Journal of Science and Technology – Smart Systems and Devices
Volume x, Issue x, Month Year, ppp-ppp

Privacy-Conscious Synthetic Fingerprint Generation for Generalizable Presentation Attack Detection

Linh Dang Tuan*, Huy Ho Tuan, An Nguyen Binh, Yen Le Hai, Minh Nguyen Tuan

Hanoi University of Science and Technology, Ha Noi, Vietnam
**Corresponding author email: linhdt@soict.hust.edu.vn*

Abstract

Fingerprint Presentation Attack Detection (FPAD) is essential for securing biometric systems against spoofing attacks. However, the development of robust models is often hindered by the scarcity and limited material diversity of existing spoof datasets. This paper presents a comprehensive empirical study on the utility of material-specific synthetic fingerprints for enhancing FPAD generalization. Specifically, we employ StyleGAN2-ADA to generate a large-scale dataset of 7,200 spoof images across 72 distinct material categories, modeled after real-world textural and optical characteristics. We evaluate the proposed dataset using a ResNet-50 backbone under an unseen-material protocol, shifting the evaluation focus from standard accuracy to security-critical metrics, including APCER, BPCER, and ACER (ISO/IEC 30107-3). Our findings demonstrate that, in this comparison, StyleGAN2-ADA introduces more challenging variations that may lower in-distribution F1-scores relative to the CycleGAN-based baseline, while providing superior structural integrity and closer distributional alignment with real fingerprints. These results indicate that high-fidelity synthetic data serves as a scalable, privacy-conscious alternative to manual collection, enabling the detection of novel presentation attacks in real-world security applications.

Keywords: Fingerprint presentation attack detection, privacy-conscious biometrics, StyleGAN2-ADA, synthetic fingerprint generation, unseen-material generalization.

1. Introduction

Fingerprint recognition remains one of the most widely adopted biometric modalities in modern security systems due to its high distinctiveness, permanence, and cost-effectiveness. It is extensively deployed in applications ranging from mobile device authentication and access control to large-scale identification systems in government and border security. However, the increasing ubiquity of fingerprint-based authentication has also made such systems attractive targets for presentation attacks, where fake fingerprints are used to deceive the sensor. As a result, Fingerprint Presentation Attack Detection (FPAD) has become a critical component for ensuring the reliability and trustworthiness of fingerprint recognition systems.

Recent advances in FPAD show that data-driven methods, especially deep learning, are much better than traditional feature-based methods at distinguishing real from fake fingerprints. However, the success of these approaches depends heavily on the availability of large-scale, diverse, and representative datasets. These datasets allow models to learn key textural and structural cues that manual features can miss, making them vital for defense against new spoofing attacks.

Despite their importance, collecting high-quality spoof fingerprint data remains a major challenge: manual acquisition is expensive and time-consuming, privacy regulations restrict large-scale collection from human subjects, and existing datasets are often limited in material diversity. This situation may lead to poor generalization to unseen attack types. These constraints highlight a critical research gap: the need for scalable, privacy-compliant alternatives that enable FPAD systems to reliably detect novel presentation attacks in real-world deployments.

The contributions of this paper are as follows :

1) We construct and publicly release a large-scale, material-diverse synthetic spoof fingerprint dataset spanning 72 material categories (7,200 images). The dataset is privacy-conscious by construction, since the samples are drawn from a learned generative distribution rather than from identifiable human subjects, and can therefore be shared without the human-subject data-protection constraints that govern real spoof collections.
2) We conduct a controlled empirical study under an unseen-material protocol, evaluating fingerprint presentation attack detection with security-critical metrics (APCER, BPCER, and ACER following ISO/IEC 30107-3 [1]).
3) We show empirically that distributional realism and in-distribution classifier separability are distinct properties that can diverge: a more realistic synthetic dataset can yield a lower in-distribution F1 score while achieving superior security generalization to

p-ISSN 3093-3285
e-ISSN 3093-3315
https://doi.org/10.51316/jst.xxx.ssad.xxxx.xx.x.x
Received: Mar 26, 2026; Revised: Jul 28, 2026;
Accepted: Jul 30, 2026; Online: Aug 26, 2026

unseen attacks, as evidenced by lower APCER and ACER. This finding reframes how synthetic FPAD data should be assessed, shifting the emphasis from classification accuracy to security-oriented evaluation.

2. Related Work

The development and benchmarking of robust FPAD algorithms are linked to the availability of diverse, high-quality datasets. For over a decade, the landscape of FPAD research has been defined primarily by the Liveness Detection (LivDet) [2, 3] competition series, which established standard protocols for contact-based sensors, and more recently by datasets such as COLFISPOOF [4] that address the emerging paradigm of contactless acquisition.

2.1. Existing Fingerprint PAD Datasets

The LivDet series, spanning from 2009 to 2025 [2, 3] , represents the benchmark for the field. A defining characteristic of these datasets is the methodology used to generate presentation attacks (PAs). The majority of spoof artifacts in the LivDet samples are created using the "consensual method". In this cooperative process, a subject presses their live finger into a soft material-such as dental impression compound, Play-Doh, or plaster-to create a negative mold. A liquid casting agent, typically silicone, gelatin, wax, or latex, is then poured into this mold. Once hardened, the resulting artificial finger retains the precise topographic structure and natural distortions of the original fingerprint, posing a significant challenge to detection algorithms.

In response to the ubiquity of mobile devices, the COLFISPOOF [4] dataset represents a significant departure from traditional contact-based scanning. Unlike the grayscale images typical of the LivDet series, COLFISPOOF utilizes high-resolution smartphone cameras (Huawei P20 Pro and Samsung Galaxy S9+) to capture RGB images. This dataset addresses the unique domain of contactless PAD, incorporating a wide array of attack instruments ranging from simple paper printouts to complex 3D artifacts modeled from Dragonskin, Silly Putty, and translucent glues. This shift introduces new variables, such as color information and environmental lighting, which were absent in legacy datasets. Table 1 summarizes the materials and sensor configurations of commonly used fingerprint PAD datasets.

Critically, none of these datasets were designed with privacy-compliant redistribution in mind, nor do they cover the breadth of material substrates required for robust generalization-gaps that directly motivate the synthetic generation framework proposed in this work.

Table 1. Comparison of fingerprint PAD datasets

Dataset	**Materials**	**Sensors**
LivDet Series	Silicone, Gelatin, Latex, Play-Doh, Wood Glue, Ecoflex, Body Double.	Contact-based: Primarily Optical and Capacitive sensors (e.g., CrossMatch, Biometrika, Green Bit).
COLFIS-POOF	72 distinct variations including Dragon Skin, Silly Putty, modeling clay, paper printouts, and traditional casting agents.	Contactless: High-resolution smartphone cameras (Huawei P20 Pro, Samsung Galaxy S9+).

2.2. The Fingerprint Presentation Attack Detection (FPAD) Problem

The evolution of FPAD algorithms can be broadly categorized into two distinct eras: traditional methods relying on hand-crafted features and modern approaches utilizing deep learning.

2.2.1. Traditional hand-crafted features

Early research focused on identifying specific liveness cues that distinguish biological tissue from synthetic materials. One line of investigation focused on physiological features, grounded in the biological reality that living fingers possess perspiration systems and fine-grained pore structures. Another study [5] proposed an analysis based on fingerprint pores and sweat activity. Their method postulates that real fingers exhibit dynamic signal changes due to perspiration and static high-frequency details, whereas artificial materials like silicone or plastic remain static and lack this micro-topology.

Parallel to physiological analysis, a significant body of work concentrated on texture and local quality features. The hypothesis suggests that creating a fake artifact will introduce noise, lower ridge clarity, or change the surface texture compared to a real finger. To capture these anomalies, researchers have employed low-level feature descriptors. For instance, combinations of SURF, PHOG, and Gabor wavelets [6] have been utilized to analyze gradient and shape characteristics from a single image. By using classifiers, these systems can tell the difference between the waxy or coarse texture of a fake finger and real skin. Despite their interpretability, such hand-crafted approaches are inherently brittle: their effectiveness degrades significantly when encountering novel spoofing

materials or sensor types not represented during feature design, highlighting the need for more adaptive detection strategies.

2.2.2. The Shift to deep learning

While hand-crafted features provided a strong foundation, they often struggled to generalize across different sensor types or novel spoofing materials. This limitation necessitated a shift toward Deep Learning techniques, which automate the feature extraction process. Nogueira *et al.* [7] demonstrated the powerful potential of Convolutional Neural Networks (CNNs) for FPAD. Their work highlighted that CNNs, particularly when leveraging transfer learning from pre-trained models, could learn complex, hierarchical representations of liveness that far exceed the performance of manual texture descriptors. This transition marked a turning point in the field, establishing Deep Learning as the dominant methodology for handling the increasing complexity of presentation attacks. However, deep learning-based approaches require large amounts of fingerprint data, which can pose challenges due to limited datasets. This data dependency underscores a critical bottleneck: without large-scale, material-diverse training data, deep learning models remain vulnerable to unseen attack types, directly motivating the synthetic data generation approach proposed in this work.

2.3. Generation Techniques in Biometrics

The synthesis of biometric data has emerged as a research priority to mitigate challenges associated with data scarcity, privacy regulations, and the requirement for large-scale annotated datasets. This section evaluates the evolution of generative frameworks, focusing on their capacity to model fingerprint morphology and material-specific textures.

2.3.1. Adversarial and probabilistic frameworks

Generative Adversarial Networks (GANs) [8] frame image synthesis as an adversarial game between a generator and a discriminator, enabling the approximation of complex real-data distributions. In the domain of fingerprint synthesis, early GAN applications focused on structural integrity. For example, Bontrager et al. demonstrated that adversarial networks could generate DeepMasterPrints [9] capable of undermining biometric systems. However, these models frequently failed to capture intra-class variations and the fine-grained textures necessary for PAD. Alternative probabilistic approaches such as VAEs [10] also prove inadequate, as their pixel-wise reconstruction losses introduce blurring artifacts that degrade fine ridge details critical for spoof synthesis. These limitations motivate the adoption of style-based architectures, which offer superior disentanglement of identity and material properties, enabling more faithful and controllable synthesis of spoof-specific textures.

2.3.2. Style-based architectures and StyleGAN2-ADA

The StyleGAN family, such as StyleGAN2 [11] and its successor StyleGAN2-ADA [12], represents a departure from traditional convolutional generators. The architecture employs a mapping network to transform a latent code $z \in \mathscr{Z}$ into an intermediate style space $w \in \mathscr{W}$, which subsequently modulates the synthesis network. A critical refinement in StyleGAN2 is the replacement of Adaptive Instance Normalization (AdaIN) with weight demodulation, which eliminates droplet artifacts and improves image quality. Following Karras *et al.* [11, 12], the modulation and demodulation of weights w_{ijk} are formulated as:

$$w'_{ijk} = s_i \cdot w_{ijk}, \quad w''_{ijk} = \frac{w'_{ijk}}{\sqrt{\sum_{i,k}(w'_{ijk})^2 + \varepsilon}} \qquad (1)$$

where s_i denotes the scale factor derived from the style vector.

This hierarchical design aligns well with the multi-scale structure of fingerprints: coarse synthesis layers govern global ridge flow and minutiae-like geometry (Levels 1–2), while fine layers modulate local surface texture and pore details (Level 3). Stochastic per-pixel noise injection further enables simulation of material-specific surface roughness. Critically, the $\mathscr{W}$ latent space allows independent manipulation of "material style" and "fingerprint identity," facilitating the synthesis of multiple spoof impressions per identity which is a capability constrained in earlier architectures such as Finger-GAN [13].

For the synthesis of material-specific fingerprints, the Adaptive Discriminator Augmentation (ADA) [12] is essential. ADA addresses discriminator overfitting that occurs when the training dataset is limited. This situation is a common occurrence when dealing with specific categories like gelatin or silicone. By dynamically augmenting both real and synthetic samples without leaking augmentations to the generated output, StyleGAN2-ADA maintains training stability and convergence on small-scale biometric datasets.

Unlike prior works that apply StyleGAN2-ADA to general image synthesis, we specifically exploit the W latent space to independently modulate material-specific surface properties (e.g., silicone translucency, gelatin roughness) from underlying ridge topology-a disentanglement property that is uniquely suited to the FPAD augmentation objective.

2.3.3. Recent advancements in biometric synthesis

Recent research has integrated domain-specific constraints into the generative process. Engelsma *et al.* (PrintsGAN) [14] incorporated a multi-stage process to ensure the synthesis of biometrically consistent fingerprints, demonstrating that deep networks pretrained on synthetic data achieve superior True

Accept Rates (TAR) when fine-tuned on real samples. In contrast to PrintsGAN [14], which synthesizes bona fide identities to improve recognition (TAR), our objective is to synthesize presentation attacks with material-specific spoof textures for PAD, which a distinct task that PrintsGAN does not address. Style-based architectures like StyleGAN2-ADA remain the standard for large-scale fingerprint synthesis due to their superior latent disentanglement and inference efficiency.

3. Methodology

A StyleGAN-based generative framework is proposed to synthesize fingerprint images on multiple material substrates for FPAD. The methodology is designed to preserve identity-consistent ridge structures while explicitly modeling material-dependent artifacts that characterize spoof presentations.

3.1. Overall Framework

The proposed approach follows a structured four-stage pipeline comprising: (i) acquisition of real fingerprint seed data, (ii) data pre-processing and augmentation, (iii) StyleGAN-based model training, and (iv) synthetic fingerprint generation with quality control and material-aware labeling.

StyleGAN2-ADA is adopted as the core generative architecture due to its robustness under limited data regimes, which are common in biometric and presentation attack datasets. Within this framework, the generator is optimized to jointly capture fingerprint ridge morphology and material-specific surface textures, while ADA mitigates discriminator overfitting and stabilizes adversarial training. The details of the framework can be seen in Fig 1.

3.2. Dataset Origin and Material Construction

In this study, a material is a type based on how it looks and feels, its color, shine, and texture, related to a particular presentation tool, instead of a specific way of making it. Data provenance. Real (live) images come from the public LivDet 2009 dataset. Fake samples are made using our StyleGAN2-ADA model, based on the public COLFISPOOF material space. As all synthetic samples are produced by a single generator and filtered by a single quality-control pipeline, acquisition conditions are uniform across all 72 categories. The training data for our generative framework are constructed to bridge the gap between existing benchmarks and the requirement for high material diversity. Bona fide (live) fingerprint images are sourced from the LivDet 2009 dataset to provide high-quality reference ridge structures. To model realistic spoof presentations, the generative distribution is conditioned on the textural and optical properties of the COLFISPOOF database, which focuses on contactless acquisition.

We define 72 distinct material categories based on the physical substrates used in presentation attack fabrication. These include various silicone-based compositions (e.g., Dragonskin variants with different pigments), moldable compounds (e.g., Play-Doh, Silly Putty, modeling clay), and office-based materials (e.g., various paper printouts and wood glue). All samples are standardized to a spatial resolution of 512×512 pixels and normalized to the range $[-1, 1]$ to ensure numerical stability during training.

To improve generalization, a two-level data augmentation strategy is employed. Offline horizontal flipping increases the effective dataset size, while online adaptive augmentation is applied during training via the ADA mechanism. The augmentation strength is dynamically adjusted based on discriminator feedback, preventing overfitting on small material-specific subsets without compromising the fidelity of the ridge micro-topology.

3.3. Synthetic Fingerprint Generation

3.3.1. Implementation details and hyperparameters

The generative model is implemented using the StyleGAN2-ADA architecture, which decomposes image synthesis into an 8-layer mapping network and a synthesis network. The mapping network projects a latent vector $z \in \mathscr{Z}$, sampled from a Gaussian distribution $\mathscr{N}(0, I)$, into an intermediate latent space $w \in \mathscr{W}$. The synthesis network then progressively generates fingerprint images from a 4×4 constant input up to the final 512×512 resolution.

Style modulation and stochastic noise injection are applied at each resolution level, enabling disentangled control over global ridge flow and fine-grained texture characteristics. Model optimization is performed using a non-saturating adversarial loss with R1 regularization, where the penalty strength is set to $\gamma = 10$.

To ensure reproducibility, we specify the following training configurations: the model was trained with a constant learning rate of 0.0025 and a batch size of 32 using the Adam optimizer ($\beta_1 = 0, \beta_2 = 0.99$). Training was conducted for a total of $15,000$ kimg (thousands of real images shown to the discriminator) until the Fréchet Inception Distance (FID) [15] reached convergence. The Adaptive Discriminator Augmentation (ADA) was initialized with a target probability of 0.6 to stabilize training under limited data. All experiments were executed on an NVIDIA A100 GPU, with an approximate training time of 18 hours. Finally, an exponential moving average (EMA) of the generator parameters was maintained and utilized during inference to maximize sample quality and consistency.

3.3.2. Latent space control for live and spoof variations

The intermediate latent space provides a semantically disentangled representation that supports controlled manipulation of fingerprint attributes.

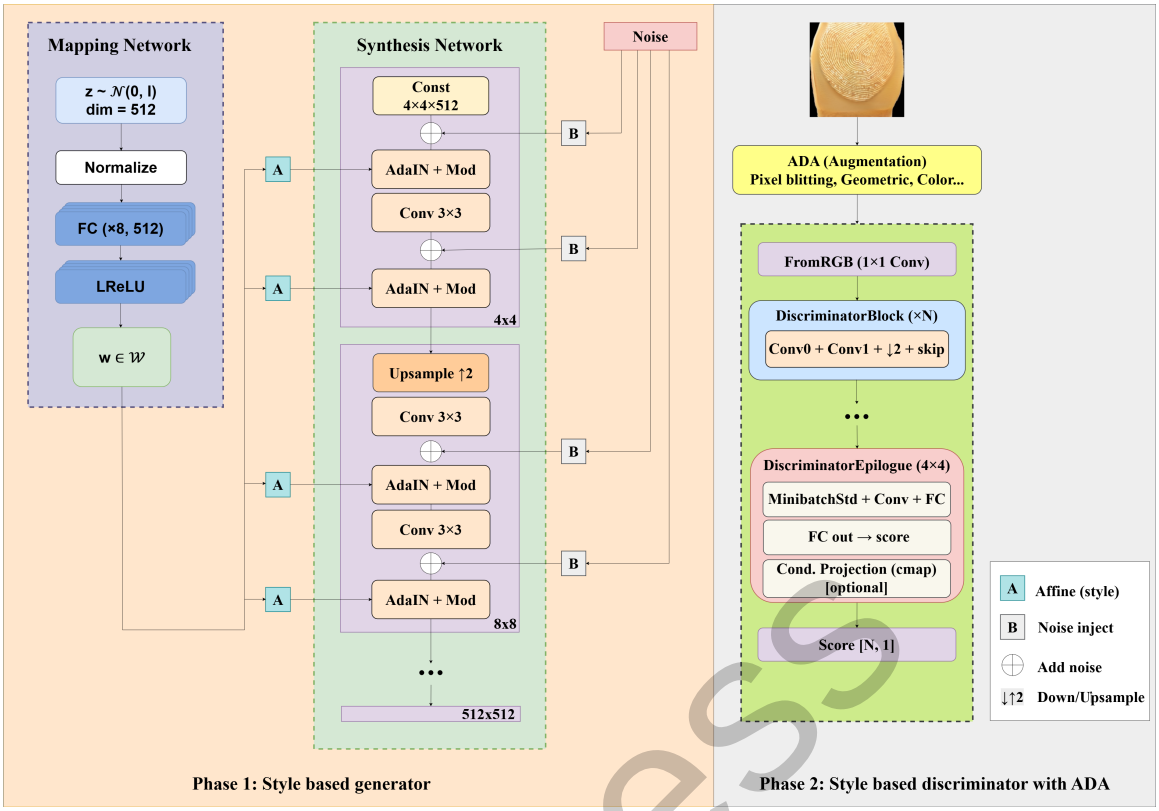

Fig. 1. Overview of the StyleGAN2-ADA architecture used for synthetic fingerprint generation. Phase 1 (left): The style-based generator transforms latent code $z \sim \mathcal{N}(0,I)$ into style space $w \in \mathcal{W}$ via the Mapping Network, then progressively synthesizes fingerprint images from 4×4 to 512×512 resolution. Phase 2 (right): The discriminator employs Adaptive Discriminator Augmentation (ADA) to prevent overfitting on limited biometric datasets

Identity-related characteristics, such as ridge flow and minutiae-like structures, are predominantly governed by coarse and intermediate synthesis layers, whereas material-dependent properties, including surface texture and noise patterns, are encoded in finer layers.

Live and spoof fingerprint variations are generated through a combination of random latent sampling, truncation, and style mixing. This design enables the synthesis of multiple material-specific spoof impressions while maintaining identity consistency, a requirement that is essential for realistic FPAD data augmentation.

3.4. Labeling and Quality Control

All generated fingerprints are subjected to an automated, rule-based quality-control stage. Rather than computing a learned quality score such as NFIQ2 [16], this stage applies explicit structural plausibility criteria that are informed by the same design principles that motivate NFIQ2, namely ridge-structure consistency and adequate fingerprint-area coverage. A generated sample is discarded only when it fails one of these criteria, for example owing to insufficient foreground or ridge area, or to a loss of coherent ridge continuity. The same criteria are applied automatically and uniformly across all 72 material categories through a single pipeline, with no per-material manual intervention, so that the quality-control stage does not introduce a category-dependent selection bias.

The stage is intentionally conservative: its purpose is to remove only structurally degenerate generations rather than to curate the distribution, since aggressive filtering toward high apparent quality would bias the synthetic set away from the realistic and more challenging spoofs that are central to our study. Material realism is additionally assessed using texture-based descriptors to confirm that retained samples remain within plausible biometric and material distributions. Samples satisfying all criteria are retained and labeled according to their corresponding material categories, and the screening implementation is released together with the dataset and code so that its behaviour is fully reproducible.

4. Experiments

4.1. Experimental Setup

To address the limited material diversity of existing spoof datasets, we used StyleGAN2-ADA to generate 7,200 synthetic fingerprint images across 72 material categories, including silicone-based materials (for example Dragonskin variants), moldable compounds (for example Play-Doh, Silly Putty), and paper-based

printouts. This material categories coverage is broader than that of current public datasets. Moreover, unlike physical spoof samples, synthetic samples can be generated on demand and shared more easily without violating privacy requirements. In the unseen-material protocol, materials are sorted by labels. The test part has categories without labels that are found in the training part. This checks how well it can handle new attacks not seen during training. Fig. 2 provides representative examples from different categories.

To examine whether synthetic data can improve PAD performance, especially for unseen attack types, we designed three training configurations:

Dataset-A combines live fingerprint images from LivDet with synthetic spoof images generated by CycleGAN. It contains 5,040 images from 24 material classes (approximately 210 images per material).

Dataset-B combines live fingerprint images from LivDet with synthetic spoof images generated by StyleGAN2-ADA. It contains 7,200 images from 72 material classes (approximately 100 images per material).

Dataset-C combines live fingerprint images from LivDet with authentic spoof images from COLFISPOOF, and serves as the real-data baseline.

Note that while Dataset-B contains more total images than Dataset-A, its per-material sample density is lower due to the deliberate prioritization of material diversity (72 vs. 24 categories). This design choice reflects the primary objective of maximizing coverage across spoofing materials rather than maximizing per-class sample size.

For a fair comparison, all three configurations were evaluated using the same PAD model: ResNet-50 initialized with ImageNet-pretrained weights and fine-tuned for 50 epochs using binary cross-entropy loss with a batch size of 32. The evaluation protocol followed a stratified split: 80% of samples for training, 10% for validation, and 10% for testing. The test set contained materials deliberately excluded from the training partition to simulate the unseen material challenge.

4.2. PAD Performance

Table 3 summarizes the classification results. Dataset-A (CycleGAN) achieved the highest overall performance, with an F1 score of 0.9589 and an AUC of 0.9788. Dataset-C (real COLFISPOOF spoofs) followed with an F1 of 0.9091, while Dataset-B (StyleGAN2-ADA) attained an F1 of 0.8980 and an AUC of 0.9561.

The superior in-distribution accuracy of Dataset-A is attributable to the nature of CycleGAN's paired style transfer: by directly transforming real live fingerprint images into spoof-style outputs, CycleGAN preserves the underlying ridge structure while applying material-specific appearance changes. This produces a sharp, well-defined live/spoof boundary that a classifier can exploit efficiently. However, this advantage comes at a cost: CycleGAN-generated spoofs remain structurally anchored to their source live images, limiting the diversity of attack patterns the model encounters during training.

In contrast, StyleGAN2-ADA generates spoof samples from a learned latent distribution without relying on direct image-to-image translation. As a result, the generated samples show greater variation within the spoof class and more realistic material-specific textures, as presented in Tables 5 and 6. Because these synthetic spoofs are closer to real spoofs, it is more difficult for a moderate-capacity classifier such as ResNet-50 to separate from live samples. Therefore, the lower in-distribution accuracy of Dataset-B should not be interpreted as weaker synthetic data quality, but rather as evidence that Dataset-B better captures the practical difficulty of the PAD task. Data quality and classification accuracy do not measure the same thing. The purpose of this experiment is to test whether greater distributional realism of the synthetic spoofs improves operational security rather than only in-distribution accuracy. Using the security metrics of ISO/IEC 30107-3 [1] in Table 4, we compare three detectors trained respectively on Dataset-A (CycleGAN), Dataset-B (StyleGAN2-ADA), and Dataset-C (real COLFISPOOF spoofs) under the unseen-material protocol, and we additionally report the same metrics on the original full COLFISPOOF test set as a real-world difficulty reference. We emphasize that TAR@FAR is reported here as a PAD-classifier operating point, with FAR denoting the proportion of spoof samples incorrectly accepted as live; this differs from the matcher-based TAR@FAR convention used in synthetic-fingerprint identity-verification studies such as PrintsGAN and SpoofGAN, where FAR instead denotes the rate of false matches across distinct identities.

Under the unseen-material protocol, Dataset-B attains the lowest APCER (2.06%) and ACER (5.49%) among the three configurations, despite its lower in-distribution F1, compared with 3.92% and 6.73% for Dataset-A, respectively. Relative to the authentic Dataset-C, Dataset-B similarly achieves lower APCER and ACER (2.45% and 6.01%, respectively); however, Dataset-C attains a lower EER (5.22% versus 6.10% for Dataset-B), indicating that the security advantage of Dataset-B over the real-data baseline is confined to the APCER and ACER operating points rather than uniform across all metrics. At a fixed false accept rate of 1 percent, Dataset-B achieves the highest true accept rate of 52.2 percent, compared with 20.27 percent for Dataset-A (41.31 percent for Dataset-C). Because APCER is the fraction of attacks wrongly accepted as genuine, the detector trained on the more realistic data is clearly more resistant to unseen attacks, whereas the high in-distribution F1 of Dataset-A reflects

(a) DS Brown (b) DS Brown Darkened (c) DS Brown Transparent (d) DS Dark Red (e) Gelafix

(f) Mouldable (g) Woodglue (h) Knetosil45 (i) Playdoh Pink (j) Printout Blue

Fig. 2. Representative synthetic spoof fingerprint samples generated by StyleGAN2-ADA across ten material categories, illustrating material-specific variations in color, surface reflectance, and ridge texture characteristics

overfitting to a narrow band of attack appearances whose CycleGAN spoofs remain anchored to their source live images.

On the original full COLFISPOOF test set, all error rates rise substantially, reaching an APCER of 23.88 percent and an ACER of 16.70 percent. This is expected, since the original set is larger and more heterogeneous than the held-out unseen-material partition, and it provides a conservative estimate of the gap that remains between synthetic training data and unconstrained real deployment. Taken together, these results support the central premise of this study, namely that greater distributional realism yields measurable security generalization to unseen attacks despite lower in-distribution separability, and they directly address the concern that image-quality metrics alone cannot establish attack effectiveness.

To verify that the detector generalizes beyond the contactless COLFISPOOF-derived domain, we additionally evaluate the Dataset-B detector on a contact-based LivDet-only test partition (bona fide and spoof samples from LivDet 2009), reported in Table 2. On this distinct sensor domain it attains an APCER of 0.45%, a BPCER of 9.53%, an ACER of 4.99%, an EER of 3.10%, and an AUC of 0.9893. The low APCER confirms that the detector remains resistant to attacks in a sensor domain different from the contactless source distribution, providing direct cross-sensor support for the generalization claim. The substantially higher error on the heterogeneous combined partition (APCER 23.88%, ACER 16.70%) is retained as a conservative real-world difficulty reference.

Table 2. Cross-sensor evaluation of the Dataset-B detector. Lower APCER, BPCER, ACER, EER are better; higher AUC and TAR are better

Metric	**LivDet only**	**LivDet + COLFIS-POOF**
Accuracy (%)	94.34	80.34
APCER (%)	0.45	23.88
BPCER (%)	9.53	9.53
ACER (%)	4.99	16.70
EER (%)	3.10	13.33
AUC	0.9893	0.9092
TAR@FAR=1% (%)	58.57	12.18

Table 3. In-Distribution Classification Performance of PAD Models Trained on Different Synthetic and Real Datasets

Metric	**Dataset-A (CycleGAN)**	**Dataset-B (StyleGAN2)**	**Dataset-C (COLFISPOOF)**
F1 Score	0.9589	0.8980	0.9091
AUC	0.9788	0.9561	0.9632
Precision	0.9596	0.9016	0.8333
Recall	0.9531	0.8988	0.8897

Table 4. FPAD Performance under ISO/IEC 30107-3 Metrics. Lower APCER, BPCER, ACER and EER are better, while higher TAR is better. Note that TAR@FAR here denotes the PAD classifier's operating point, where FAR refers to the rate at which spoof samples are incorrectly accepted as live; this usage is distinct from the matcher-based TAR@FAR reported in identity-verification studies (e.g., PrintsGAN, SpoofGAN), where FAR denotes false matches between different identities

Metric	Dataset-A (CycleGAN)	Dataset-B (StyleGAN2)	Dataset-C (real COLFISPOOF)	Original (COLFISPOOF)
APCER (%)	3.92	2.06	2.45	23.88
BPCER (%)	9.53	8.93	9.56	9.53
ACER (%)	6.73	5.49	6.01	16.70
EER (%)	7.96	6.10	5.22	13.33
TAR@FAR=1% (%)	20.27	52.2	41.31	12.18

To better understand these classification results, we next analyze the quality of the synthetic data. The generation quality metrics in Section 4.3 therefore complement, rather than contradict, the classification results.

4.3. Synthetic Data Quality

To rigorously assess the fidelity and distributional alignment of the synthetic datasets with real fingerprint data, a suite of complementary metrics was employed, covering distributional congruence (FID, KID), perceptual realism (LPIPS), pixel-level fidelity (SSIM, PSNR), and structural integrity (gradient and edge statistics).

4.3.1. Distributional and perceptual fidelity

Metrics such as FID and KID measure how close synthetic images are to real data. If synthetic spoof images look more like real spoofs, they can also become harder for a classifier to separate from live fingerprints. This is especially true when the classifier is not very large and the training data is limited. In contrast, a model like CycleGAN changes real live fingerprints into spoof-like images, so the generated samples still keep much of the original fingerprint structure. This can make live and spoof images easier to separate, which can improve classification accuracy, but it does not fully reflect the real difficulty of PAD. In real applications, a PAD system must detect spoof attacks that look as realistic as possible. From this view, more realistic synthetic data is more useful, even if it leads to slightly lower classification accuracy.

Dataset-B (StyleGAN2-ADA) achieved an FID of 63.37, much lower than Dataset-A (CycleGAN), which had 261.74. We interpret this value in relative rather than absolute terms: it indicates that, among the two generators compared, the synthetic images of Dataset-B are closer to the real fingerprint distribution than those of Dataset-A, rather than asserting state-of-the-art absolute photorealism. This relative ordering is corroborated by the KID, LPIPS, SSIM, PSNR, and gradient/edge statistics reported below, all of which point in the same direction.

The KID results showed the same trend. Dataset-B achieved 0.0354 ± 0.0011, while Dataset-A had 0.1778 ± 0.0068. The lower value and lower variance of Dataset-B suggested that it produced more stable and more realistic samples across the dataset.

LPIPS results also supported this finding. Dataset-B achieved an intra-class LPIPS of 0.2472 ± 0.1296 and an inter-class LPIPS of 0.6031 ± 0.1266. This means that samples from the same material still had enough variation, while samples from different materials remained clearly different. Its diversity ratio was 2.44×, compared with 2.34× for Dataset-A, showing that Dataset-B gave a better balance between variation within each material and separation across materials. The comparative results are summarized in Table 5.

4.3.2. Structural and textural integrity

As presented in Table 5, dataset-B achieved an SSIM of 0.4824 ± 0.2027 and a PSNR of 11.56 dB, outperforming Dataset-A (SSIM: 0.3804 ± 0.3106, PSNR: 9.69 dB). Although the absolute values are not high, this is expected in generative tasks. The goal of these tasks is to realistic samples rather than exact pixel-level reconstruction. Nonetheless, the higher SSIM and PSNR of Dataset-B suggest that it preserves fingerprint ridge patterns and local contrast better than Dataset-A.

Gradient and edge analysis showed the same pattern. Dataset-B had an average gradient magnitude of 4.08 and an edge density of 0.0714, closely aligned with the real COLFISPOOF reference values of 3.22 and 0.0507. In contrast, Dataset-A had lower values of 1.85 and 0.0184, indicating smoother transitions and weaker ridge definition. These results suggested that StyleGAN2-ADA, with its noise injection and multi-resolution synthesis, can generate the fine texture details needed for realistic fingerprint images. Overall, Dataset-B was numerically closer to the real dataset in both gradient magnitude and edge density.

4.4. Limitations and Future Work

While the proposed synthetic dataset demonstrates substantial improvements, several limitations remain. First, the current 72 material categories do not yet

Table 5. Distributional and perceptual quality metrics

Metric	**Dataset-A (CycleGAN)**	**Dataset-B (StyleGAN2)**
FID	261.74	63.37
KID	0.1778 ± 0.0068	0.0354 ± 0.0011
LPIPS (Intra-class)	0.3037 ± 0.2457	0.2472 ± 0.1296
LPIPS (Inter-class)	0.7110 ± 0.3127	0.6031 ± 0.1266
Diversity Ratio	$2.34\times$	$2.44\times$
SSIM	0.3804 ± 0.3106	0.4824 ± 0.2027
PSNR (dB)	9.69	11.56

Table 6. Structural and gradient-based quality metrics

Dataset	**Avg. Gradient Magnitude**	**Avg. Edge Density**
Dataset-A (CycleGAN)	1.85	0.0184
Dataset-B (StyleGAN2)	4.08	0.0714
Dataset-Real (COLFIS-POOF)	3.22	0.0507

cover some new attack types, such as 3D-printed molds, transparent conductive gels, and hybrid silicone composites. Future study should include these materials, ideally with support from forensic laboratories. Second, as the synthetic data is conditioned on contactless COLFISPOOF imagery, a domain gap persists relative to contact-based optical and capacitive sensors. Thus, future work should explore domain adaptation or multi-modal synthesis frameworks to bridge this gap. Third, the dataset represents a static snapshot of each material, omitting temporal degradation effects and multi-session environmental variability that characterize real-world deployment. Incorporating time-dependent degradation models would enhance operational realism. Finally, early experiments with diffusion models have shown promising FID scores below 50, but their high inference cost makes large-scale use difficult. A hybrid architectures combining StyleGAN's latent disentanglement with diffusion-based refinement represent a compelling direction for generating pixel-level realistic fingerprints. In this regard, our current evaluation contrasts a paired style-transfer generator (CycleGAN), an unconditional style-based generator (StyleGAN2-ADA), and a real-data reference (COLFISPOOF); a broader empirical comparison against more recent generative approaches, including diffusion-based models, is therefore an important direction for future work once their inference cost becomes practical for large-scale material-specific synthesis. We further note that the unseen-material categories are held out entirely from training and evaluated only at test time, so there is no random train/test split of these categories to repeat over; robustness is instead assessed by evaluating the same detector across distinct test distributions, namely the unseen-material partition, the contact-based LivDet-only partition (Table 2), and the full COLFISPOOF set, with the principal ordering preserved across all three. A related caveat concerns the comparison between Dataset-A and Dataset-B: the two configurations differ not only in material diversity (24 vs. 72 categories) but also in generative architecture (CycleGAN paired style transfer vs. StyleGAN2-ADA unconditional latent sampling), so the observed security advantage of Dataset-B cannot, on the present evidence, be attributed to material diversity alone rather than to this architectural difference. A formal analysis reporting mean $\pm$ std over multiple seeds, together with a matched-class ablation that holds the generator fixed while varying material count and per-class density, is identified as future work; this ablation should additionally include a matched-architecture arm (e.g., StyleGAN2-ADA restricted to 24 materials) so that the architectural and diversity factors can be disentangled. Addressing these limitations will be important for improving the robustness and transferability of synthetic FPAD datasets.

5. Conclusions

This paper presented an empirical study on generating material-diverse synthetic fingerprints for FPAD. By synthesizing 7,200 images across 72 material categories using StyleGAN2-ADA, we demonstrated that high-fidelity synthetic data can achieve performance comparable to, and on security-critical metrics even exceed, real spoof datasets, while providing a privacy-conscious alternative to manual collection. Our evaluation using security-critical metrics (APCER, ACER) confirms that the StyleGAN2-ADA-based synthesis pipeline employed in this study, which jointly varies material diversity and generative architecture relative to the CycleGAN baseline, offers a more rigorous training signal for unseen-material generalization than the paired style-transfer approach. As Dataset-A and Dataset-B differ concurrently in material diversity and in generative architecture, this

advantage is attributed to the synthesis pipeline as a whole rather than to either factor in isolation; disentangling the two, via a matched-architecture, matched-class ablation, is identified as future work. Future work will also focus on bridging the domain gap between contactless and contact-based sensors using multi-modal generative frameworks.

6. Data Availability

The synthetic fingerprint dataset generated and analysed in this study is publicly available at this repository or via DOI: 10.6084/m9.figshare.32754600. The generation, quality-control, and evaluation code, together with the training configuration files, split definitions, and random seeds, are available from the corresponding author upon reasonable request. The LivDet 2009 and COLFISPOOF source datasets are available from their respective original publications.

7. Acknowledgments

This research was funded by the Ministry of Education and Training of Vietnam under project code CT2025.EA.BKA.02.

References

[1] I. Goicoechea-Telleria, M. B. Fernandez-Saavedra, J. Liu-Jimenez, and R. Sanchez-Reillo, An evaluation of presentation attack detection of fingerprint biometric systems applying ISO/IEC 30107-3, in International Biometric Performance Conference, 2016.

[2] G. L. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli, and S. Schuckers, First international fingerprint liveness detection competition—LivDet 2009, pp. 12–23, 2009. https://doi.org/10.1007/978-3-642-04146-4_4

[3] G. Orrù, M. Micheletto, R. Casula, S. Zedda, D. Fenu, L. Igene, J. Priesnitz, C. Busch, C. Rathgeb, S. C. Schuckers, and G. Luca Marcialis, LivDet2025: Toward Robust and Generalizable Fingerprint Presentation Attack Detection, pp. 1–9, 2025. https://doi.org/10.1109/IJCB65343.2025.11411004

[4] 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, Jan. 2023, pp. 653–661

[5] G. L. Marcialis, F. Roli, and A. Tidu, Analysis of fingerprint pores for vitality detection, in 2010 20th international conference on pattern recognition, Aug. 2010, pp. 1289–1292.

[6] R. K. Dubey, J. Goh, and V. L. Thing, Fingerprint liveness detection from single image using low-level features and shape analysis, IEEE transactions on information forensics and security, vol. 11, no. 7, pp. 1461–1475, Jul. 2016. https://doi.org/10.1109/TIFS.2016.2535899

[7] R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado, Fingerprint liveness detection using convolutional neural networks, IEEE transactions on information forensics and security, vol. 11, no. 6, pp. 1206–1213, Jan. 2016. https://doi.org/10.1109/TIFS.2016.2520880

[8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial nets, Advances in neural information processing systems, vol. 27, 2014.

[9] P. Bontrager, A. Roy, J. Togelius, N. Memon, and A. Ross, Deepmasterprints: Generating masterprints for dictionary attacks via latent variable evolution, in 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Oct. 2018, pp. 1–9.

[10] P. K. Diederik and W. Max, An introduction to variational autoencoders, Foundations and Trends® in Machine Learning, vol. 12, no. 4, pp. 307–392, Nov. 2019. https://doi.org/10.1561/2200000056

[11] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, Analyzing and improving the image quality of stylegan, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Jun. 2020, pp. 8110–8119.

[12] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, Training generative adversarial networks with limited data, Advances in neural information processing systems, vol. 33, pp. 12104–12114, Jun. 2020.

[13] S. Minaee and A. Abdolrashidi, Finger-GAN: Generating realistic fingerprint images using connectivity imposed GAN, arXiv preprint arXiv:1812.10482, Dec. 2018. https://doi.org/10.48550/arXiv.1812.10482

[14] J. J. Engelsma, S. Grosz, and A. K. Jain, PrintsGAN: Synthetic fingerprint generator, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 5, pp. 6111–6124, Sep. 2022. https://doi.org/10.1109/TPAMI.2022.3204591

[15] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium, Advances in neural information processing systems, vol. 30, 2017.

[16] E. Tabassi, M. Olsen, O. Bausinger, C. Busch, A. Figlarz, G. Fiumara, O. Henniger, J. Merkle, T. Ruhland, C. Schiel, and M. Schwaiger, NIST Fingerprint Image Quality 2, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD, July 13, 2021. [Online]. Available: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=920087.