SynthASpoof: Developing Face Presentation Attack Detection Based on Privacy-friendly Synthetic Data

Meiling Fang[1,2], Marco Huber[1,2], Naser Damer[1,2]
[1]Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany
[2]Department of Computer Science, TU Darmstadt, Darmstadt, Germany
Email: meiling.fang@igd.fraunhofer.de

Abstract

Recently, significant progress has been made in face presentation attack detection (PAD), which aims to secure face recognition systems against presentation attacks, owing to the availability of several face PAD datasets. However, all available datasets are based on privacy and legally-sensitive authentic biometric data with a limited number of subjects. To target these legal and technical challenges, this work presents the first synthetic-based face PAD dataset, named SynthASpoof, as a large-scale PAD development dataset. The bona fide samples in SynthASpoof are synthetically generated and the attack samples are collected by presenting such synthetic data to capture systems in a real attack scenario. The experimental results demonstrate the feasibility of using SynthASpoof for the development of face PAD. Moreover, we boost the performance of such a solution by incorporating the domain generalization tool MixStyle into the PAD solutions. Additionally, we showed the viability of using synthetic data as a supplement to enrich the diversity of limited authentic training data and consistently enhance PAD performances. The SynthASpoof dataset, containing 25,000 bona fide and 78,800 attack samples, the implementation, and the pre-trained weights are made publicly available [1].

1. Introduction

Due to its outstanding performance, face recognition has been widely used in various aspects of our daily lives, such as access control, phone unlocking, and mobile payments. However, face recognition is vulnerable to presentation attacks (PAs) including print attacks, video replay attacks, and 3D mask attacks [1,7,32,45]. Therefore, face presentation attack detection (PAD), referring to the process of identifying whether a face presented to the system is a bona fide (live) or PA (spoof), is essential to secure face recognition from PAs.

With the advancements in deep learning technology, face PAD algorithms have made great progress. One of the main contributors to this advance is the face PAD datasets [1, 7, 29, 44, 45]. However, these datasets utilized for developing data-driven PAD solutions are built on authentic biometric data, which might raise ethical and legal challenges. This concern has recently been discussed in both the face recognition [31] and face morphing attack detection [10] communities. Given the legal privacy regulations, the collection, use, share, and maintenance of face data for biometric processing is extremely challenging [8]. For example, several large-scale face recognition datasets, such as VGGFace2 [6], MS-Celeb-1M [19], and MegaFace [26], were withdrawn by their creators with privacy and proper subjects consent issues being the main drive. One of the main candidate solutions for this issue is the use of synthetic data [8]. This has been very recently and successfully proposed for the training of face recognition [4, 5, 31] and morphing attack detection [10, 12, 22], among other processes such as model quantization [2]. Synthetic data for PAD development has, besides the privacy and legal motivations, a major advantage when it comes to scale and diversity. While most existing face PAD datasets are of a small-scale with a limited number of subjects, creating a synthetic-based PAD development data enables to produce a large-scale dataset in terms of both, the number of samples and the number of different faces.

Motivated by the legal and ethical challenges in using, sharing, and collecting authentic biometric data along with the limitation in the scale and diversity in existing datasets, this work poses the question of "can synthetic data be used for the development of face PAD?". This is based on our assumption that learning to detect the differences between bona fide and attack samples of a synthetic origin can be used to detect these differences between authentic bona fide and attacks, and thus perform PAD. Towards that, we introduce **the first privacy-friendly synthetic-based face PAD (Anti-Spoofing) dataset, SynthASpoof**, consisting of

[1]https://github.com/meilfang/SynthASpoof.git

25,000 bona fide and 78,800 attack samples. The bona fide samples are created by using StyleGAN2-ADA [24], while the attack samples are collected by presenting these synthetic samples as printed/replayed attacks to varied capture sensors. Samples of the SynthASpoof are shown in Fig. 1. Based on the SynthASpoof dataset, we then conduct extensive experiments to **explore the feasibility of using synthetic data for the development of face PADs**. Subsequently, we propose to **adapt MixStyle [46] to enhance the generalizability of models trained on SynthASpoof**. Furthermore, we **successfully propose supplementing the authentic training data with the synthetic SynthASpoof to achieve even higher PAD performances**.

2. Related Work

In the last decade, many face PAD datasets [1, 7, 28, 29, 32, 39, 44, 45] have been collected and made available to support the development of PAD algorithms. The face PAD datasets can be broadly categorized into four groups based on the type of attacks and sensors: multi-modal 3D attacks [18], multi-modal 2D attacks [32, 43], single-modal 3D attacks [28], and single-modal 2D attacks [1, 7, 39, 45].

Multi-modal datasets [18, 43] used multiple sensors in addition to visible cameras, such as depth and infrared cameras, providing more options for face PAD solutions. However, such datasets have limitations in real-world deployment due to the cost of sensors and computation resources. 3D attacks are more realistic than traditional 2D attacks. The HiFiMask [28] dataset is the largest and most recent 3D face mask PAD dataset, collected from 75 subjects and including three mask attacks. However, HiFiMask dataset has a limited number of subjects and mask materials due to the higher cost of 3D mask creation compared to 2D attacks. Most 2D face PAD datasets (as seen in Table 1) are outdated due to their acquisition equipment and have limited numbers of subjects and samples, leading to a potential over-fitting risk. SiW-M [29], PADISI-Face [32], and CelebA-Spoof [44] are relatively up-to-date and large-scale datasets, where CelebA-Spoof and part of the SiW-M dataset were collected from the web. Many of the existing face PAD datasets have ethical and legal issues that limit their public availability and raise concerns about sharing and reusing biometric information of individuals, driving some researches to keep their developed datasets private [13, 15]. For example, SiW-M is currently inaccessible. In addition to privacy issues, CelebA-spoof has several limitations: 1) numerous label noise, 2) low quality attack samples, which contradict the fact that attackers commonly use highly sophisticated artifacts to maximize their impersonation success probability, and 3) no consent of all involved individuals.

Overall, existing face PAD datasets have two primary limitations. First, the collection, use, and share of such data pose ethical and legal challenges [8]. Second, the scale of the existing face PAD datasets may not be sufficient to develop over-parameterized deep learning based PAD solutions. This highlights the need for face PAD development datasets that prioritize the privacy of individuals, the shareability of data in the research community, and the reproducibility and continuity of face PAD research. To address these concerns, we propose the use of synthetic data for the development of face PAD. Our synthetic SynthASpoof data contains of 25,000 bona fide samples and 78,800 attack samples (details in Section 3).

3. SynthASpoof Dataset

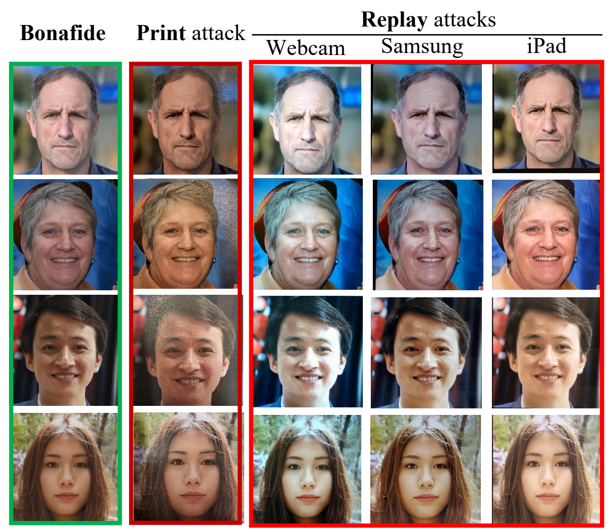


Figure 1. Samples of the SynthASpoof dataset. The left column shows bona fide samples. The second to last column show different attack samples collected from the corresponding bona fide images. In the case of replay attacks, three sensors (webcam, Samsung phone, and iPad) were used to capture the attacks displayed on different screens.

Despite the significance of publicly available datasets in promoting the progress of face PAD and being valuable sources for the research community, legal, ethical and privacy concerns, as well as the limited size and diversity of the datasets pose challenges to the development of generalized PAD solutions.

This section introduces our SynthASpoof dataset (samples are shown in Fig. 1), which is the first synthetic-based face PAD dataset. The dataset is built based on the image synthesis and selection procedure presented in [10]. To follow realistic attack scenarios appearing in authentic data attacks, the attack samples are created based on the synthetic bona fide data by presenting printed and replayed images to capture sensors. This aims at fulfilling our assumption that the difference between authentic bona fide and authentic-

Dataset	Year	# Bona fide/attack	# Sub	Attack types
CASIA-FASD [45]	2012	150 / 450 (V)	50	1 Print, 1 Replay
Replay-Attack [7]	2012	200 / 1,000 (V)	50	1 Print, 2 Replay
MSU-MFSD [39]	2015	70 / 210 (V)	35	1 Print, 2 Replay
OULU-NPU [1]	2017	1,980 / 3,960 (V)	55	2 Print,2 Replay
SiW-M [29]	2019	660 / 968 (V)	493*	1 Print, 1 Replay, 5 3D Mask, 3 Make Up, 3 Partial
CelebA-Spoof [44]	2020	184,407 / 377,168(I)	10,177*	3 Print, 3 Replay, 1 3D, 3 Paper Cut
PADISI-Face [32]	2021	1,105 / 924 (V)	360	1 Print, 4 Mask, 1 Makeup, 1 Tattoo, 2 Partial
SynthASpoof	2023	25,000 / 78,800 (I&V)	25,000	1 Print, 3 Replay

Table 1. Summary of the public face PAD datasets. V and I are shorthand for video and image, respectively. Subject number with '*' denotes the subjects are partially or all from the web. Note the limited scale of most datasets and the fact that the larger ones are based on web-collected images.

based attacks induced by the attack process can also be induced by the same attack process using synthetic data. Thus, learning to detect this difference on synthetic data will enable detecting it in authentic-based attacks.

Bona Fide Samples: First, 125,000 images were created by using the StyleGAN2-ADA [24] trained on Flickr-Faces-HQ dataset (FFHQ) [25]. The pretrained model produced a synthetic face data for each latent vector that was randomly non-repeatedly generated based on Gaussian noise. These images were then filtered automatically by using the CR-FIQA [3] face image utility assessment approach, where extreme non-frontal poses and largely occluded images were mostly removed by removing the images with the lowest utility score. This helps simulate the real log-in face recognition scenario that is commonly targeted by PAs. Finally, SynthASpoof contains 25,000 bona fide samples.

Attack Samples: SynthASpoof contains two attack types, print and replay attacks. For the print attacks 3,800 videos of distinct synthetic subjects were captured using a Samsung Galaxy Tablet S6. For the more challenging replay attack, we introduce diverse display and capture setups. First, attacks displayed on a MacBook Air 2020 screen were captured using both a Samsung Galaxy A71 and an iPad Pro 10.5 (both with a resolution of 1920×1080). Additionally, attacks displayed on a Dell UltraSharp 24 display were captured using a Creative Labs webcam with a resolution of 720×480. All the 25,000 images were used as an attack on each setup resulting in a total of 75,000 replay attack clips. All attack captures (print and replay) were cropped so that they do not include any region outside of the displayed attack image (e.g. screen border). All captured attacks are videos with a duration ranging from 3 to 5 seconds, from each of these videos the single frame in the middle of the video is also extracted as a single image attack used in our training. Both, the videos and the images of the SynthaSpoof dataset are publicly released and can be used to develop PAD solutions based on synthetic data.

Comparing to existing face PAD datasets, the proposed SynthASpoof dataset provides three advantages: **1) Privacy-friendly:** SynthASpoof is the first synthetic face PAD dataset which relaxes the pure dependence on the legally and ethically challenging use of authentic development data. **2) Large-scale and high-quality samples**: As discussed in Section 2, most existing datasets are of small scale, the only relatively larger and diverse dataset is the CelebA-Spoof [44]. However, some of its bona fide samples might be falsely annotated and many of the attack samples exhibit severe distortion and low quality, as they were collected from the web (which is a orivacy issue by itself) without proper control or post-processing checks. In contrast, the bona fide samples in SynthASpoof were checked by face image quality control and the attack samples were collected in a controlled manner to reflect the fact that attackers usually use highly sophisticated artifacts to maximize their success in impersonation. **3) Extensibility:** researchers can build subsequent synthetic-based face PAD datasets by increasing the diversity of attack types.

4. PAD Solutions

To assess the suitability of using SynthASpoof for the development of face PAD, we adopt two of the commonly used face PAD backbones, ResNet [20] and PixBis [17]. The selection of these two backbones was based on their wide use, representing two common supervision strategies in face PAD, and reported good performance in previous studies [14, 17, 20, 41].

4.1. Base Presentation Attack Detectors

ResNet [20] is one of the most popular backbone architectures used in face PAD algorithm design [11, 14, 41, 44]. We report the results of a model trained from scratch based on the ResNet-18 model architecture. A cross-entropy loss function is used in the training phase and formulated as follows:

$$\mathcal{L}_{CE} = -[y \cdot \log p + (1 - y) \cdot \log(1 - p)], \qquad (1)$$

where y is the ground truth (1 for bona fide and 0 for attack in our case) and p is the predicted score.

PixBis [17] employs a binary supervisory strategy at pixel-level to simplify the problem and obviate the need for a computationally intensive synthesis of depth maps. Two dense blocks of DenseNet121 [21] are utilized as the model backbone and a combination of two binary cross-entropy loss functions is used to train the model for both pixel-wise and binary output. The combined loss equation for the training of all models is formed as:

$$\mathcal{L}_{PixBis} = \mathcal{L}_{CE}^{pixel-wise} + \mathcal{L}_{CE}^{binary}. \quad (2)$$

where $\mathcal{L}_{CE}^{pixel-wise}$ refers to the loss based on the pixel-wise output and $\mathcal{L}_{CE}^{binary}$ refers to the loss based on the binary output.

4.2. MixStyle

Recent studies in synthetic-based face recognition revealed a domain gap between synthetic and authentic face images through the examination of the performance divergence between face recognition models trained on synthetic and authentic data. This performance gap has also been observed in our work, as will be detailed in Section 6.1. To narrow the domain gap between synthetic and authentic face PAD data, we adapt a recently proposed domain generalization method, MixStyle [46]. MixStyle mixes the feature statistics of two samples to synthesize novel domains inspired by the observation that the feature statistics encode style/domain-related information. To adapt the face PAD model from synthetic data to authentic data, we utilized the labeled synthetic SynthASpoof and unlabeled authentic face PAD data to perform MixStyle within a mini-batch and with a controlled probability during the training process.

Mathematically, the MixStyle adapted in our case can be formulated as follows:

$$\begin{aligned} \gamma &= \lambda\sigma(x_s) + (1-\lambda)\sigma(x_a) \\ \beta &= \lambda\mu(x_s) + (1-\lambda)\mu(x_a) \end{aligned} \quad (3)$$

Where x_s and x_a refer to synthetic and authentic face PAD data, respectively. $\lambda \in \mathbb{R}^B$ are weights sampled from the Beta distribution. The final mixed feature statistic is applied to the styled normalized synthetic face PAD data x_s as:

$$MixStyle(x_s) = \gamma \frac{x_s - \mu(x_s)}{\sigma(x_s)} + \beta \quad (4)$$

It is important to note that: 1) The loss is calculated only on the synthetic face PAD data output when conducting the domain adaptation experiment in Section 6.3. 2) In Section 6.4, the face PAD model is trained on a combination of the SynthASpoof and an authentic face PAD dataset to address the problem of limited training data. MixStyle is used there to reduce the difference between the synthetic and authentic training data, which is demonstrated later by the performance on unseen authentic data. 3) MixStyle is removed during the inference process, and thus does not require additional computational overhead while using the PAD.

In our experiment, MixStyle is inserted after the first and second ResNet blocks and after the first dense block of the PixBis model, based on the fact that features at higher layers have a stronger correlation with class labels as opposed to the domain information [46].

5. Experiments

5.1. Datasets

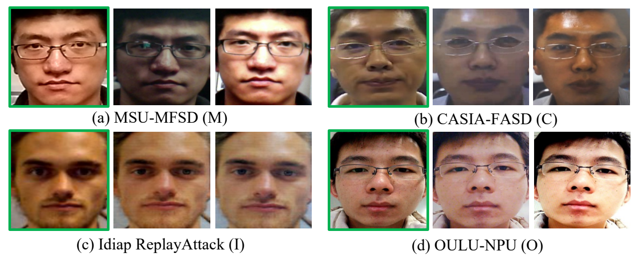


Figure 2. Samples of four authentic face PAD datasets. Images with green bounding box are bona fide, while others are attack samples.

To assess the feasibility of using SynthASpoof to develop face PAD, the performance of models trained on SynthASpoofis evaluated on four authentic face PAD benchmarks: MSU-MFSD [39] (denoted as M), CASIA-MFSD [45] (denoted as C), Idiap Replay-Attack [7] (denoted as I), and OULU-NPU [1] (denoted as O), which are widely used in PAD research [9, 14, 38, 47]. The data samples are shown in Fig. 2.

The **MSU-MFSD** [39] dataset consists of 440 videos captured from 35 subjects using two different resolutions of cameras. The dataset contains two types of attacks, printed photo attacks and replay attacks. The **CASIA-MFSD** [45] dataset is comprised of 600 videos from 50 subjects and includes three types of attacks: warped photo attack, cut photo attack, and video replay attack. The **Idiap Replay-Attack** dataset [7] contains 300 videos from 50 subjects captured under various sensors and illumination conditions. The dataset includes two attack types: print attacks and replay attacks. The **Oulu-NPU** [1] is a mobile face PAD dataset designed for assessing the generalizability of PAD methods in a realistic mobile scenario. OULU-NPU consists of 5940 video clips from 55 subjects using six different mobile phones.

The performance of models trained on synthetic and authentic face PAD data is analyzed over these datasets. As SynthASpoof database is specifically created for the purpose of training face PAD models, the entire dataset is used in the training phase. The trained models are then further tested on other authentic face PAD datasets.

5.2. Implementation Setup

Following [14, 17, 41] and to make up for the relatively smaller size of the authentic datasets, 25 frames (per video) were sampled evenly across the duration of each video in the four authentic face PAD datasets, while only one frame was considered from each video in the SynthASpoof database as detailed in Sec. 3 and Tab. 1. The faces were then detected and cropped using the MTCNN method [42] and resized to $224 \times 224 \times 3$ pixels, following [14,17,38,41,44]. During training, a weighted sampling was performed to insure a bona fide-attack ratio of 1:1. The Stochastic Gradient Descent (SGD) optimizer with a momentum of 0.9 and weight decay of 5e-4, and an exponential learning scheduler with a gamma of 0.998 was applied in all training processes. The initial learning rate for training the ResNet and PixBis models on the SynthASpoof database was set to 0.01. The batch size in the training phase was 128 and the training epoch was set to 70. Conventional data augmentation techniques: horizontal flipping, scaling and rotating, random gamma adjustment, RGB shifting, and color gittering, were used. The effects of these techniques are explored in Section 6.2. In the testing phase, a final PAD decision score of a video is a fused score (mean-rule fusion) of all frames, following [14, 17, 41].

5.3. Evaluation Metrics

Following existing cross-domain face PAD methods [14, 27, 33, 34], we report the Half Total Error Rate (HTER), which is the mean of Bona fide Presentation Classification Error Rate (BPCER) [23] and Attack Presentation Classification Error Rate (APCER) [23] and Area under the Receiver Operating Characteristic (ROC) Curve (AUC) value for cross-dataset face PAD evaluation. Additionally, ROC curves are illustrated, where the x-axis is APCER and the y-axis is 1-BPCER.

6. Results

We first explore if the synthetic data can be used to develop PAD solutions. We further compare the performance between models trained on the synthetic and the authentic data and tested on different authentic face PAD datasets. Then, we investigate the effect of cropping and data augmentation on PAD performance. Furthermore, we prove the sanity of adapting MixStyle to enhance the performance of PAD models trained on synthetic data. We also explore the usability of SynthASpoof as supplementary training data to enhance the diversity of authentic training data. Finally, a visual analysis and a final discussion is presetned.

6.1. SynthASpoof PAD

To study the feasibility of using SynthASpoof for developing face PAD solutions, we train the models on our SynthASpoof dataset and test on the different authentic datasets from real-world scenarios. We also train the models on the authentic data by following the cross-dataset (the evaluation data is unknown) evaluation protocols in recent face PAD works [38, 40, 47].

In these works [38, 40, 47], one face PAD dataset is used for the training and the remaining three datasets are separately used as testing data. Therefore, we conduct experiments upon the following 12 scenarios: C $\rightarrow$ I, C $\rightarrow$ M, C $\rightarrow$ O, I $\rightarrow$ C, I $\rightarrow$ M, I $\rightarrow$ O, M $\rightarrow$ C, M $\rightarrow$ I, M $\rightarrow$ O, O $\rightarrow$ M, O $\rightarrow$ I, and O $\rightarrow$ C. Two face PAD models, ResNet-18 and PixBis, are trained following these 12 protocols to evaluate the real-world scenarios. The results are shown in Tab. 2. In general, training on SynthASpoof dataset obtains comparable results to the training on authentic data. For example, the average HTER values of ResNet trained on the authentic data and the synthetic data are 25.01% and 26.96%, respectively. When testing on M, C, and O, the models trained on authentic data achieved better performance than the models trained on synthetic data. The possible reason of this observation is that there is a domain gap between the trained synthetic and the tested authentic images. When testing on I, models trained on the SynthASpoof obtain significantly better results than models trained on authentic data, because the SynthASpoof dataset includes a diverse range of replay attacks (more than print ones), enabling the model to generalize well on the Idiap ReplayAttack , which mainly consist of replay attacks. This result indicate the applicability of our SynthASpoof dataset for face PAD.

We also visualize the feature distribution using t-SNE [35] on the most challenging case, CASIA, and the best-performing dataset, Idiap ReplayAttack. As shown in Fig. 4 (a) and (c), we have the following observations: 1) Different attack types are clustered separately (represented by different shades of blue), indicating potentially poor generalization on unseen attacks, as evidenced by the results on the CASIA dataset. 2) There is a clear distance between the SynthASpoof and the authentic datasets (represented by different shapes), implying the domain gap between synthetic and authentic data.

Both quantitative and qualitative results demonstrate the high viability of using SynthASpoof for the development of face PAD algorithms, especially when containing the same type of PA. The observable distance between the synthetic and authentic data will be reduced later in Section 6.3 when incorporating MixStyle.

6.2. Effect of Cropping and Data Augmentation

We explore the impact of data augmentation and adding a margin to the face crop by conducting experiments on the synthetic (training) and the authentic data (test). The results are shown in Tab. 3. We use the following data augmenta-

Method	Training data	C → I	C → M	C → O	I → C	I → M	I → O	M → C	M → I	M → O	O → M	O → C	O → I	Average
ResNet	Authentic	38.85	18.10	17.94	42.22	18.81	28.42	27.11	16.30	30.49	15.71	23.11	23.10	25.01 ± 8.74
	SynthASpoof	8.90	25.48	34.23	39.22	25.48	34.23	39.22	8.90	34.23	25.48	39.22	8.90	26.96 ± 12.04
PixBis	Authentic	25.05	11.19	20.72	34.22	21.67	36.57	39.11	13.65	32.58	15.00	28.11	21.80	24.97 ± 9.27
	SynthASpoof	7.50	38.33	38.70	38.44	38.33	38.70	38.44	7.50	38.70	38.33	38.44	7.50	30.73 ± 14.01

Table 2. The comparison results of models trained on SynthAspoof and authentic datasets, presented as HTER (%). Models trained on SynthASpoof dataset achieve comparable performances to models trained on authentic datasets in many cases, indicating the usability of SynthASpoof for the development of face PAD.

Margin	Aug	M		C		I		O		Average	
		HTER(%) ↓	AUC(%) ↑	HTER(%) ↓	AUC(%) ↑	HTER(%) ↓	AUC(%) ↑	HTER(%) ↓	AUC(%) ↑	HTER(%) ↓	AUC(%) ↑
0%	w/o	21.43	79.96	42.00	59.33	15.90	91.36	36.33	60.27	28.92	72.73
0%	w/	24.52	82.22	39.22	62.00	8.90	96.96	30.91	74.03	**25.89**	**78.80**
5%	w/	24.29	79.18	47.00	52.33	11.45	94.99	35.77	67.98	29.63	73.62

Table 3. The impact of margin extension of face bounding box (extracted from MTCNN) and the effect of using data augmentation by training models on synthetic data and testing separately on four authentic face PAD datasets (M, C, I, and O). The results show that models trained on face images without bounding box extension outperformed models trained on slightly extended face regions. Moreover, applying data augmentation resulted in a better generalized model.

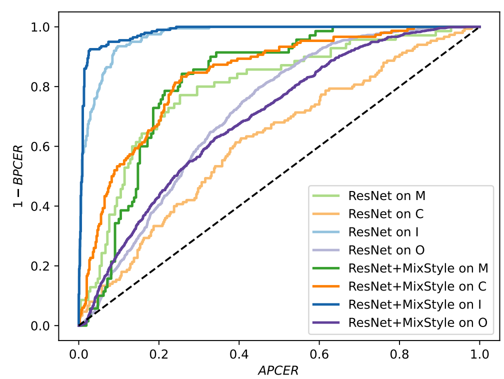
tion [2] techniques: horizontal flipping, scaling and rotation with a limit of 0.1%, random gamma adjustment within a gamma range from 80 to 180, RGB shifting with a limit of 20, and color jittering with a limit of 0.1%. As shown in Tab. 3, using a combined augmentation operation obtains a significant average performance improvement, decreasing the average HTER values from 28.92% to 25.89%.

As previous work have shown that the consideration area beyond the face is beneficial for PAD performance [30], we argue that this enhancement might be related to properties of specific limited dataset and might not generalize well on unknown data, therefore we study the inclusion of such an area. We compare the results between cropping face region without extension and with 5% extension of bounding box extracted from MTCNN [42]. The results in Tab. 3 indicate that cropping faces with extension leads to a lower PAD performance on unknown datasets.

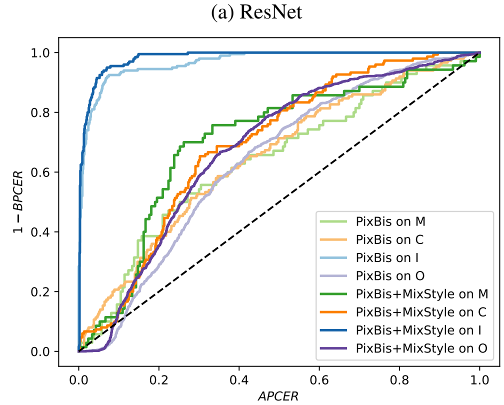
As a result, in the following experiments, all models are trained on the cropped faces without bounding box extension and using data augmentation to enhance the PAD generalizability.

6.3. Effect of MixStyle

As discussed in Sec. 6.1 and showed in Fig. 4, there is a distance between the training synthetic and the testing authentic samples. Motivated by this, MixStyle [46] is adapted to transfer the domain-related information from the authentic data to the synthetic data. In the training process, the authentic face PAD dataset is used without label only for calculating feature statistics, i.e., the loss is only computed based on the synthetic data.



(a) ResNet



(b) PixBis

Figure 3. ROC curves of ResNet (a) and PixBis (b) trained on SynthASpoof dataset and tested on four authentic face PAD datasets (M, C, I, and O). The light colors represent the baseline model and the heavy colors indicate the models trained with the help of MixStyle, mostly leading to better performance (higher curves).

[2]The used augmentation library: Albumentation - https://albumentations.ai/

Method	M		C		I		O		Average	
	HTER(%)↓	AUC(%)↑	HTER(%)↓	AUC(%)↑	HTER(%)↓	AUC(%)↑	HTER(%)↓	AUC(%)↑	HTER(%)↓	AUC(%)↑
ResNet	25.48	79.54	39.22	62.00	8.90	96.96	**34.23**	**71.48**	26.96	77.50
+MixStyle	**21.43**	**81.97**	**22.78**	**83.66**	**6.70**	**98.30**	36.07	69.52	**21.75**	**83.36**
PixBis	38.33	63.87	38.44	64.79	7.50	96.88	35.77	63.50	30.74	72.26
+MixStyle	**30.48**	**70.38**	**32.00**	**69.36**	**6.10**	**98.29**	**34.46**	**67.71**	**25.76**	**76.44**

Table 4. The ablation study of MixStyle for adapting models from the synthetic domain to authentic domain. Note that the target domain (authentic face PAD) is used in the training process for MixStyle without labels. The bold number indicates the best performance for each method, pointing out that the usage of MixStyle resulted in an enhanced PAD performance in general.

Tab. 4 shows that applying MixStyle leads to better model generalizability in most cases. The average HTER values on four testing authentic face PAD datasets decreased from 26.96% to 21.75% by ResNet and from 30.74% to 25.76% by PixBis, while the AUC values increased from 77.50% to 83.36% by ResNet and from 72.26% to 76.44% by PixBis. The ROC curves shown in Fig. 3 illustrate the consistent observation for the baseline model and the model with MixStyle.

To provide a more detailed understanding of the benefit of MixStyle, we visualize the feature space by ResNet models on our most challenging dataset CAISA and the best performed dataset Idiap ReplayAttack in Fig. 4. We have the following observations: 1) Features of different types of synthetic attacks (different blue squares) are clustered more closely by applying MixStyle than baseline models, as well as for bona fide, indicating a better generalizability on unknown attack types. 2) Features of authentic data are clustered more closely within the same class of data by applying MixStyle than baseline models. 3) The distance between features of synthetic (■) and authentic data (✖) is visually reduced by using MixStyle.

The quantitative and visual results discussed above demonstrate that MixStyle helps to enhance the PAD performance of models trained on the synthetic data.

6.4. Effect of a supplementing Authentic Data with SynthASpoof

As models trained on limited data can easily over-fit the training data and thereby generalize poorly to other domains, we investigate the effect of using the SynthASpoof dataset as a supplementary training data to enhance the diversity of authentic training data. *+ SynthASpoof* in Tab. 5 refers to that the SynthASpoof is combined with the authentic data in the training process.

As shown in Tab. 5, including synthetic data in the training process improves the generalizability of the PAD models. For example, the average HTER value of ResNet decreases from 25.01% to 23.34%, and of PixBis decreases from 24.97% to 22.35%. These results suggest that adding the SynthASpoof dataset increases the diversity of training samples and thus leads to better representation learning. Despite the overall improvement, the inclusion of synthetic data did not improve performance in all scenarios. A performance degradation is observed in five out of 12 scenarios when training the ResNet model and in four cases with the PixBis model. This might be caused by the distance between synthetic and authentic face data as we discussed in Sec. 6.1. Therefore, we also utilized MixStyle in a combined training process, aiming to narrow the domain gap between the synthetic and authentic data, just as we did in Sec. 6.3 but with the PAD training here including authentic data.It can be seen that the model trained with MixStyle generalizes better on unseen test data than the one trained without MixStyle, e.g., the average HTER value of ResNet decreases from 23.34% to 19.58% and of PixBis decreases from 22.35% to 19.16% with the help of MixStyle. With MixStyle, suplementing the authentic data with SynthASpoof improved the PAD performance in 10 out 12 experimental setups for the ResNet-based PAD.

In summary, incorporating the SynthASpoof dataset seems to diversify the training data, alleviating the overfitting issue caused by limited training data. Furthermore, MixStyle narrows the domain gap between synthetic and authentic data, leading to improved model generalizability.

6.5. Visualization and Analysis

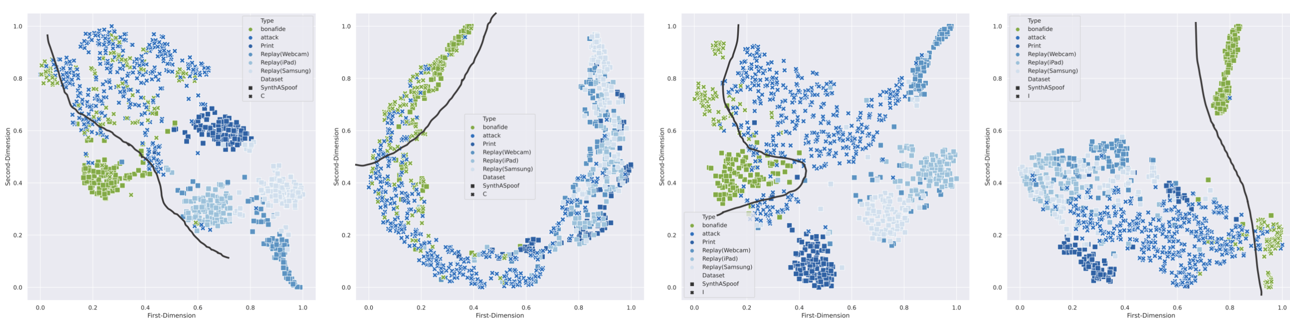
We visualized the feature distribution learned by ResNet without MixStyle and with MixStyle in Fig. 4 by considering the most challenging case, SynthASpoof → CASIA, and the best performing case SynthASpoof → Idiap ReplayAttack (both with results presented in Tab. 4). To avoid the possible overlapping region and obtain a clear observation, we randomly select 500 samples from each dataset and illustrate their distribution by using t-SNE [35]. Comparing Fig. 4 (a) and (b), and (c) and (d), we found that samples obtained by the model with MixStyle are clustered more closely than baseline models, indicating the effectiveness of MixStyle. Furthermore, applying MixStyle results in a clearer decision boundary given the perspective of the discriminative capability.

6.6. Discussion

A extensive experiments successfully demonstrated the feasibility of using the SynthASpoof dataset for the development of face PAD solutions by training as a stand-alone

Method	C → I	C → M	C → O	I → C	I → M	I → O	M → C	M → I	M → O	O → M	O → C	O → I	Average
Binary CNN [40]	45.80	25.60	36.40	44.40	48.60	45.40	50.10	49.90	31.40	30.20	41.20	47.40	41.37 ± 8.42
ADA [36]	17.50	*9.30*	29.10	41.60	30.50	39.60	17.70	5.10	31.20	31.50	19.80	26.80	24.98 ± 11.28
DR-MD-Net [37]	26.10	20.20	24.70	39.20	23.20	33.60	34.30	8.70	31.70	22.00	21.80	27.60	26.09 ± 8.05
DR-UDA [38]	*15.60*	**9.00**	28.70	34.20	29.00	38.50	*16.80*	**3.00**	30.20	27.40	19.50	25.40	23.11 ± 10.50
SDFANet [47]	**15.50**	12.14	*17.08*	46.11	24.29	41.56	**13.33**	11.36	**18.92**	**11.67**	19.33	18.71	20.83 ± 11.44
ResNet	38.85	18.10	17.94	42.22	*18.81*	28.42	27.11	16.30	30.49	15.71	23.11	23.10	25.01 ± 8.74
+ SynthASpoof	36.80	13.10	20.88	33.33	*18.81*	31.65	28.67	11.85	26.53	19.05	20.44	18.95	23.34 ± 7.97
+SynthASpoof + MixStyle	22.50	12.86	19.49	28.56	19.52	*26.96*	17.44	11.10	*20.95*	*14.76*	**18.67**	22.15	*19.58* ± **5.20**
PixBis	25.05	11.19	20.72	34.22	21.67	36.57	39.11	13.65	32.58	15.00	28.11	21.80	24.97 ± 9.27
+ SynthASpoof	20.15	24.05	26.55	*27.67*	23.10	32.41	26.56	8.10	28.18	15.95	*19.00*	*16.50*	22.35 ± *6.72*
+SynthASpoof + MixStyle	22.02	11.19	**16.48**	**23.00**	**14.05**	**23.74**	36.56	*4.05*	28.71	15.71	21.44	**12.95**	**19.16** ± 8.63

Table 5. The results of PAD models trained on a combined training dataset, presented as HTER (%). Combining SynthASpoof data and the authentic PAD images boost the generalizability of PAD models. Moreover, incorporating MixStyle into the training process leads to even a better generalized PAD models. In comparison to existing works, supplementing the authentic data with SynthASpoof and using MixStyle leads to comparable results and an average performance that goes beyond the latest PAD solutions.



(a) ResNet on CASIA (b) ResNet+**MixStyle** on CASIA (c) ResNet on Idiap ReplayAttack (d) ResNet+**MixStyle** on Idiap ReplayAttack

Figure 4. Visualization of the feature distribution by using t-SNE [35] for the training on synthetic (■) and test on authentic face PAD (✖) samples, including the most challenging dataset CASIA in our case and the best performing dataset Idiap ReplayAttack. The bona fide samples in both datasets are illustrated by green, while different attack types in the SynthASpoof dataset and attacks in the authentic dataset are represented in various shades of blue. Fig. (a) and (c) demonstrate a clear distance between different attack in the SynthASpoof dataset and a distance between synthetic and authentic data for both classes, bona fides and attack. Fig. (b) and (c) indicate that MixStyle helps to reduce the distance between the synthetic and the authetic data, i.e., samples within the same class are clustered more closely.

dataset and serving as a supplement to increasing the diversity of limited training data. Although the goal of this work is not to achieve state-of-the-art PAD performances, a comparison to recent major works using the same experimental protocol is presented in Tab. 5. This comparison shows that our PAD trained with the supplement of SynthASpoof and with MixStyle actually outperforms these works in many experimental settings, even leading to a better overall (average) performance. Due to its privacy-friendly characteristic, the SynthASpoof dataset is made publicly available to the research community. Therefore, SynthASpoof can be extended by collecting more attack data to increase the diversity of attacks (printing, screen and capture devices). Moreover, researchers can build PAD datasets to tackle research problems, e.g., the PAD fairness issue [16] by ensuring a demographically-diverse training data.

7. Conclusion

To address the ethical and legal challenges associated with the usage, reuse, and sharing of authentic biometric data and motivated by the need for large-scale and diverse PAD development datasets, this work introduced the first privacy-friendly and synthetic-based dataset, SynthASpoof. The dataset consists of 25,000 bona fide and 78,000 presentation attack samples, which is made publicly available for the research community. We successfully proved the usability of the SynthASpoof dataset for the development of face PADs. We also showed that SynthASpoof enhanced the generalizability of PAD models by enriching the diversity of the limited authentic data. Furthermore, MixStyle helped to decrease the distance between the synthetic and authentic data, resulting in a more robust and generalized presentation attack detector.

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

References

[1] Zinelabdine Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. OULU-NPU: A mobile face presentation attack database with real-world variations. In *FG*, pages 612–618. IEEE Computer Society, 2017. 1, 2, 3, 4

[2] Fadi Boutros, Naser Damer, and Arjan Kuijper. QuantFace: Towards Lightweight Face Recognition by Synthetic Data Low-bit Quantization. In *ICPR*, pages 855–862. IEEE, 2022. 1

[3] Fadi Boutros, Meiling Fang, Marcel Klemt, Biying Fu, and Naser Damer. CR-FIQA: Face Image Quality Assessment by Learning Sample Relative Classifiability. In *CVPR*. IEEE Computer Society, 2023. 3

[4] Fadi Boutros, Marco Huber, Patrick Siebke, Tim Rieber, and Naser Damer. SFace: Privacy-friendly and Accurate Face Recognition using Synthetic Data. In *IJCB*, pages 1–11. IEEE, 2022. 1

[5] Fadi Boutros, Marcel Klemt, Meiling Fang, Arjan Kuijper, and Naser Damer. Unsupervised face recognition using unlabeled synthetic data. In *FG*, pages 1–8. IEEE, 2023. 1

[6] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *FG*, pages 67–74. IEEE Computer Society, 2018. 1

[7] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face antispoofing. In *BIOSIG*, volume P-196 of *LNI*, pages 1–7. GI, 2012. 1, 2, 3, 4

[8] César Augusto Fontanillo López and Abdullah Elbi. On synthetic data: a brief introduction for data protection law dummies, 2022. 1, 2

[9] Naser Damer and Kristiyan Dimitrov. Practical view on face presentation attack detection. In *BMVC*. BMVA Press, 2016. 4

[10] Naser Damer, César Augusto Fontanillo López, Meiling Fang, Noémie Spiller, Minh Vu Pham, and Fadi Boutros. Privacy-friendly Synthetic Data for the Development of Face Morphing Attack Detectors. In *CVPR Workshops*, pages 1605–1616. IEEE, 2022. 1, 2

[11] Meiling Fang, Hamza Ali, Arjan Kuijper, and Naser Damer. Patchswap: Boosting the generalizability of face presentation attack detection by identity-aware patch swapping. In *IJCB*, pages 1–10. IEEE, 2022. 3

[12] Meiling Fang, Fadi Boutros, and Naser Damer. Unsupervised Face Morphing Attack Detection via Self-paced Anomaly Detection. In *IJCB*, pages 1–11. IEEE, 2022. 1

[13] Meiling Fang, Fadi Boutros, Arjan Kuijper, and Naser Damer. Partial Attack Supervision and Regional Weighted Inference for Masked Face Presentation Attack Detection. In *FG*, pages 1–8. IEEE, 2021. 2

[14] Meiling Fang, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Learnable Multi-level Frequency Decomposition and Hierarchical Attention Mechanism for Generalized Face Presentation Attack Detection. In *WACV*, pages 1131–1140. IEEE, 2022. 3, 4, 5

[15] Meiling Fang, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Real masks and spoof faces: On the masked face presentation attack detection. *Pattern Recognit.*, 123:108398, 2022. 2

[16] Meiling Fang, Wufei Yang, Arjan Kuijper, Vitomir Struc, and Naser Damer. Fairness in face presentation attack detection. *CoRR*, abs/2209.09035, 2022. 8

[17] Anjith George and Sébastien Marcel. Deep pixel-wise binary supervision for face presentation attack detection. In *ICB*, pages 1–8. IEEE, 2019. 3, 4, 5

[18] Anjith George, Zohreh Mostaani, David Geissenbuhler, Olegs Nikisins, André Anjos, and Sébastien Marcel. Biometric face presentation attack detection with multi-channel convolutional neural network. *IEEE Trans. Inf. Forensics Secur.*, 15:42–55, 2020. 2

[19] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *ECCV (3)*, volume 9907, pages 87–102. Springer, 2016. 1

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778. IEEE Computer Society, 2016. 3

[21] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *CVPR*, pages 2261–2269. IEEE Computer Society, 2017. 4

[22] Marco Huber, Fadi Boutros, Anh Thi Luu, Kiran B. Raja, Raghavendra Ramachandra, Naser Damer, Pedro C. Neto, Tiago Gonçalves, Ana F. Sequeira, Jaime S. Cardoso, João Tremoço, Miguel Lourenço, Sergio Serra, Eduardo Cermeño, Marija Ivanovska, Borut Batagelj, Andrej Kronovsek, Peter Peer, and Vitomir Struc. SYN-MAD 2022: Competition on face morphing attack detection based on privacy-aware synthetic training data. In *IJCB*, pages 1–10. IEEE, 2022. 1

[23] International Organization for Standardization. ISO/IEC DIS 30107-3:2016: Information Technology – Biometric presentation attack detection – P. 3: Testing and reporting, 2017. 5

[24] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. In *NeurIPS*, 2020. 2, 3

[25] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, pages 4401–4410. IEEE, 2019. 3

[26] Ira Kemelmacher-Shlizerman, Steven M. Seitz, Daniel Miller, and Evan Brossard. The megaface benchmark: 1 million faces for recognition at scale. In *CVPR*, pages 4873–4882. IEEE Computer Society, 2016. 1

[27] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C. Kot. Domain generalization with adversarial feature learning. In *CVPR*, pages 5400–5409. IEEE Computer Society, 2018. 5

[28] Ajian Liu, Chenxu Zhao, Zitong Yu, Jun Wan, Anyang Su, Xing Liu, Zichang Tan, Sergio Escalera, Junliang Xing, Yanyan Liang, Guodong Guo, Zhen Lei, Stan Z. Li, and Du Zhang. Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection. *IEEE Trans. Inf. Forensics Secur.*, 17:2497–2507, 2022. 2

[29] Yaojie Liu, Joel Stehouwer, Amin Jourabloo, and Xiaoming Liu. Deep Tree Learning for Zero-Shot Face Anti-Spoofing. In *CVPR*, pages 4680–4689. IEEE, 2019. 1, 2, 3

[30] Pedro C. Neto, Ana F. Sequeira, and Jaime S. Cardoso. Myope Models - Are face presentation attack detection models short-sighted? In *WACV (Workshops)*, pages 390–399. IEEE, 2022. 6

[31] Haibo Qiu, Baosheng Yu, Dihong Gong, Zhifeng Li, Wei Liu, and Dacheng Tao. Synface: Face recognition with synthetic data. In *ICCV*, pages 10860–10870. IEEE, 2021. 1

[32] Mohammad Rostami, Leonidas Spinoulas, Mohamed E. Hussein, Joe Mathai, and Wael Abd-Almageed. Detection and continual learning of novel face presentation attacks. In *ICCV*, pages 14831–14840. IEEE, 2021. 1, 2, 3

[33] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C. Yuen. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *CVPR*, pages 10023–10031. IEEE, 2019. 5

[34] Rui Shao, Xiangyuan Lan, and Pong C. Yuen. Regularized fine-grained meta face anti-spoofing. In *AAAI*, pages 11974–11981. AAAI Press, 2020. 5

[35] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. 5, 7, 8

[36] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Improving cross-database face presentation attack detection via adversarial domain adaptation. In *ICB*, pages 1–8. IEEE, 2019. 8

[37] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Cross-domain face presentation attack detection via multi-domain disentangled representation learning. In *CVPR*, pages 6677–6686. IEEE, 2020. 8

[38] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection. *IEEE Trans. Inf. Forensics Secur.*, 16:56–69, 2021. 4, 5, 8

[39] Di Wen, Hu Han, and Anil K. Jain. Face spoof detection with image distortion analysis. *IEEE Trans. Inf. Forensics Secur.*, 10(4):746–761, 2015. 2, 3, 4

[40] Jianwei Yang, Zhen Lei, and Stan Z. Li. Learn convolutional neural network for face anti-spoofing. *CoRR*, abs/1408.5601, 2014. 5, 8

[41] Zitong Yu, Xiaobai Li, Jingang Shi, Zhaoqiang Xia, and Guoying Zhao. Revisiting pixel-wise supervision for face anti-spoofing. *IEEE Trans. Biom. Behav. Identity Sci.*, 3(3):285–295, 2021. 3, 5

[42] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Proc. Lett.*, 23(10):1499–1503, 2016. 5, 6

[43] Shifeng Zhang, Xiaobo Wang, Ajian Liu, Chenxu Zhao, Jun Wan, Sergio Escalera, Hailin Shi, Zezheng Wang, and Stan Z. Li. A dataset and benchmark for large-scale multi-modal face anti-spoofing. In *CVPR*, pages 919–928. IEEE, 2019. 2

[44] Yuanhan Zhang, Zhenfei Yin, Yidong Li, Guojun Yin, Junjie Yan, Jing Shao, and Ziwei Liu. Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations. In *ECCV (12)*, volume 12357 of *Lecture Notes in Computer Science*, pages 70–85. Springer, 2020. 1, 2, 3, 5

[45] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z. Li. A face antispoofing database with diverse attacks. In *ICB*, pages 26–31. IEEE, 2012. 1, 2, 3, 4

[46] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. In *ICLR*, 2021. 2, 4, 6

[47] Lifang Zhou, Jun Luo, Xinbo Gao, Weisheng Li, Bangjun Lei, and Jiaxu Leng. Selective Domain-Invariant Feature Alignment Network for Face Anti-Spoofing. *IEEE Trans. Inf. Forensics Secur.*, 16:5352–5365, 2021. 4, 5, 8