A	N/A	N/A
2016	[15]	OCT, double bright peaks + autocorrelation	APCER = 0% BPCER = 0%	3	28	540
2018	[16]	SWIR spectral signatures + SVM	APCER = 5.7% BPCER = 0%	12	47	13
	[17]	SWIR + CNN	APCER = 0% BPCER = 0%	12	47	13
	[18]	LSCI + SVM BSIF, LBP, HOG, histogram	APCER = 15.5% BPCER = 0.2%	32	225	545
	[19]	SWIR, LSCI + patch-based CNN	APCER = 0% BPCER = 0%	17	227	551
	[20]	Weighted score fusion + SVM SWIR, LSCI, Finger vein	APCER = 6.6% BPCER = 0.2%	35	442	4,291
2019	[21]	OCT peak analysis	APCER = 0% BPCER = 0%	4	24,000	12,000
	[22]	LSCI CNN + LSTM patch-based	APCER ≤ 0.14% BPCER ≤ 0.11%	6	218	3,743
	[23]	OCT patch-CNN (sensor captures no fingerprints)	APCER = 0.27% BPCER = 0.2%	8	357	3,413
	[24]	LSCI benchmark + Fusion BSIF, LBP, HOG, histogram	APCER = 9.01% BPCER = 0.05%	35	442	4,291
	[25]	Fusion of: SWIR + CNN and LSCI + hand-crafted features	APCER ≤ 3% BPCER ≤ 0.1%	35	442	4,291
2020	[26]	Finger vein images Gaussian pyramid + SVM	APCER = 11.61% BPCER = 0.68%	32	224	542
	[27]	SWIR CNN fusion (pre-trained and from scratch)	APCER ≈ 7% BPCER = 0.1%	35	442	4,291
	[28]	SWIR CNN Fusion 4D-input pre-processing	APCER = 1.16% BPCER = 0.2%	41	3,310	8,214
	[29]	1310 nm laser sequences LSTM + CNN benchmark	APCER = 3.71% BPCER = 0.2%	45	4,339	17,730

significantly improves PAD performance in contrast to the manual pre-processing used in [27].

On the other hand, the technique of laser speckle contrast imaging (LSCI) [32] is able to visualise blood movement underneath the skin. For this purpose, a laser illuminates the desired area and a sequence (i.e., 1 second) of images is captured. Since this laser slightly penetrates the skin, subtle movements within blood tissues change the reflected speckle pattern over time [33]. Utilising this principle for fingerprint PAD, Keilbach *et al.* [18] compute the temporal contrast in order to obtain a single LSCI image for feature extraction. Those handcrafted features (e.g., LBP, BSIF) are then classified by support vector machines (SVMs). This approach was

later benchmarked in [24] with eight additional classifiers on a larger dataset in order to evaluate the best PAD performance by fusing different schemes. However, similar to the work on vein patterns, thin and transparent overlays are often wrongly classified as bona fide. In the case that the material of the PAI is thin enough for the laser to still penetrate into the skin below, bona fide properties are captured and thus the PAI is not detected. Finally, Mirzaalian *et al.* [22] applied deep learning methods on these laser sequences. Next to more traditional CNNs, the authors propose the usage of long short-term memory (LSTM) networks, which are able to remember a temporal state and can directly process the temporal information within sequences. The results show a slight advantage of the

TABLE II
ONE-CLASS PAD METHODS ACROSS DIFFERENT MODALITIES

Year	Ref.	Modality	Description
2016	[34]	Fingerprint	multiple OC-SVMs + PA refining
2018	[35]	Face	OC-SVM + OC-GMM vs. two-class versions
2019	[36]	Face	Autoencoder + multi-layer perceptron
2019	[37]	Fingerprint	Three GANs based on DCGAN architecture [38]

LSTM towards the four CNNs tested. A more extensive benchmark on LSTMs and CNNs in [29] comes to the conclusion that both temporal analysis of the LSTMs and spatial analysis of some CNNs are partly complementary and detect different PA samples.

Given the promising concepts of SWIR and LSCI data for fingerprint PAD, fusions of both approaches have been published in [19], [20], [25]. These multimodal approaches prove that PAD benefits from additional sensors. The weaknesses of one technology can be covered by another and the combination of different methods significantly improves the overall detection accuracy. Additionally, fused systems are more robust against unseen PAI species in the test set.

B. One-Class Presentation Attack Detection

Unlike traditional classification problems, the motivation behind one-class classifiers is learning the structure of data samples belonging to a single class. Therefore, in case of PAD, one-class classifiers are trained only on bona fide samples. New and unseen samples are classified as PAs if their structure differs from those bona fide samples used in the training phase. In this context, the main challenge is to find an optimal threshold to ensure that sophisticated PAs can still be distinguished from those bona fides that deviate from normality. Due to the environmental conditions and interaction factors (data subject with respect to the capture device) a significant intra-class variation for the bona fide class must be expected. Since the majority of published PAD approaches are based on two-class classification, this section reviews one-class publications across modalities as summarised in Table II. Due to the different modalities and datasets used, a comparison of performance metrics is not included.

Generally, one-class classifiers can be split into generative and non-generative approaches [35]. Generative methods aim to approximate the distribution function of the bona fides (e.g., a Gaussian model). Non-generative approaches focus on learning an optimal hypersphere that defines a decision boundary to separate bona fides from PAs.

One non-generative fingerprint PAD approach has been presented by Ding and Ross [34], who introduced an ensemble of multiple one-class support vector machine (OC-SVM) classifiers, each of which is trained on different feature sets.

The main goal of all OC-SVMs is to find the smallest possible hypersphere around the majority of training samples. Once the boundaries of the hyperspheres are found, they are refined using a small number of PA samples. Finally, in order to obtain a single prediction, the scores of all OC-SVMs are fused by majority voting. With regard to unknown attacks not seen in the training phase, the authors reported an averaged APCER of 15.3% vs. an averaged BPCER of 10.8% on the LivDet 2011 database [39].

Another non-generative approach for face PAD has been proposed by Nikisins *et al.* [36], who use a combination of pre-trained autoencoders (AEs) and a simple multi-layer perceptron (MLP) for the final classification. The AEs are used to extract features from multi-channel input data, which in this case is a stack of greyscale, near-infrared, and depth facial images (BW-NIR-D). Each of the AEs are only trained on bona fide samples, thereby learning the appearance of real faces. Instead of collecting a lot of training data, Nikisins *et al.* use transfer learning techniques to transmit the knowledge of facial images from the RGB to the BW-NIR-D domain. The CNN model was pretrained on the CelebA [40] database containing RGB facial images, which they fine-tuned on the Wide Multi-Channel Presentation Attack database (WMCA) [41]. Only the subsequent MLP is trained on both bona fide and PA samples for the final classification of the face images. The authors report a BPCER of 7.3% vs. an APCER of 1%.

In another work on face PAD, Nikisins *et al.* [35] implemented and tested both one-class Gaussian mixture models (OC-GMM) (generative) and OC-SVMs (non-generative), benchmarking their results, with two-class approaches as well. For their experiments, the authors employed an aggregated database as a composition of three publicly available databases: Replay-Attack [42], Replay-Mobile [43], and MSU MFSD [44]. Their results show a significant better detection performance for the OC-GMM approach compared to the OC-SVM. Particularly, they emphasise the ability of the OC-GMM to have better generalisation properties to unknown attack types as compared to the two-class classifiers and the OC-SVMs. Both models were trained on the image quality metric features introduced in [44] and [45].

Lastly, Engelsma and Jain [37] present another one-class approach using generative adversarial networks (GANs) for fingerprint PAD. Specifically, they trained three different GAN models using the DCGAN architecture proposed by Radford *et al.* [38]. As part of their work, they collected a dataset comprising 12 different PAIs and 11,800 bona fide samples. The experimental evaluation reports an APCER of 15.6% for a BPCER of 0.2%.

III. CAPTURE DEVICE

The camera-based fingerprint capture device [46] that was used for data collection is depicted in Fig. 1. One camera (Basler acA1300-60gm) takes finger photos in the visible spectrum to extract the fingerprint for legacy compatibility. This camera is also able to capture finger vein images, when only the near-infrared (NIR) LEDs above the finger are switched on. A second camera (100 fps Xenics Bobcat 320) captures
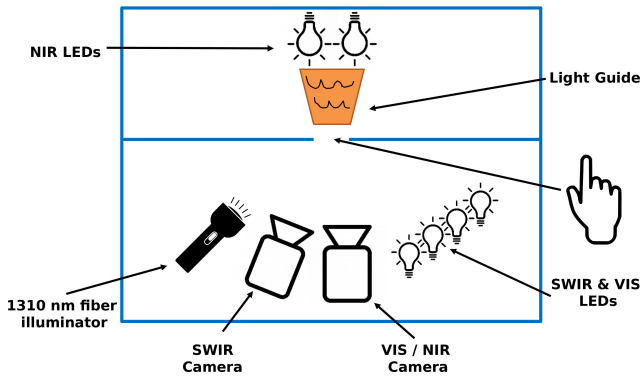
Fig. 1.   The capture device is a closed box with only one free slot for the finger. Two cameras and multiple illuminations are able to capture the fingerprint and additional PAD data.
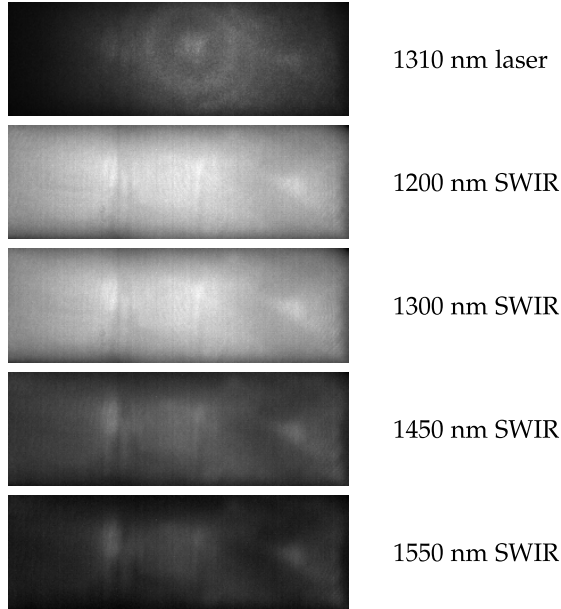

Fig. 2.   Bona fide samples acquired at five different wavelengths.

PAD data in wavelengths between 900 nm and 1700 nm. Both cameras are placed in a closed box next to multiple illumination sources with only one finger slot at the top. Once a finger is placed on this slot, all ambient light is blocked and only the desired wavelengths illuminate the finger. The invisible SWIR wavelengths of 1200 nm, 1300 nm, 1450 nm, and 1550 nm are especially suited for PAD because all skin types in the Fritzpatrick scale [47] reflect in the same way as shown by Steiner *et al.* [48] for face PAD. Hence, SWIR images are captured in each of these wavelengths. Additionally, a 1310 nm laser diode illuminates the finger area and a sequence of 100 frames is collected within one second. Stemming from biomedical applications, this laser sequence is used to image and monitor microvascular blood flow [32]. Since the laser scatters differently when penetrating human skin in contrast to artificial PAIs, this technique qualifies for PAD as well.

Example frames of a bona fide presentation acquired at the aforementioned wavelengths are shown in Fig. 2. For the laser sequence data, only one frame is depicted since the

subtle temporal changes are not visible in steady pictures. Nevertheless, we can recognise a circle where the laser focuses the finger. On the other hand, the LEDs achieve a much more consistent illumination for the SWIR images, where the skin reflections get darker for increasing wavelengths. The region of interest for all samples comprises 100×300 pixels due to the fixed size of the finger slot.

IV. PROPOSED PAD METHOD

This section introduces our one-class fingerprint PAD scheme based on a convolutional autoencoder, which is described in Section IV-A. Since AEs measure the reconstruction error, this concept is subsequently discussed in detail in Section IV-B. Finally, this scheme is combined with fingerprint PAD in Section IV-C.

A. Convolutional Autoencoder

A convolutional autoencoder is a neural network optimised to copy its input data. The model consists of two components: the encoder function $h = f(x)$ and the decoder function $x' = g(h)$, both of which are implemented as a multi-layer CNN. This means that the AE maps an input image x to an output image x'. The output h of the encoder function f is a lower dimensional latent representation of the original image x. Out of this latent variable, the decoder function g tries to reconstruct the original image x. In order to force the model to learn correct parameters for decoding the latent representation, a loss function needs to be minimised:

$$L(x, g(f(x))) \tag{1}$$

This loss function penalises $g(f(x))$ if it is dissimilar to x. The choice of the loss function thus plays a decisive role in the performance of convolutional AEs. In order to increase the efficiency of the learning process, the loss value can be calculated on a randomly selected subset called *Batch*. However, one important requirement is to design the architecture of an AE in an undercomplete way. In other words, the dimension of h needs to be smaller than the original dimension of input x. This forces the AE to only extract the most relevant features from the training data. Furthermore, it prevents the model to be in danger of learning the identity function $id(x) = x$ [49]. Once the model is trained, it is able to encode and reconstruct images x', which resemble the training data. In case of an input image that is dissimilar to the ones involved in training, the reconstruction fails and leads to a high reconstruction error (see Eq. (1)). The high input sensitivity of an AE can be exploited to detect images that differ from the ones being used during training. For this reason, AEs became very popular in the field of anomaly detection (e.g., [36], [50]). Transferred to the domain of fingerprint PAD, the AE is only trained on bona fide samples. Later, the model can be used to detect unknown PAs by comparing the reconstruction error against a threshold.

B. Reconstruction Error (RE)

A common approach to compute the reconstruction error is to use the mean squared error (MSE) [51] as loss function,

which is defined as

$$L(x, x') = \frac{1}{B}\sum_{j=1}^{B}\frac{1}{WHI}\sum_{w=1}^{W}\sum_{h=1}^{H}\sum_{i=1}^{I}\left(x_{whi}^{j} - x_{whi}^{\prime j}\right)^2$$

$$= \frac{1}{B}\sum_{j=1}^{B}\frac{1}{WHI}\sum_{w=1}^{W}\sum_{h=1}^{H}\sum_{i=1}^{I}e_{whi}^{j}(x, x') \qquad (2)$$

where B, W, H and I denote the number of data samples involved in one batch iteration, the image width, height and the number of input channels of an input image x. The usage of MSE is convenient since it is easy understandable and often pre-implemented. However, there is also a major drawback in case of random noise occurring in the data. Since the calculation of the MSE involves squaring the difference between every pixel of the input image, single outliers have a huge impact on the reconstruction error. This inevitably leads to an increased rate of bona fide samples erroneously classified as PAs. This lack of robustness against outliers is a well known challenge in the deep learning domain and is referred to as *robust estimation* [52]. The idea of increasing the robustness of an AE model for anomaly detection was studied by Ishii and Takanashi [50], who introduced a weighted version of the MSE (wMSE):

$$L_{Ishii}(x, x') = \frac{1}{B}\sum_{j=1}^{B}w^{j} \cdot mse^{j}(x, x') \qquad (3)$$

where

$$mse^{j}(x, x') = \frac{1}{WHI} \cdot \sum_{w=1}^{W}\sum_{h=1}^{H}\sum_{i=1}^{I}e_{whi}^{j}(x, x') \qquad (4)$$

and w^{j} is defined as

$$w^{j} = \begin{cases} 1, & mse^{j}(x, x') \leq C \\ 0, & mse^{j}(x, x') > C. \end{cases} \qquad (5)$$

Here, C refers to the α-th quantile of $mse = [mse^{1}, \ldots, mse^{B}]$. The approach of Ishii and Takanashi ignores training samples during the optimisation process as soon as their measured MSE exceeds a defined threshold C. Translated to the problem of fingerprint PAD, that means that a certain percentage of bona fides is ignored during the training phase. The authors state that their proposed loss function is useful to cope with unknown outliers within the training set, since they will not distort the resulting model. Unknown outliers can occur, for example, if the data is not labelled. Therefore, it is difficult to differentiate them from normal data samples. However, in our case the training data contains no PAs. That means that excluding bona fide samples from the training process could potentially lead to a loss of information.

For that reason, the proposed loss function of Ishii and Takanashi is adjusted within this work. The main idea is to integrate the weight factor such that it excludes pixel values that the AE is systematically not able to reconstruct. In other words, this means that the AE is optimised to reconstruct the most meaningful areas of the images while ignoring random

noise. The adjusted loss function is defined as follows:

$$L_{Prop}(x, x') = \frac{1}{B} \cdot \sum_{j=1}^{B}\frac{1}{WHI}\sum_{w=1}^{W}\sum_{h=1}^{H}\sum_{i=1}^{I}w_{whi}^{j}e_{whi}^{j}(x, x') \qquad (6)$$

with

$$w_{whi}^{j} = \begin{cases} 1, & e_{whi}^{j}(x, x') \leq mse^{j}(x, x') + C \cdot std^{j} \\ 0, & e_{whi}^{j}(x, x') > mse^{j}(x, x') + C \cdot std^{j} \end{cases} \qquad (7)$$

and

$$std^{j} = \sqrt{\frac{1}{WHI} \cdot \sum_{w=1}^{W}\sum_{h=1}^{H}\sum_{i=1}^{I}\left(e_{whi}^{j}(x, x') - mse^{j}(x, x')\right)^2} \qquad (8)$$

Generally speaking, every pixel value is compared to a threshold that is a linear combination of both mean and standard deviation of the squared error. Thus, exceeding pixels are ignored and contrary to the MSE, it is assumed that this approach prevents random noise from increasing the overall reconstruction error of the bona fide samples. The remaining challenge however consists in finding the optimal constant value of C. By choosing a too low threshold, the model might tend to over-generalise such that decisive patterns that are important for distinguishing between bona fides and PAs are not extracted anymore. On the other hand, if C is too high, noisy data might be involved in both training and testing, which leads to a less robust model and consequently increases error rates. This problem is related to the typical trade-off between bias and variance.

C. PAD Scheme

We study three different architectures of an AE, as illustrated in Fig. 3, in order to find the best suited approach for fingerprint PAD. The four SWIR images are concatenated to a single 4-channel image such that one AE can work on all information simultaneously. Taking the first, middle, and last frame of the laser sequence, a second AE is trained on a 3-channel input image. In contrast to a LSTM [53], the AE is not designed to learn temporal correlation, and since the changes within this sequence are subtle, we decided to take into account only the three furthermost frames in a similar way as the SWIR images are used. Due to the hardware changes of the capture device, computing the contrast of the laser sequence data does not work anymore as opposed to previous work [18], [24]. Hence, we discard the term LSCI and refer to laser sequences (or laser) in this work.

We denote the three architecture types as **Conv-AE**, **Pooling-AE**, and **Dense-AE** (top to bottom in Fig. 3). The names refer to the type of layers which were successively added to the architecture. The **Conv-AE** is composed of convolutional layers with a stride value of two in order to reduce the dimension during the encoding phase. In the **Pooling-AE**, the stride value of the convolutional operations was changed to one, followed by a max pooling operation to reduce the dimension. The last modification **Dense-AE** added a Fully Connected Neural Network (Fully-Connected NN) between
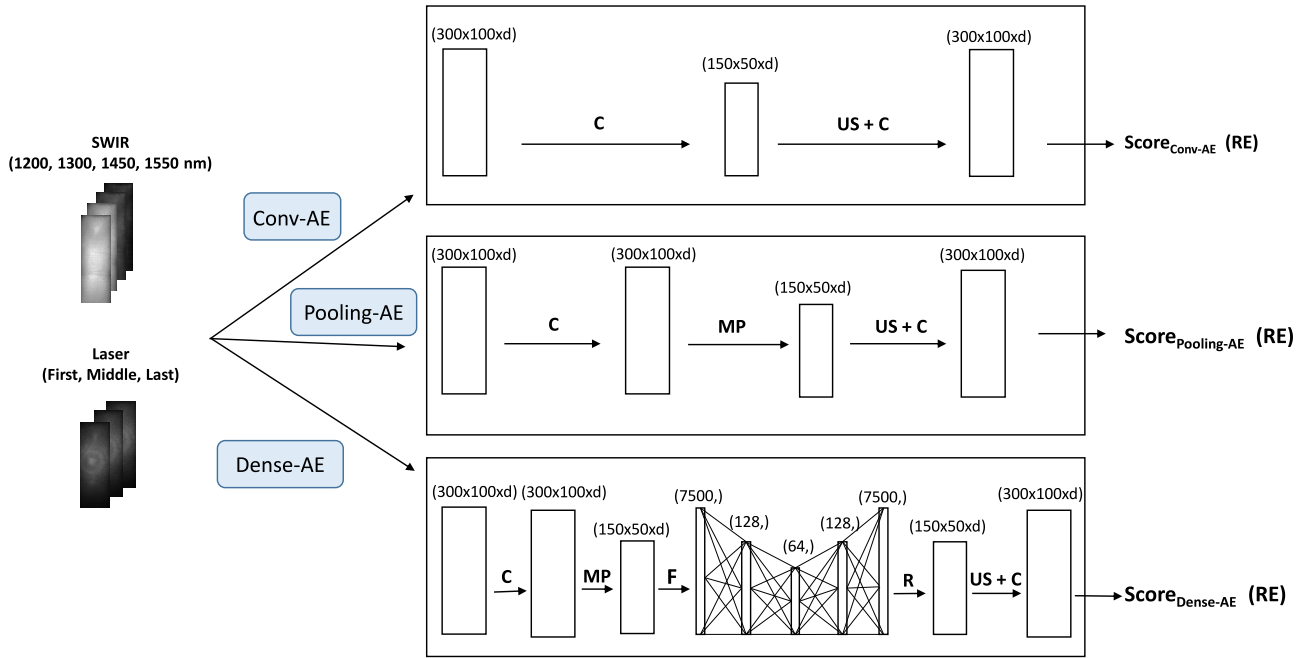
Fig. 3. The baseline architectures are defined as *Conv-AE*, *Pooling-AE*, and *Dense-AE*. The following operations are involved: C = Convolution, US = Upsampling, MP = Max Pooling, F = Flatten, R = Reshape. d is the image dimension: $d = 4$ for SWIR and $d = 3$ for laser samples.

the encoding and decoding phase to reduce the dimension of the original image down to a 64-dimensional vector. All baseline architectures include a single convolutional / max pooling layer in the encoding phase.

The distinction between the Conv- and Pooling AE is grounded on the findings of Springenberg *et al.* [54], who claim that the max pooling operation can simply be replaced by a convolutional layer with an increased stride without significant loss in accuracy. On the other hand, Goodfellow *et al.* [55] state that the max pooling operation leads to an invariance of translations in smaller regions. Finally, the Dense-AE is inspired by Ke *et al.* [56], who emphasise the ability of the Fully-Connected NN to combine local features and to find interdependent patterns within the feature maps. Across all architectures the *relu* activation function is used in all layers except for the very last convolutional layer, where the *sigmoid* function proved to be the better choice. The convolutional layer includes twelve filters and MSE (Eq. (2)) is used to measure the reconstruction error.

In a second step, we evaluate the influence of the reconstruction error. In particular, we take the best-performing architecture and compare the MSE approach to the wMSE approach by analysing different constant values C for the threshold computation. Hence, for each adaptation a new model is trained, since the loss function changes the learned weights during training.

Finally, we are interested in the best fusion of both AE types, based on SWIR and laser data, since previous approaches [19], [20], [25] show a significant improvement in PAD performance. For this reason, we compute different weighted fusions and compare the results in order to find the one best suited for our fingerprint PAD approach.

V. EXPERIMENTAL EVALUATION

Starting with a description of the utilised dataset and protocol, this section provides the details of the experimental evaluation. Subsequently, the results of our PAD method are presented and finally benchmarked with additional classifiers.

A. Database and Experimental Protocol

The data was collected in four acquisition sessions in two distinct locations within a timeframe of nine months. Subjects could participate multiple times and presented six to eight fingers per capture round including thumb, index, middle, and ring fingers. Fingers were presented as they were, which resulted in samples with different levels of moisture, dirt, or ink. Further details about the capture process are given in [46]. The combined database contains a total of 24,050 samples comprising 19,711 bona fides and additional 4,339 PAs stemming from 45 different PAI species. These PAI species include full fake fingers and more challenging overlays as summarised in Table III. The printouts were also worn as overlays and conductive paint was applied to some PAIs. Note that the project sponsor indicated to make the complete dataset available in the near future for reproducibility and comparison.[1]

The combined database is split into non-overlapping training, validation, and test sets, where subjects who participated multiple times are included in only one of the sets. This ensures a fair evaluation on unseen samples at the test stage. Randomly assigning 30% of the subjects to the training and additional 20% to the validation set results in the partitioning shown in Table IV.

Our implementation is done with Keras [57], which is a python based deep learning library that facilitates the

[1]https://www.isi.edu/projects/batl/data

TABLE III
SUMMARY OF PAIs IN THE DATABASE WITH THEIR CORRESPONDING
GROUP. THE NUMBER OF TOTAL SAMPLES AND THE NUMBER OF
VARIATIONS IS GIVEN. VARIATIONS INCLUDE, E.G., DIFFERENT
COLOURS AND CONDUCTIVE AUGMENTATIONS

PAI Group	PAI	#variations	#samples
Fakefinger	3D printed	2	72
	dental material	1	33
	dragon skin	3	477
	ecoflex	4	291
	latex	2	147
	playdoh	4	116
	silly putty	3	55
	wax	1	74
Overlay opaque	bandage plaster	1	14
	dental material	1	51
	dragon skin	1	17
	ecoflex	2	1035
	gelatin	1	194
	printout paper	1	49
	silicone	4	752
	urethane	1	72
Overlay transparent	dragon skin	1	106
	gelatin	1	107
	glue	2	27
	latex	1	34
	printout foil	1	64
	silicone	1	157
	wax	1	18
Overlay semi	dragon skin	1	47
	ecoflex	1	24
	glue	2	146
	silicone	1	160

TABLE IV
PARTITION OF TRAINING, VALIDATION, AND TEST DATASETS

	Samples	BF Samples	PA Samples
Training set	5,717	5,717	0
Validation set	3,553	3,553	0
Test set	14,780	10,441	4,339
Total	24,050	19,711	4,339

Fig. 4. DET curves for the three evaluated AE architectures on laser (top) and SWIR (bottom) data.

definition, training and evaluation of various deep learning model types. For training the parameters, we used the pre-implemented *RMSprop* [58] as an adaptive optimiser.

The PAD performance is shown in detection error trade-off (DET) curves between the BPCER and the APCER. For further comparison the partial area under curve (pAUC) of up to 20% error rate is computed for each curve. It should be noted that the PAD threshold can be adjusted depending on the use case: A low BPCER represents a very convenient system, while a low APCER is more important for high security applications. Furthermore, the detection equal error rate (D-EER) is the point where APCER = BPCER.

B. PAD Method Evaluation

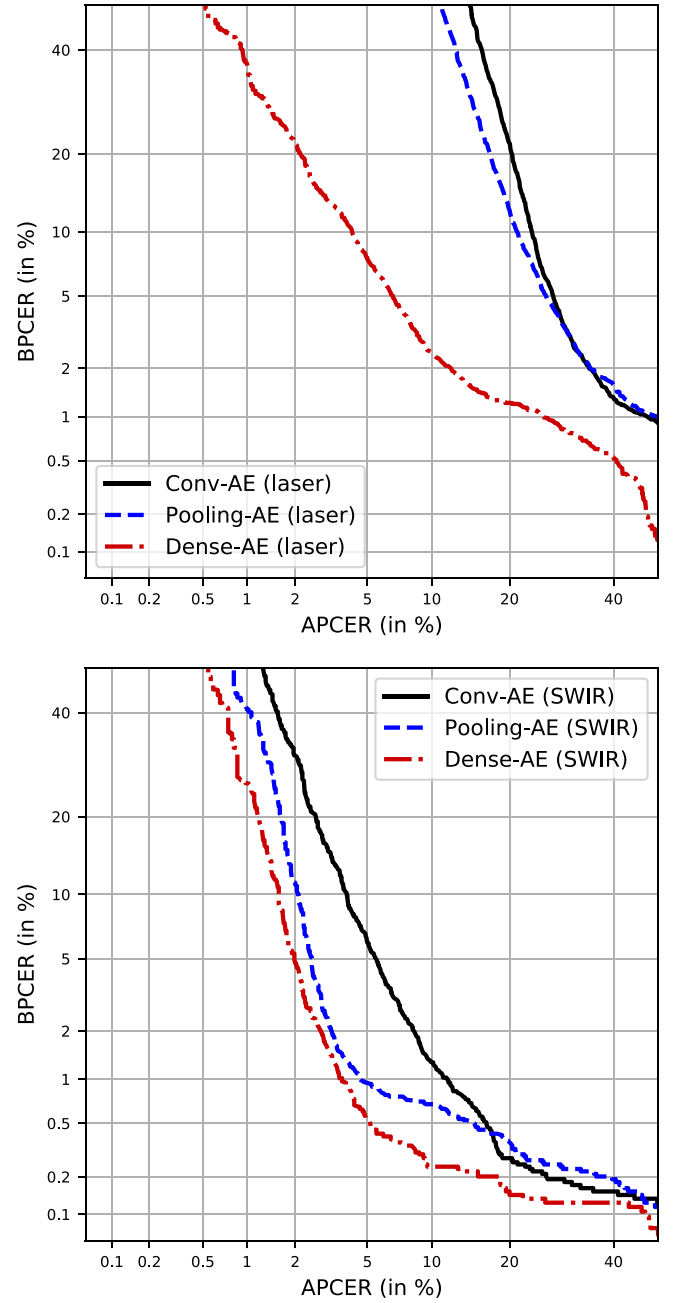The first part of our experiments compares the three baseline architectures: Conv-AE, Pooling-AE, and Dense-AE. The corresponding DET curves for both laser (top) and SWIR (bottom) input data are shown in Fig. 4. In both cases, the Dense-AE (red) achieves the best performance at all thresholds. Therefore, it can be concluded that the Dense-AE is better capable of extracting relevant latent features of the given input data, that can be reconstructed to the original image.

In the next step, the MSE (Eq. (2)) has been replaced by our proposed wMSE (Eq. (6)). Since the wMSE involves another hyperparameter C, Fig. 5 depict the DET curves for different parameter choices for laser and SWIR data, respectively. Also, the best performing baseline model has been added (Dense-AE with MSE) in order to directly compare it with the new settings. Looking at the graphs and the pAUC values, the
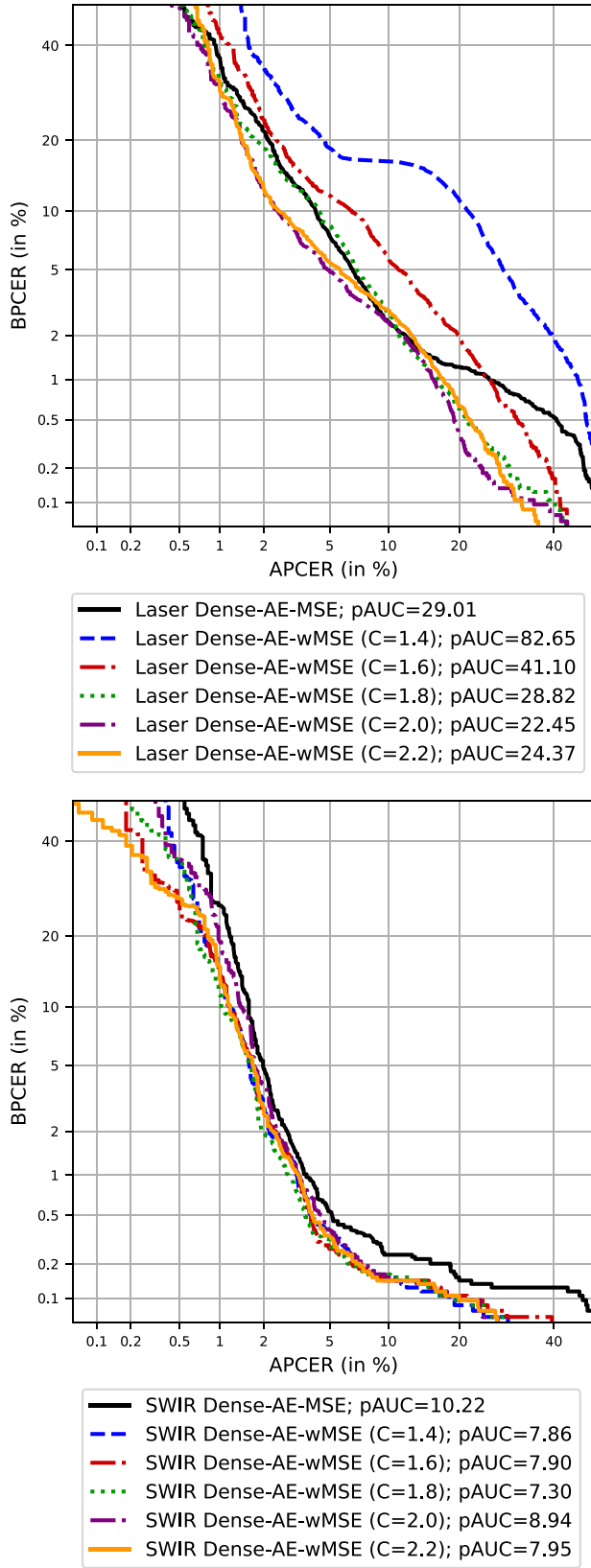
Fig. 5. DET curves and pAUC (%) of the Dense-AE comparing the MSE approach to the wMSE optimisation on laser (top) SWIR (bottom) data.
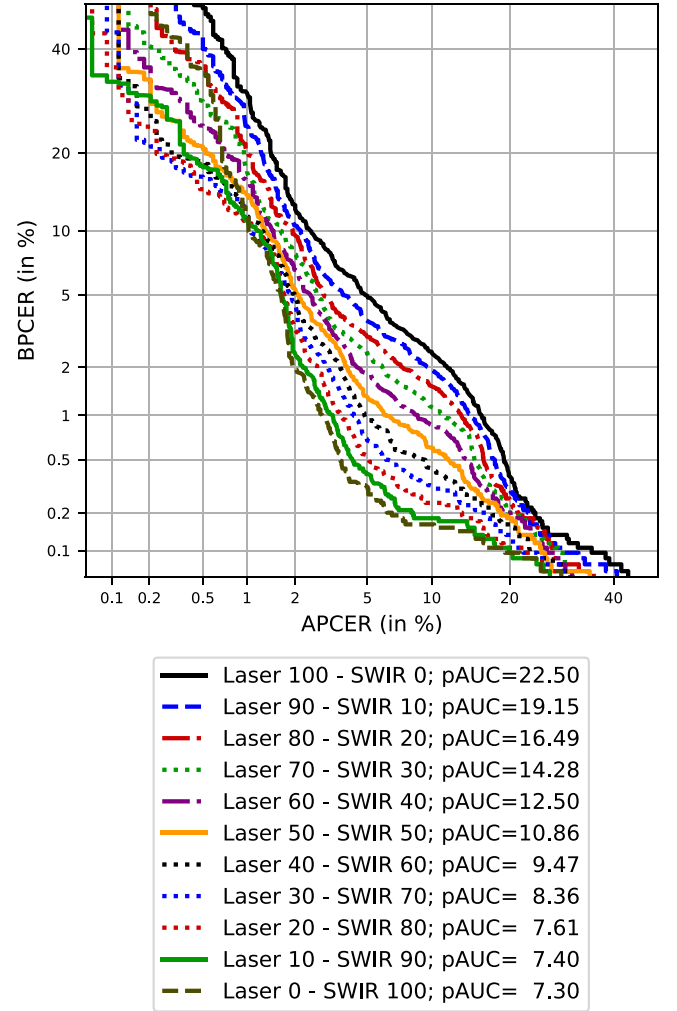

Fig. 6. DET curves and pAUC (%) from different weighted score-level fusions of the best-performing wMSE Dense-AEs.

phenomenon can be observed up to values of $C = 2.2$ (laser) and $C = 2.0$ (SWIR), where the performance decreases again. Choosing C values that are too high leads to thresholds, that non of the pixel-wise REs exceed. Therefore, too few areas are excluded from the training process. Hence, in our experiments, values of $C = 2.0$ (laser) and $C = 1.8$ (SWIR) proved to be good choices.

To evaluate whether the laser and SWIR AE models complement each other, we applied a weighted score fusion and the resulting DETs are depicted in Fig. 6. The given pAUCs show that the performance constantly decreases for higher weights on the laser scores. Thus, the optimal setting is to only use the SWIR scores as any inclusion of the laser scores has a negative effect on the classification results. On the other hand, for a possible high security application (e.g., APCER = 0.1%) the fusion benefits from the laser-based PAD. However, the BPCER values are above our 20% pAUC mark and thus not considered in computing the pAUC.

When analysing the occurring APCEs for a convenient BPCER = 0.2%, we found that all falsely classified PA samples of the SWIR AE are also misclassified by the laser AE. This includes mostly transparent overlays of clear dragon skin
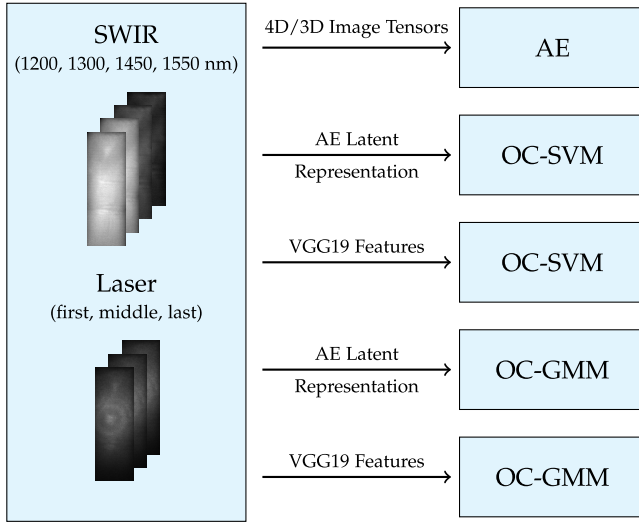
performance increases for growing values of C. This indicates that by choosing C too low, the excluded image areas are too large, which in turn leads to a loss of information. This

Fig. 7. Overview of the additionally benchmarked one-class classifiers and their corresponding input features.

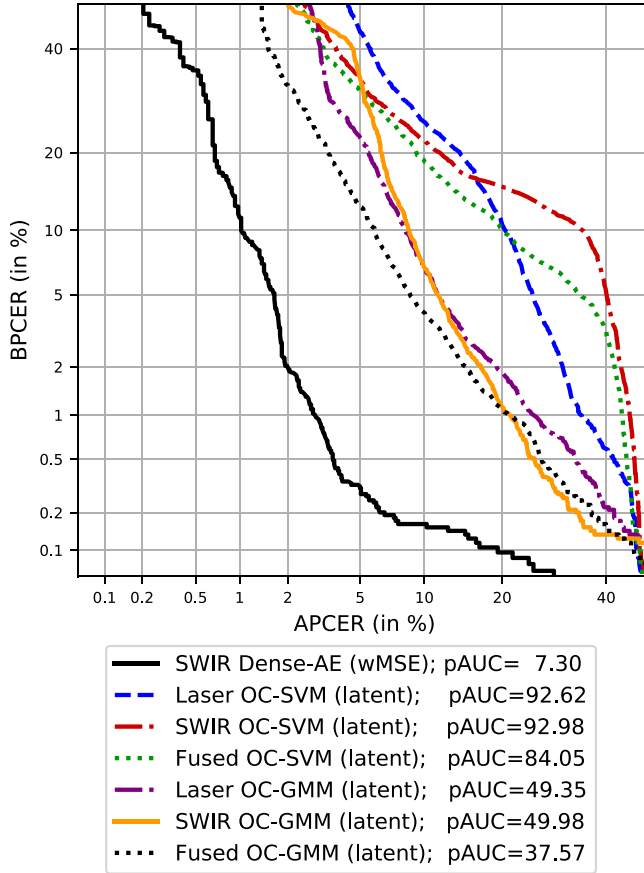Fig. 9. DET curves and pAUC (%) from the benchmark of SWIR wMSE Dense-AE to other classifiers trained on features extracted by VGG19.

Fig. 8. DET curves and pAUC (%) from the benchmark of SWIR wMSE Dense-AE to other classifiers trained on latent AE representations.

to the already mentioned APCEs, the laser AE further fails to detect full finger PAIs of dragon skin, ecoflex, and monster latex as well as overlays out of gelatin, school glue, ecoflex, gelatin, and monster latex. Since the laser samples are all captured in the same wavelength, PAIs are more likely to resemble bona fide samples.

C. Benchmark With Other One-Class Classifiers

Summarising the results so far, the best performance could be obtained with the Dense-AE trained on the SWIR dataset using the proposed wMSE. To put these numbers into context, we benchmark our proposed AE with further one-class classifiers. In this context, we train and test an OC-SVM [59] and an OC-GMM [60] on two different feature representations of the input images. One is the latent feature representation as a result of the encoding phase from our Dense-AE and the other method utilises the VGG19 [61] CNN pre-trained on [29] to only extract features from the given input. This results in a total of four combinations of classifiers and features for each SWIR and laser data as depicted in Fig. 7. Finally, the laser and SWIR approaches are also fused to enhance their detection accuracy. Fig. 8 and Fig. 9 visualise how the AE benchmarks against other one-class classifiers. The first graph contains the performance of OC-SVMs and OC-GMMs trained on the latent representations of the AE. The second graph shows the DET curves of both classifiers trained on features extracted

and two part silicone or full finger PAIs in yellow and orange playdoh. Also previous works [16], [28] on SWIR PAD had troubles with orange playdoh since its reflections are nearly identical to skin within the SWIR spectrum. The other APCEs are still close enough to bona fide representations that the reconstruction errors could not be distinguished. In addition
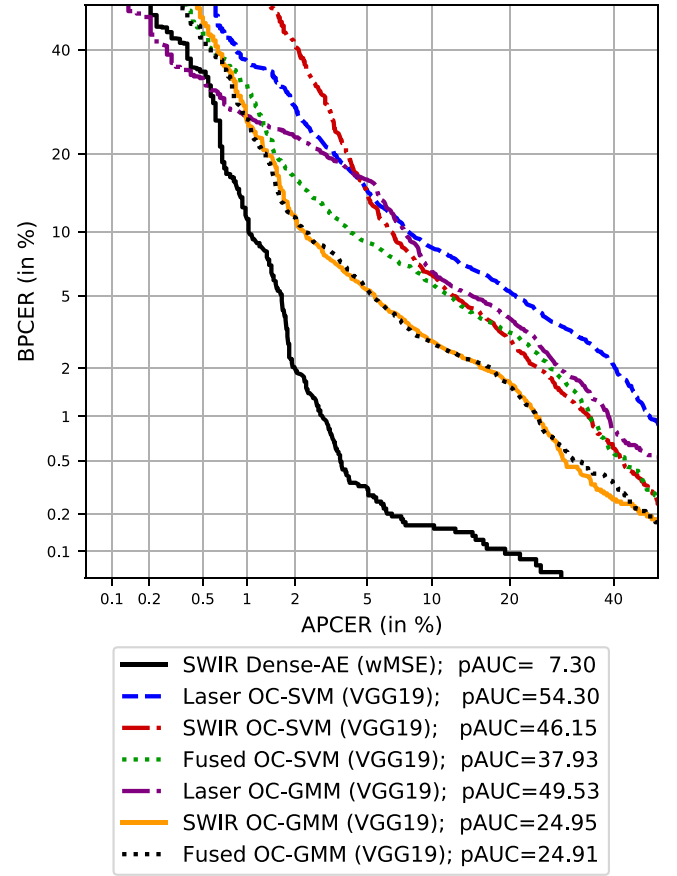
TABLE V
OVERVIEW OF APCERS FOR A FIXED BPCER OF 0.2% AND D-EERS

Algorithm	APCER	D-EER
Laser Dense-AE (wMSE)	24.33%	4.96%
SWIR Dense-AE (wMSE)	**6.59%**	**2.00%**
Laser OC-GMM (latent)	41.25%	8.80%
SWIR OC-GMM (latent	34.17%	8.98%
Fused OC-GMM (latent)	36.88%	7.09%
Laser OC-SVM (latent)	45.95%	16.21%
SWIR OC-SVM (latent)	47.27%	16.18%
Fused OC-SVM (latent)	45.86%	13.93%
Laser OC-GMM (VGG19)	63.98%	8.64%
SWIR OC-GMM (VGG19)	47.43%	5.21%
Fused OC-GMM (VGG19)	47.41%	5.16%
Laser OC-SVM (VGG19)	65.12%	9.07%
SWIR OC-SVM (VGG19)	52.82%	7.96%
Fused OC-SVM (VGG19)	55.55%	7.17%

with a pre-trained CNN (see Section IV). The AE performs significantly better than both other approaches since its curves are well below the other methods. Interestingly, the fused OC-GMM performs second-best with a pAUC of 37.57% (latent) and 24.91% (VGG19). Contrary to the AE, the performances of the OC-SVMs and OC-GMMs can be improved by fusing the laser and SWIR scores. Thus, in contrast to the AE, there is a complementary effect measurable.

In order to connect our results with the state-of-the-art, Table V shows the APCER values of this final benchmark for a very convenient BPCER of 0.2%. Furthermore, the D-EER is depicted as the point where APCER = BPCER. While above pAUC evaluated the general performance, those two operation points (APCER at BPCER=0.2%, D-EER) suit better to report specific results. The values show that both Dense-AEs outperform the other classifiers and prove the superiority of our SWIR Dense-AE. While our method achieves an APCER of 6.59% at a BPCER of 0.2% and a D-EER of 2.00%, the best performance of all other classifiers is an APCER of 34.17% at a BPCER of 0.2% for the SWIR OC-GMM based on the latent representation, and a D-EER of 5.16% for the fused OC-GMM based on the VGG19 representation.

VI. CONCLUSION

In this article, we have proposed a one-class PAD method based on convolutional autoencoders. Specifically, we compared three different AE architectures (Conv-AE, Pooling-AE, and Dense-AE). Based on our experiments, we can conclude that the Dense-AE performs significantly better than the other model architectures on both laser and SWIR input images.

Additionally, we proposed the wMSE as an extension of the MSE with the idea of ignoring disturbing image areas (e.g., illumination interference) during both training and testing. With the MSE replaced by the wMSE, the pAUC values could further be improved from 29.01% to 22.45% (laser) and from 10.22% to 7.30% (SWIR). The weighted fusion of the laser and SWIR scores did not improve the results. Therefore,

in contrast to related work applying two-class approaches, the two AEs do not complement each other.

Finally, two additional well-established one-class classifiers (OC-SVMs and OC-GMMs) have been trained on two different feature inputs. The first set of OC-SVMs and OC-GMMs were trained on the latent representations of the best performing AE. The second features have been extracted with a VGG19 [61] CNN pre-trained on [29]. None of the alternative one-class classifiers achieved a comparable performance to our proposed Dense-AE, which proves the soundness of the approach. Nevertheless, both alternative methods benefit from information fusion of laser and SWIR data.

Future work will focus on further optimising the wMSE. In our work, every pixel-wise RE gets an individual weight (zero or one) depending on whether it exceeds the chosen threshold C or not. This binary criterion could be loosened to allow the weights to have values between zero and one. Additionally, the concept of the Dense-AE can be applied to further PAD tasks as face and iris PAD, or software-based fingerprint PAD on the LivDet datasets.

REFERENCES

[1] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. London, U.K.: Springer, 2009.

[2] ISO/IEC JTC1 SC37 Biometrics, *Information Technology—Biometric Presentation Attack Detection—Part 1: Framework*, ISO/IEC Standard 30107-1, 2016.

[3] S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans, *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection*. Cham, Switzerland: Springer, 2019.

[4] D. Yambay, L. Ghiani, G. L. Marcialis, F. Roli, and S. Schuckers, "Review of fingerprint presentation attack detection competitions," in *Handbook of Biometric Anti-Spoofing*. Cham, Switzerland: Springer, 2019, pp. 109–131.

[5] D. Yambay, A. Czajka, K. Bowyer, M. Vatsa, R. Singh, and S. Schuckers, "Review of Iris presentation attack detection competitions," in *Handbook of Biometric Anti-Spoofing*. Cham, Switzerland: Springer, 2019, pp. 169–183.

[6] O. Kanich, M. Drahansky, and M. Mézl, "Use of creative materials for fingerprint spoofs," in *Proc. Int. Workshop Biometrics Forensics (IWBF)*, 2018, pp. 1–8.

[7] J. M. Singh, A. Madhun, G. Li, and R. Ramachandra, "A survey on unknown presentation attack detection for fingerprint," 2020. [Online]. Available: arXiv:2005.08337.

[8] D. M. J. Tax, "One-class classification: Concept learning in the absence of counter-examples," Ph.D. dissertation, Dept. Faculty Elect. Eng., Mathematics & Comput. Sci., TU Delft, Delft, The Netherlands, 2002.

[9] C. Sousedik and C. Busch, "Presentation attack detection methods for fingerprint recognition systems: A survey," *IET Biometrics*, vol. 3, no. 1, pp. 1–15, 2014.

[10] E. Marasco and A. Ross, "A survey on antispoofing schemes for fingerprint recognition systems," *ACM Comput. Surveys*, vol. 47, no. 2, pp. 1–36, 2014.

[11] ISO/IEC JTC1 SC37 Biometrics, *Information Technology—Biometric Presentation Attack Detection—Part 3: Testing and Reporting*, ISO/IEC Standard 30107-3, 2017.

[12] R. K. Rowe, K. A. Nixon, and P. W. Butler, "Multispectral fingerprint image acquisition," in *Advances in Biometrics: Sensors, Algorithms and Systems*. London, U.K.: Springer, 2008, pp. 3–23.

[13] C. Hengfoss, A. Kulcke, G. Mull, C. Edler, K. Püschel, and E. Jopp, "Dynamic liveness and forgeries detection of the finger surface on the basis of spectroscopy in the 400–1650 nm region," *Forensic Sci. Int.*, vol. 212, no. 1, pp. 61–68, 2011.

[14] M. Drahansky, M. Dolezel, J. Vana, E. Brezinova, J. Yim, and K. Shim, "New optical methods for liveness detection on fingers," *BioMed Res. Int.*, vol. 2013, Sep. 2013, Art. no. 197925.

[15] L. N. Darlow, L. Webb, and N. Botha, "Automated spoof-detection for fingerprints using optical coherence tomography," *Appl. Opt.*, vol. 55, no. 13, pp. 3387–3396, 2016.

[16] M. Gomez-Barrero, J. Kolberg, and C. Busch, "Towards fingerprint presentation attack detection based on short wave infrared imaging and spectral signatures," in *Proc. Norwegian Inf. Security Conf. (NISK)*, 2018, pp. 1–10.

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

[18] P. Keilbach, J. Kolberg, M. Gomez-Barrero, C. Busch, and H. Langweg, "Fingerprint presentation attack detection using laser speckle contrast imaging," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, 2018, pp. 1–6.

[19] M. 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 *Proc. IEEE Workshop Inf. Forensics Security (WIFS)*, 2018, pp. 1–8.

[20] M. Gomez-Barrero, J. Kolberg, and C. Busch, "Towards multi-modal finger presentation attack detection," in *Proc. Int. Workshop Ubiquitous Implicit Biometrics Health Signals Monitor. Person Centric Appl. (UBIO)*, 2018, pp. 547–552.

[21] F. Liu, G. Liu, and X. Wang, "High-accurate and robust fingerprint anti-spoofing system using optical coherence tomography," *Expert Syst. Appl.*, vol. 130, pp. 31–44, Sep. 2019.

[22] H. Mirzaalian, M. Hussein, and W. Abd-Almageed, "On the effectiveness of laser speckle contrast imaging and deep neural networks for detecting known and unknown fingerprint presentation attacks," in *Proc. Int. Conf. Biometrics (ICB)*, 2019, pp. 1–8.

[23] T. Chugh and A. K. Jain, "OCT fingerprints: Resilience to presentation attacks," 2019. [Online]. Available: arXiv:1908.00102.

[24] J. Kolberg, M. Gomez-Barrero, and C. Busch, "Multi-algorithm benchmark for fingerprint presentation attack detection with laser speckle contrast imaging," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, 2019, pp. 1–5.

[25] M. Gomez-Barrero, J. Kolberg, and C. Busch, "Multi-modal fingerprint presentation attack detection: Looking at the surface and the inside," in *Proc. Int. Conf. Biometrics (ICB)*, 2019, pp. 1–8.

[26] J. Kolberg, M. Gomez-Barrero, S. Venkatesh, R. Ramachandra, and C. Busch, "Presentation attack detection for finger recognition," in *Handbook of Vascular Biometrics*. Cham, Switzerland: Springer, 2020, pp. 435–463.

[27] R. Tolosana, M. Gomez-Barrero, C. Busch, and J. Ortega-Garcia, "Biometric presentation attack detection: Beyond the visible spectrum," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1261–1275, Aug 2019, doi: 10.1109/TIFS.2019.2934867.

[28] M. Gomez-Barrero, R. Tolosana, J. Kolberg, and C. Busch, "Multispectral short wave infrared sensors and convolutional neural networks for biometric presentation attack detection," in *AI and Deep Learning in Biometric Security: Trends, Potential and Challenges*. Boca Raton, FL, USA: CRC Press, 2020.

[29] J. Kolberg, A. C. Vasile, M. Gomez-Barrero, and C. Busch, "Analysing the performance of LSTMs and CNNs on 1310 nm laser data for fingerprint presentation attack detection," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, 2020, pp. 1–7.

[30] R. Breithaupt, C. Sousedik, and S. Meissner, "Full fingerprint scanner using optical coherence tomography," in *Proc. Int. Workshop Biometrics Forensics (IWBF)*, 2015, pp. 1–6.

[31] Y. Moolla, L. Darlow, A. Sharma, A. Singh, and J. van der Merwe, "Optical coherence tomography for fingerprint presentation attack detection," in *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection*, S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans, Eds. Cham, Switzerland: Springer, 2019, pp. 49–70.

[32] J. Senarathna, A. Rege, N. Li, and N. V. Thakor, "Laser speckle contrast imaging: Theory, instrumentation and applications," *IEEE Rev. Biomed. Eng.*, vol. 6, pp. 99–110, 2013.

[33] P. G. Vaz, A. Humeau-Heurtier, E. Figueiras, C. Correia, and J. Cardoso, "Laser speckle imaging to monitor microvascular blood flow: A review," *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 106–120, 2016.

[34] Y. Ding and A. Ross, "An ensemble of one-class SVMs for fingerprint spoof detection across different fabrication materials," in *Proc. IEEE Int. Workshop Inf. Forensics Security (WIFS)*, 2016, pp. 1–6.

[35] O. Nikisins, A. Mohammadi, A. Anjos, and S. Marcel, "On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing," in *Proc. IEEE Int. Conf. Biometrics (ICB)*, 2018, pp. 75–81.

[36] O. Nikisins, A. George, and S. Marcel, "Domain adaptation in multichannel autoencoder based features for robust face anti-spoofing," in *Proc. IEEE Int. Conf. Biometrics (ICB)*, 2019, pp. 1–8.

[37] J. J. Engelsma and A. K. Jain, "Generalizing fingerprint spoof detector: Learning a one-class classifier," 2019. [Online]. Available: arXiv:1901.03918.

[38] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015. [Online]. Available: arXiv:1511.06434.

[39] D. Yambay, L. Ghiani, P. Denti, G. L. Marcialis, F. Roli, and S. Schuckers, "LivDet 2011 fingerprint liveness detection competition 2011," in *Proc. IEEE Int. Conf. Biometrics (ICB)*, 2012, pp. 208–215.

[40] Z. Liu, P. Luo, X. Wang and X. Tang, "Deep Learning Face Attributes in the Wild," in *Proc. Intl. Conf. Computer Vision (ICCV)*, Dec. 2015, pp. 3730–3738.

[41] 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 Trans. Inf. Forensics Security*, vol. 15, pp. 42–55, May 2019, doi: 10.1109/TIFS.2019.2916652

[42] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. IEEE Int. Conf. Biometrics Special Interest Group (BIOSIG)*, 2012, pp. 1–7.

[43] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, "The replay-mobile face presentation-attack database," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, 2016, pp. 1–7.

[44] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015.

[45] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint and face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 710–724, Feb. 2014.

[46] L. Spinoulas, H. Mirzaalian, M. Hussein, and W. AbdAlmageed, "Multimodal fingerprint presentation attack detection: Evaluation on a new dataset," 2020. [Online]. Available: arXiv:2006.07498.

[47] T. B. Fitzpatrick, "The validity and practicality of sun-reactive skin types I through VI," *Archives Dermatol.*, vol. 124, no. 6, pp. 869–871, 1988.

[48] H. Steiner, A. Kolb, and N. Jung, "Reliable face anti-spoofing using multispectral SWIR imaging," in *Proc. Int. Conf. Biometrics (ICB)*, 2016, pp. 1–8.

[49] S. Skansi, "Autoencoders," in *Introduction to Deep Learning*. Cham, Switzerland: Springer, 2018, pp. 153–163.

[50] Y. Ishii and M. Takanashi, "Low-cost unsupervised outlier detection by autoencoders with robust estimation," *J. Inf. Process.*, vol. 27, pp. 335–339, Apr. 2019.

[51] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York, NY, USA: Springer, 2013.

[52] P. J. Rousseeuw and M. Hubert, "Robust statistics for outlier detection," *Wiley Interdiscipl. Rev. Data Min. Knowl. Disc.*, vol. 1, no. 1, pp. 73–79, 2011.

[53] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[54] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2014. [Online]. Available: arXiv:1412.6806.

[55] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[56] M. Ke, C. Lin, and Q. Huang, "Anomaly detection of logo images in the mobile phone using convolutional autoencoder," in *Proc. IEEE Int. Conf. Syst. Informat. (ICSAI)*, 2017, pp. 1163–1168.

[57] F. Chollet *et al.* (2015). *Keras*. [Online]. Available: https://keras.io

[58] V. Bushaev, "Understanding RMSprop—Faster neural network learning," in *Towards Data Science*, 2018. [Online]. Available: https://towardsdatascience.com/understanding-rmsprop-faster-neural-network-learning-62e116fcf29a

[59] Y. Chen, X. S. Zhou, and T. S. Huang, "One-class SVM for learning in image retrieval," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, 2001, pp. 34–37.

[60] D. A. Reynolds, "Gaussian mixture models," *Encyclopedia of Biometrics*, vol. 741. Berlin, Germany: Springer, 2009.

[61] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. [Online]. Available: arXiv:1409.1556.
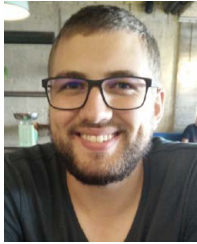
Jascha Kolberg (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in IT-security/information technology from Ruhr-University Bochum in 2014 and 2017, respectively. Since 2017, he has been a Ph.D. student with da/sec, National Research Center for Applied Cybersecurity (ATHENE). His current research focuses on biometric template protection and presentation attack detection for fingerprint recognition sensors.

Marcel Grimmer received the B.Sc. and M.Sc. degrees in applied mathematics and data science from the University of Applied Sciences Darmstadt in 2018 and 2020. Since 2020, he has been a Ph.D. student with the Norwegian Biometrics Laboratory, Norwegian University of Science and Technology (NTNU). During his studies, his research focused on presentation attack detection for fingerprints. His main research is dedicated to image quality assessment of facial images.

Marta Gomez-Barrero (Member, IEEE) received the M.Sc. degree in computer science and mathematics and the Ph.D. degree in electrical engineering from the Universidad Autonoma de Madrid, Spain, in 2011 and 2016, respectively. She is a Professor for IT-security and technical data privacy with Hochschule Ansbach, Germany. From 2016 to 2020, she was a Postdoctoral Researcher with the National Research Center for Applied Cybersecurity (ATHENE)–Hochschule Darmstadt, Germany. She has coauthored more than 70 publications, chaired special sessions and competitions at international conferences. Her current research focuses on security and privacy evaluations of biometric systems, presentation attack detection methodologies, and biometric template protection schemes. She has also received a number of distinctions, including the EAB European Biometric Industry Award 2015, the Best Ph.D. Thesis Award by Universidad Autonoma de Madrid in 2015 and 2016, the Siew-Sngiem Best Paper Award at ICB 2015, the Archimedes Award for young researches from Spanish MECD, and the Best Poster Award at ICB 2013. She is an Associate Editor for the *EURASIP Journal on Information Security*, and represents the German Institute for Standardization in ISO/IEC SC37 JTC1 SC37 on biometrics.

Christoph Busch received the Diploma degree from the Technical University of Darmstadt (TUD), Darmstadt, Germany, and the Ph.D. degree in computer graphics from TUD, in 1997. He joined the Fraunhofer Institute for Computer Graphics, Darmstadt, in 1997. He is a Member of the Norwegian Biometrics Laboratory, Norwegian University of Science and Technology, Norway, and holds a joint appointment with the Faculty of Computer Science, Hochschule Darmstadt. He has been lecturing a course on biometric systems with DTU, Copenhagen, since 2007. He has coauthored over 500 technical papers, and has been a speaker at international conferences. His research includes pattern recognition, multimodal and mobile biometrics, and privacy enhancing technologies for biometric systems. He is the Co-Founder of the European Association for Biometrics and Convener of WG3 in ISO/IEC JTC1 SC37 on Biometrics.