Attack-Agnostic Deep Face Anti-Spoofing

Ajian Liu[1], Zichang Tan[2], Yanyan Liang[3], Jun Wan[1,3*]
[1]MAIS, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]Institute of Deep Learning, Baidu Research, Beijing, China
[3]School of Computer Science and Engineering, Faculty of Innovation Engineering,
Macau University of Science and Technology, Macau
ajianliu92@gmail.com, jun.wan@ia.ac.cn

Abstract

The task of face anti-spoofing (FAS) is to determine whether the captured face from a face recognition system is live or fake. Current methods which are trained with existing fake faces ignore the generalization and perform poorly on unseen attacks. To tackle this problem, a novel Attack-agnostic Face Anti-spoofing framework is proposed. Different from previous methods that can be treated as a defense system, we regard face anti-spoofing as a unified framework with the attack and defense systems and optimize the defense system against unseen attacks via adversarial training with the attack system. Concretely, the attack system consists of two modules: an Adversarial learning-based Attack Pattern Generation (Adv-APG) module and a Supervised learning-based Attack Pattern Drift (Sup-APD) module. The Adv-APG module generates a series of spoofing samples by recombining a live face with known attack patterns in a generative way. The Sup-APD module pulls the generated spoofing samples in a supervised way to an unknown domain that makes the defense system ineffective. The defense system is free to choose and compatible with our attack system. Extensive experiments are conducted by using three different defense architectures to verify that the proposed attack system can improve the performance on both seen- and unseen attacks on multiple datasets.

1. Introduction

Face Presentation Attack Detection (PAD) is a technology for defending the face recognition system from malicious attacks, such as print attack, replay attack, or mask. It has become an increasingly critical concern [4, 27, 34, 56] recently due to its wide applications in financial payment, phone unlocking, and face surveillance. However, the generalization of unseen attacks is still a challenging problem, which has not been perfectly solved by these algorithms.

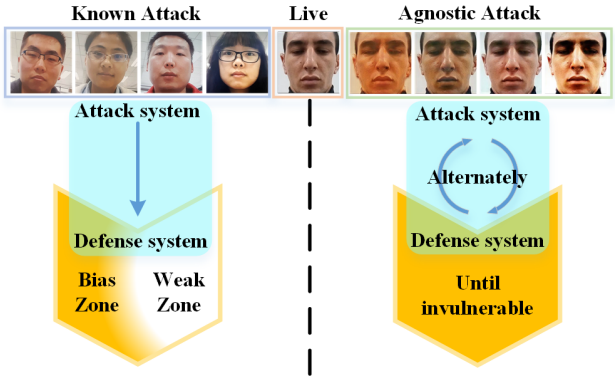

Figure 1. Comparison of two anti-spoofing systems. **Left**: the traditional anti-spoofing system that only considers the defense unit. **Right**: the proposed framework unifies the attack and defense units. Note that these samples are from the OULU-NPU dataset [3].

Some early temporal-based face PAD works attempt to detect the evidence of liveness (*e.g.*, eye-blinking), which requires a constrained human interaction. However, these methods become vulnerable if someone presents a replay attack or a print attack with cut eye/mouth regions. Other works are based on static texture analysis [21]. However, these algorithms are not accurate enough because of the use of handcrafted features, such as LBP [5], HoG [50] and GLCM [39], that do not necessarily capture the most discriminative information associated with the data. Recently, CNN-based face PAD methods [22, 25, 33, 56] and challenges [23, 28, 30] have shown impressive progress due to the excellent performance of deep neural networks and the availability of large datasets [3, 26, 33, 58–60]. With the maturity of 3D printing technology, face mask has become a new type of PA to threaten face recognition systems' security. Compared with traditional 2D Presentation Attacks (PAs), face masks are more realistic in terms of color,

*Contact person

texture, and geometry structure, making it easy to fool a face PAD system designed based on coarse texture [49] and facial depth information [33]. Fortunately, some works have been devoted to 3D mask attacks, including design of datasets [9, 13, 31] and algorithms [11, 31, 32, 44]. Although these methods achieve near-perfect performance in intra-database experiments, they are still vulnerable when facing unseen attacks and cross-database experiments.

Fake faces that are forged from spoof mediums introduce special texture differences compared with a live face, such as color distortions, specular highlights, Moiré patterns, and so on. Based on this observation, both traditional methods (extracting features with handcrafted operators, *e.g.*, LBP [5]) and deep CNN methods (extracting features by utilizing deep networks) focus on the texture differences to distinguish the live faces from spoofing samples. **We regard these existing texture differences in one dataset as known attack patterns (or clues)**. Two limitations severely hinder their performance when facing unseen attacks. First, the lack of live faces (ground truth) that are strictly aligned with fake faces makes face anti-spoofing becomes a very challenging problem. One model might mistake the face subject, pose, expression, or background as a clue to separate the live face from a fake one, but not the faithful attack clue [33]. Second, the attack clues in the training set are known and limited in prior works. While in a real-world face recognition scene may encounter a wide variety of attack clues that are unknown and unpredictable. One model trained on these datasets is very easy to remember the limited attack clues, which leads to poor generalization. Therefore, how to obtain strictly aligned sample pairs (*i.e.*, live-fake faces) and how to introduce unknown attack clues in the training time are effective ways to solve the above deficiencies.

Essentially, face anti-spoofing is a unified system of attack and defense. However, as shown on the left in Fig. 1, the previous works [21, 49, 56] try to make the defense system capable of fighting only the attack clues that have been seen at the training time, but rarely consider how to strengthen the weakest zone of the defense system from the perspective of the attack system. This work is established based on an attack system and introduces unseen fake faces plus the seen fake faces during the training time. As shown on the right in Fig. 1, these introduced fake faces are not only strictly aligned with live faces but also contain unseen attack clues that can attack the weakest zone of the defense system. Therefore, the generalization of our model is improved through the adversarial optimization of these two subsystems to alleviate the bias of the model on the fixed attack clues. To sum up, the contributions are summarized below:

- It is an attempt to address face anti-spoofing by using a unified framework composed of both attack and defense systems. The latter can be alternately optimized against unseen attacks via adversarial training with the proposed attack system.

- Two novel modules, namely Adv-APG and Sup-APD, are proposed in the attack system. The Adv-APG module generates a series of spoofing samples by recombining a live face with known attack clues, while the Sup-APD module pulls the generated spoofing samples to an unknown domain along the direction of disabling the defense system.

- Extensive experiments demonstrate that the proposed attack system has pushed the state-of-the-art performances on several benchmarks, especially for unseen attacks.

2. Related work

2.1. Attacks

Face spoofing (*e.g.*, presentation attacks) is the typical physical attack to deceive the face recognition systems, where attackers present faces from spoof mediums, such as a photograph, screen, or mask, instead of a living human [52]. According to the spoof mediums, we can roughly classify the existing attacks into 2D [3,26,33,58] and 3D attacks [8, 9, 31].

Replay-Attack [5] and CASIA-FASD [61] are two widely used datasets in the face anti-spoofing community. The attack clues in the former are introduced by electronic screens, while the latter is introduced by an additional printing device. Recently, with the widespread application of face recognition in mobile phones, there are also some datasets recorded by replaying face video with a smartphone, such as Replay-Mobile [7], OULU-NPU [3], and SiW [33]. CelebA-Spoof [60] introduces rich attribute annotation information, which can be used as an auxiliary task to improve the generalization of the model in various attacks. With the cost reduction of multi-spectral sensors and the popularity of use scenarios, some new sensors have been introduced to provide more possibilities for face PAD methods. Holger *et al.* [44] uses multi-spectral short wave infrared (SWIR) imaging to ensure the authenticity of a face even in the presence of partial disguises and masks. Zhang *et al.* [58] collects a CASIA-SURF dataset with 3 modalities (*i.e.*, RGB, Depth, and NIR) using Intel RealSense SR300 camera, and proposes a multi-modal multi-scale fusion method for face anti-spoofing. Similarly, work, Liu *et al.* [26] introduces a CASIA-SURF CeFA dataset, covering 3 ethnicities, $1,607$ subjects. As attack techniques are constantly upgraded, some new types of attacks have emerged. Such as a large-scale HiFiMask [31] dataset has been collected. Specifically, it consists of a total amount of

$54,600$ videos which are recorded from 75 subjects with 7 kinds of sensors.

Although the imaging quality of spoofing samples is increasing high (*i.e.*, resolution from 320×240 [5] to $1,920 \times 1,080$ [33]), imaging devices (*i.e.*, containing 6 mobile phones in OULU-NPU [3]) and spoofing mediums (*i.e.*, including 4 display devices in SiW [33]) are more and more diverse, the attack clues are still very limited. Since an anti-spoofing system may encounter a wide variety of spoof types, even unpredictable attacks. It is impractical to accurately model all possible attack clues during the model training, but it is necessary to simulate as much as possible the effective attack clues for the defense system. As compared to the above existing attacks, we design an attack system that continuously produces spoofing samples that effectively attack the weak zone of the defense system.

2.2. Defenses

The essence of face anti-spoofing is a defensive measure for face recognition systems and has been studied for over a decade. Early works were mainly based on color texture [2, 5] and motion analysis [37, 47]. The former is based on the consideration that the fake face is different from the live face in texture details, such as color distortions, and specular highlights, due to the intervention of spoofing mediums. However, these algorithms are not accurate enough because of the use of handcrafted features, such as LBP [5], HoG [39], and SURF [1]. The latter analyzes the attack samples as static or non-rigid motion compared with live faces from the perspective of motion. Unfortunately, these methods become vulnerable if someone presents a replay attack or a print attack with cut eye/mouth regions.

Instead of using pre-defined features such as LBP and HOG, CNN-based methods [49] design a unified framework of feature extraction and classification in an end-to-end manner. However, they treat face anti-spoofing as a binary classification task, and will highly depend on the liveness-unrelated cues, such as color distortion, shape deformation, or background information. Intuitively, the live faces in any scene have consistent face-like geometry. Inspired by this, another work [33, 48, 56] leverages the physical-based depth information instead of binary classification loss as supervision, which are more faithful attack clues in any domain. Liu *et al.* [33] design a CNN-RNN model to leverage the Depth map and rPPG signal as supervision. Similarly, Wang *et al.* [48] take deep spatial gradient and temporal information to assist depth map regression and Yu *et al.* [56] propose a novel frame-level FAS method based on Central Difference Convolution (CDC), which can capture intrinsic detailed patterns via aggregating both intensity and gradient information. Yu *et al.* [53] treat FAS as a material recognition problem and combine it with classical human material perception [41], intending to extract discriminative and robust features for FAS task. Although these CNN-based methods achieve near-perfect performance under known attack clues, they still show poor generalization in the face of unknown attacks.

Another works [20, 29, 34, 42, 57] treat FAS as a feature disentangled representation learning. There are also some methods [40, 46] that focus on improving the generalization of FAS in unknown domains. However, these methods can be treated as a defense system that is trained with existing fake faces and ignore the generalization and perform poorly on unseen attacks. Inspired by the success of Transformers in natural language processing (NLP), convolution-free models that only build on transformer blocks have flourished in computer vision. In FAS community, ViTranZFAS [12] uses the pure ViT to solve the zero-shot anti-spoofing task for the first time. TransRPPG [54] proposes a pure rPPG transformer framework for mining the global relationship within MSTmaps for liveness representation and gives a binary prediction for 3D mask detection. MA-ViT [24] adopts the early fusion to aggregate all the available training modalities' data and enables flexible testing of any given modal samples with a Modality-Agnostic Transformer Block (MATB). ViTAF [17] introduce the ensemble adapters module and feature-wise transformation layers to adapt to different domains with a few samples. In this work, we employ a visual task universal network, *e.g.*, ResNet50 [16], and two FAS task networks, *e.g.*, Auxiliary [33], and CDCN [56] as backbones. At the same time, they combine the proposed attack system to treat anti-spoofing as a unified task of attack and defense to alleviate the bias of the defense system on known attack clues.

3. Proposed Method

The proposed method in this work mainly solves the poor performance of the current anti-spoofing systems in the face of unknown attacks by introducing agnostic attack patterns in the training process. In this part, we first introduce the overall system and then describe each of the subsystems in detail.

3.1. Overall System

As shown in Fig. 2, we regard face anti-spoofing as an integrated attack and defense system. The attack system is composed of an **Adv-APG** module and a **Sup-APD** module. First, Adv-APG translates a live face (abbreviated as $\mathbf{L}$) into a spoofing sample with known attack clues (abbreviated as $\mathbf{S}^k$), and Sup-APD shifts the generated spoofing sample into the agnostic domain (abbreviated as $\mathbf{S}^a$). The defense system can be an arbitrary face anti-spoofing backbone network, such as ResNet50 [16], Auxiliary [33], and CDCN [56]. We take the commonly used Auxiliary [33] method which includes a **Depth Estimator** as an example.

It takes $\mathbf{L}$, $\mathbf{S}^a$ and $\mathbf{S}$ as input, and uses the depth estimator under $\mathcal{L}_2$ distance supervision to regress the corresponding depth maps of input, where $\mathbf{S}$ means the spoofing samples in the original dataset.

Inspired by the generative adversarial learning [14], we organically optimize these two systems by adversarial training manner, *i.e.*, by $\mathbf{max-minimize}$ the loss of depth estimator in defense system, where the attack system and defense system are similar to a generator and a discriminator, respectively. On the one hand, the defense system forces the attack system to generate agnostic attacks with unseen clues, that can attack its weakest zones. And on the other hand, the defense system tries to distinguish the live face from the spoofing samples as much as possible. These two subsystems are trained alternately until they converge to a balanced state.

3.2. Attack System

The attack system plays an important role in our framework. It generates spoofing samples with agnostic attack cues during the training time and can attack the weakest zones of the defense system. However, directly performing random forgery operations on live faces to mimic agnostic attack clues, such as adding random noise [43], blurring image, or applying a perspective transformation (with random parameters) [51], cannot truly simulate the attack clues of the existing fake faces. Because the noise on an existing fake face is a combination of sensor, medium, image content, and environment. Therefore, how to generate as fidelity and unseen spoofing samples as possible is the main purpose of the attack system.

In this work, we sequentially complete this task through two modules: (1) Adv-APG. It translates a live face $\mathbf{L}$ into a spoofing sample $\mathbf{S}^k$ with known attack clues. (2) Sup-APD. It further pushes the translated spoofing sample $\mathbf{S}^k$ to sample $\mathbf{S}^a$ with an agnosic-attack clue.

Adv-APG. One key distinction of our method from the above [43, 51] is that we do not generate our attacks from a fixed pool of common attack clues. Instead, the spoofing examples are generated from a trainable Adv-APG module. See from in Fig. 2, the Adv-APG essentially is a GANs [14] associated with a generator (abbreviated as $\mathcal{G}_{en}$) and a discriminator (abbreviated as $\mathcal{D}_{is}$).

Traditionally, we train $\mathcal{G}_{en}$ to translate a live face into a spoofing sample with a special attack clue i, $\mathcal{G}_{en} : \{\mathbf{L}\} \rightarrow \mathbf{S}^{k_i}$ (i means the index of attack clues). To make the generated spoofing sample indistinguishable from the existing fake faces, we adopt an adversarial loss

$$\begin{aligned}\mathcal{L}^i_{GAN} = &\mathbb{E}_{\mathbf{S}^i}[log\mathcal{D}_{is}(\mathbf{S}^i)] \\ &+ \mathbb{E}_{\mathbf{L}}[log(1-\mathcal{D}_{is}(\mathcal{G}_{en}(\mathbf{L})))]\end{aligned} \tag{1}$$

where $\mathcal{G}_{en}$ tries to minimize this objective against an adversarial $\mathcal{D}_{is}$ that tries to maximize it, *i.e.*, $\mathcal{G}^*_{en} = arg\ min_{\mathcal{G}_{en}}\ max_{\mathcal{D}_{is}}\ \mathcal{L}^i_{GAN}$. However, each of the models is tailored for a specific attack clue translation, such as 4 models are needed to generate samples (*i.e.*, $\mathbf{S}^{i=1}$ for Print1, $\mathbf{S}^{i=2}$ for Print2, $\mathbf{S}^{i=3}$ for Replay1, $\mathbf{S}^{i=4}$ for Replay2) containing all attack clues for a live face from OULU-NPU [3].

In this work, our goal is to train a single generator that maps a live face to a spoofing sample with any specified attack clue. To achieve this, we explore GANs in conditional setting (cGANs) [10] by learning an attack clue-conditional generative model, that translates a live face into a spoofing sample in a generative way under the condition of the specified attack clue. At the same time as an inverse mapping, the existing fake face $\mathbf{S}$ will also be translated to a live face under the condition of the specified live label.

Specially, we first use a one-hot label to classify the attack clues and randomly produce a target label i during training time, where i is the index of one-hot vector, and $i > 0$ means one attack clue, $i = 0$ means the label of corresponding live face for symbolic unity. Such as for OULU-NPU [3], we use $\mathbf{S}^{i=1}$, $\mathbf{S}^{i=2}$, $\mathbf{S}^{i=3}$, and $\mathbf{S}^{i=4}$ to represent Print1, Print2, Replay1, and Replay2, respectively. Then, the Adv-APG module can take in as inputs both image and label information to achieve multi-label image-to-image translation. On the one hand suitable for our task, it learns to translate the live face $\mathbf{L}$ into a spoofing sample by the randomly produced attack label $i(i > 0)$, *i.e.*, $\mathcal{G}_{en} : \{\mathbf{L}, i\} \rightarrow \mathbf{S}^{k_i}$ (briefly named Mapping1). On the other hand for an inverse mapping, it learns to translate the fake face $\mathbf{S}$ into a corresponding live face by the given live label $i(i = 0)$, *i.e.*, $\mathcal{G}_{en} : \{\mathbf{S}, i\} \rightarrow \mathbf{L}$ (briefly named Mapping2). The objective for the Mapping1 (similar to the Mapping2) of a conditional GAN can be expressed as

$$\begin{aligned}\mathcal{L}_{cGAN} = &\mathbb{E}_{\mathbf{L}}[log\mathcal{D}_{is}(\mathbf{L})] \\ &+ \mathbb{E}_{\mathbf{L},i}[log(1-\mathcal{D}_{is}(\mathcal{G}_{en}(\mathbf{L},i)))]\end{aligned} \tag{2}$$

where $\mathcal{G}_{en}$ generates an image $\mathcal{G}_{en}(\mathbf{L}, i)$ conditioned on both the live face $\mathbf{L}$ and the attack label i, while $\mathcal{D}_{is}$ tries to distinguish between real spoof face $\mathbf{S}$ and fake spoof face $\mathbf{S}^k$ (and tries to distinguish between real live face $\mathbf{L}$ and fake live face in Mapping-2). Inspired by the AC-GANs [36], we also introduce an auxiliary classifier that allows a single classifier D_{cls} to control multiple attack clues. It aims to impose the classification loss of attack clue when optimizing the above objective. Similar to StarGAN [6], the classification loss of live faces is used to optimize $\mathcal{D}_{is}$ and defined as:

$$\mathcal{L}^l_{cls} = \mathbb{E}_{\mathbf{L},i=0}[-logD_{cls}(i|\mathbf{L})] \tag{3}$$

where $D_{cls}(i|\mathbf{L})$ represents the probability that the sample $\mathbf{L}$ belongs to a live face. By minimizing this loss, $\mathcal{D}_{is}$ learns to classify the live face $\mathbf{L}$ to its corresponding original label i ($i = 0$). In contrast, the classification loss of generated

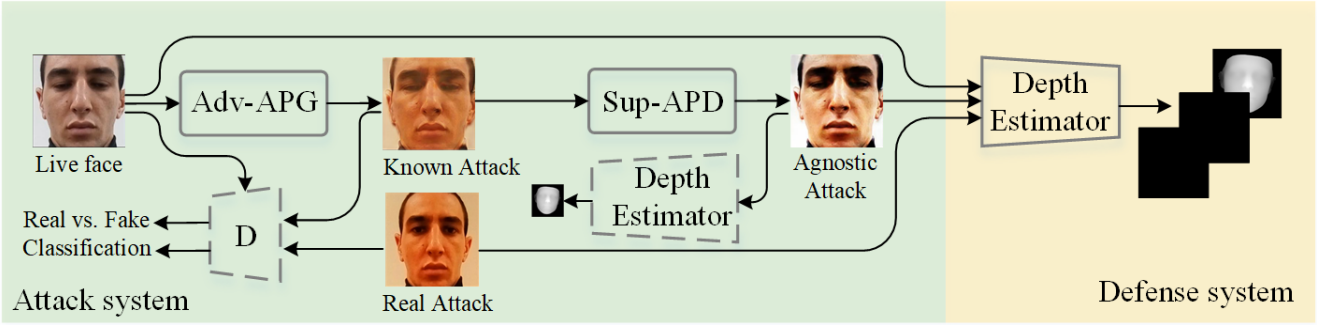

Figure 2. The overall architecture of the agnostic-attack face an anti-spoofing framework. It contains an attack system and a defense system, which are trained in an alternate and adversarial manner. The dotted line indicates that the network parameters are fixed in the current step.

spoofing samples is used to optimize $\mathcal{G}_{en}$:

$$\mathcal{L}^{s}_{cls} = \mathbb{E}_{\mathbf{L},i>0}[-logD_{cls}(i|G(\mathbf{L},i))] \tag{4}$$

where $\mathcal{G}_{en}$ tries to minimize this loss to generate spoofing samples that can be classified as the attack clue i. To further reduce the spatial diversity brought by GANs, we impose a cycle-consistency constraint [6, 62] on the translator $\mathcal{G}_{en}$. In theory, for a live face $\mathbf{L}$, the image translation cycle should be able to bring it back to the original image after two times of inverse mapping, *i.e.*, $\mathbf{L} \rightarrow \mathcal{G}_{en}(\mathbf{L}, i|i>0) \rightarrow \mathcal{G}_{en}(\mathcal{G}_{en}(\mathbf{L}, i|i>0), i|i=0) \approx \mathbf{L}$.

$$\mathcal{L}_{cyc} = \mathbb{E}_{\mathbf{L},i>0} \left\| \mathcal{G}_{en}(\mathcal{G}_{en}(\mathbf{L},i),0) - \mathbf{L} \right\|_1 \tag{5}$$

where the cycle-consistency loss $\mathcal{L}_{cyc}$ can be regarded as a regularizer to guarantee that the learned function $\mathcal{G}_{en}$ can map an individual input $\mathbf{L}$ to a desired output $\mathbf{S}^k$.

Sup-APD. Only generating spoofing samples $\mathbf{S}^k$ that are indistinguishable from the existing fake faces $\mathbf{S}$ cannot effectively attack the weakest zones of the defense system. Because trained on spoofing samples with fixed and limited attack clues can easily lead to overfitting. Therefore, we design a Sup-APD module that will drift the generated sample to an unknown domain along the direction that makes the defense system invalid.

Before exploring this direction, we first introduce the principle of the defense system in this work. See the defense system in Fig. 2, the depth estimator is essentially a depth regression network, denoted as $\mathcal{D}_{ep}$, which outputs the fitted depth map of the input face with the supervision of depth loss. It can be represented by $\mathcal{L}_{dep}$ by calculating the $\mathcal{L}_2$ distance between the depth map of the input face $\mathbf{I}$ and the ground truth $\mathbf{D}$

$$\mathcal{L}_{dep} = \left\| \mathcal{D}_{ep}(\mathbf{I}) - \mathbf{D} \right\|_2^2 \tag{6}$$

where $\mathbf{I}$ is the input face from either a live or a fake face (denoted as $\mathbf{I} \in \{\mathbf{L} \cup \mathbf{S}^a \cup \mathbf{S}\}$), $\mathbf{D}$ is the corresponding ground truth which is estimated by 3DDFA [15] in this work for live faces and 0 for fake faces, and $\mathcal{D}_{ep}(\mathbf{I})$ means the depth map, denoted as $\hat{\mathbf{D}}$, output by $\mathcal{D}_{ep}$ with fixed parameters in inference phase, respectively.

In our context, drifting the spoofing sample to an unknown domain along the direction that makes the defense system invalid is equivalent to searching out the variants of the spoofing sample in an unknown domain that makes the current defense system give the worst prediction. As shown in Fig. 2, it can be done by **maximizing** the $\mathcal{L}_{dep}$ in Eq.6. Furthermore, to ensure the diversity of drifted samples $\mathbf{S}^a$, we design a drifter (abbreviated as $\mathcal{D}_{rf}$) to push the generated spoofing samples $\mathbf{S}^k$ to any domain in as many directions as possible. This process of the Sup-APD module can be expressed as

$$\mathcal{L}_{drf} = \alpha_1 \left\| \mathcal{D}_{rf}\left(\mathbf{S}^k\right) - \mathbf{S}^k \right\|_2^2 - \alpha_2 \mathcal{L}_{dep} \tag{7}$$

where α_1 controls the the fidelity of the image content, while α_2 controls drift strength of the image style, respectively. They are set to 1, 0.1 in our experiments.

Objective for Attack System. Based on the above discussion, we can regard the Sup-APD as a constraint imposed on Adv-APG, which guides the generator $\mathcal{G}_{en}$ to generate $\mathbf{S}^a$ by minimizing the $\mathcal{L}_{drf}$ in Eq.7. Therefore, the objective functions to optimize the attack system are respectively expressed as

$$\mathcal{L}_{\mathcal{G}_{en}} = \mathcal{L}_{cGAN} + \lambda_{cls}\mathcal{L}^{s}_{cls} + \lambda_{cyc}\mathcal{L}_{cyc} + \lambda_{drf}\mathcal{L}_{drf} \tag{8}$$

$$\mathcal{L}_{\mathcal{D}_{is}} = -\mathcal{L}_{cGAN} + \lambda_{cls}\mathcal{L}^{l}_{cls} \tag{9}$$

where $\mathcal{L}_{\mathcal{G}_{en}}$ (and similar to $\mathcal{L}_{\mathcal{D}_{is}}$) is a weighted sum of the adversarial loss (Eq.2), the classification loss (Eq.3 and Eq.4), the cycle-consistency loss (Eq.5), and the pixel-based drift loss (Eq.7). λ_{cls}, λ_{cyc}, λ_{drf} are weights that control the importance of constraint terms, which are set to 1, 10, 1

in all of our experiments, respectively. In particular, the cycle consistency and classification losses focus on restoring image content and style, while the adversarial loss restores texture details and the drift loss enriches the diversity of generated spoof samples.

3.3. Defense System

In contrast to the attack system that maximizes the $\mathcal{L}_{dep}$ (if Auxiliary [33] is used as a defense system), the optimization process of the $\mathcal{D}_{ep}$ along the direction that minimizing the $\mathcal{L}_{dep}$ in the defense system. Especially, the defense system actively strengthens the weakest point of the current $\mathcal{D}_{ep}$ by adversarial training manner with the attack system. It makes the defense system a generalizable approach and provides significant robustness against unseen attacks.

4. Experiments

In this section, we conduct a series of experiments to visually and quantitatively demonstrate the effectiveness of the proposed approach. In the following, we sequentially introduce the experimental setup, implementation details, experimental results, and analysis in detail.

4.1. Experimental Setup

Datasets & Protocols. For the intra-testing experiments, two high-resolution and -quality faces anti-spoofing datasets, including OULU-NPU [3] and SiW [33] are evaluated in our experiments. According to the protocols, we report the results under known attack on Protocols 1, 3 of the OULU-NPU and Protocol 1 of the SiW, whilst under unknown attack on the remaining protocols. For the inter-testing experiments, we utilize the Replay-Attack [5] and CASIA-FASD [61] datasets to perform cross-testing between them, which is widely used as a cross-testing benchmark.

Evaluation Metrics. Especially, Attack Presentation Classification Error Rate (APCER), Bonafide Presentation Classification Error Rate (BPCER), and ACER [18] are used for the metrics. Further, Half Total Error Rate (HTER) is adopted in the cross-testing between Replay-Attack [5] and CASIA-FASD [61].

4.2. Implementation Details

Training Details. The proposed framework is implemented with Pytorch and runs on a single NVIDIA TITAN X GPU. We resize the cropped face region to 256×256. In the training stage, all models are trained with a BatchSize of 6 and an initial learning rate of 0.0001. We train models with 40 epochs from scratch via Adam solver, keep the same learning rate for the first 20 epochs and linearly decay it to 0 over the next 20 epochs.

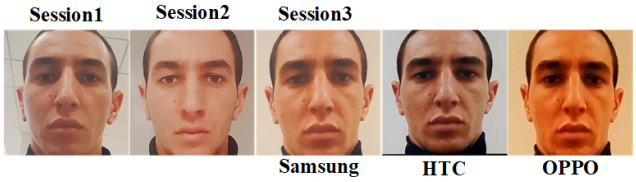

Figure 3. The domain-style display of a print attack in three sessions (Protocol 1), three acquisition devices (Protocol 3).

Network Architecture. The attack system consists of a generator $\mathcal{G}_{en}$ and a discriminator $\mathcal{D}_{is}$ with the same backbone with CycleGAN [62]. For the generator $\mathcal{G}_{en}$, it contains two stride-2 convolutions to downsample the input face, followed by 9 residual blocks, and two fractionally-strided convolutions with stride $1/2$ to upsample to ensure the input and output have the same size. Whilst for the discriminator $\mathcal{D}_{is}$, it is a 70×70 PatchGAN [19], which aims to determine whether the 70×70 image patch is real or fake. Similar to, StarGAN [6], we use instance normalization [45] for the generator but no normalization for the discriminator. Inspired by [35] the generation of watermark attack, our drifter $\mathcal{D}_{rf}$ uses a simple two-layer CNN with the structure of Conv3-LeakyReLU-Conv16. The main contribution of this work is to introduce a plug-and-play attack system to alleviate the bias of the defense to irrelevant factors. Therefore, we chose a relatively simple and effective backbone as much as possible.

For the defense system, we employ four common networks as backbone: ResNet50 [16], Auxiliary [33], and CDCN [56], respectively.

4.3. Results on Known Attack.

Results on OULU-NPU. The results of OULU-NPU are shown in Tab. 1. When the defense system uses Auxiliary [33], the proposed method reduces its ACER to 1.6% and 2.2% on Protocol 1 and 3, respectively. While when the defense system uses CDCN [33], the proposed method achieves the second-best and best performance for ACER, *i.e.*, 0.4%, and 1.7% in Protocol 1 and Protocol 3, respectively. These two protocols introduce various image domains by setting up multiple acquisition sessions and devices. As shown in Fig. 3, a print attack (the same case for other attacks) presents different domain styles in different illuminations, which are more distinct in different camera devices. The diversity of domain styles which are caused by the noise prototypes of sensors [43] poses a certain challenge to the learning of subtle attack clues.

Compared with two prior methods on Protocol 3, *i.e.*, STASN [51] and STDN [52], our approach achieves better performance with an ACER of 1.7%. In particular, by adding our attack system to the Auxiliary [33], all the performances on Protocol 3 are improved, *i.e.*, the metrics that

Table 1. Evaluation results on four protocols of OULU-NPU.

P.	Method	APCER(%)	BPCER(%)	ACER(%)
1	STASN [51]	1.2	2.5	1.9
	STDN [52]	0.8	1.3	1.1
	NAS-FAS [55]	0.4	0.0	**0.2**
	Aux. [33]	1.6	1.6	1.6
	Ours(Aux.)	1.6	1.6	1.6
	CDCN [56]	0.4	1.7	1.0
	Ours(CDCN)	0.8	0.0	0.4
2	STASN [51]	4.2	0.3	2.2
	STDN [52]	2.3	1.6	1.9
	NAS-FAS [55]	1.5	0.8	1.2
	Aux. [33]	2.7	2.7	2.7
	Ours(Aux.)	1.1	1.1	**1.1**
	CDCN [56]	1.5	1.4	1.5
	Ours(CDCN)	1.5	0.9	1.2
3	STASN [51]	4.7±3.9	0.9±1.2	2.8±1.6
	STDN [52]	1.6±1.6	4.0±5.4	2.8±3.3
	NAS-FAS [55]	2.1±1.3	1.4±1.1	1.7±0.6
	Aux. [33]	2.7±1.3	3.1±1.7	2.9±1.5
	Ours(Aux.)	1.4±1.0	3.0±6.6	2.2±3.2
	CDCN [56]	2.4±1.3	2.2±2.0	2.3±1.4
	Ours(CDCN)	1.2±1.2	2.2±1.8	**1.7±1.6**
4	STASN [51]	6.7±10.6	8.3±8.4	7.5±4.7
	STDN [52]	2.3±3.6	5.2±5.4	3.8±4.2
	NAS-FAS [55]	4.2±5.3	1.7±2.6	**2.9±2.8**
	Aux. [33]	9.3±5.6	10.4±6.0	9.5±6.0
	Ours(Aux.)	2.1±1.0	4.9±1.2	3.0±0.5
	CDCN [56]	4.6±4.6	9.2±8.0	6.9±2.9
	Ours(CDCN)	4.9±3.2	5.7±6.2	5.3±2.9

are reduced by 1.3%, 0.1%, and 0.7% for APCER, BPCER, and ACER respectively. Similar configuration for CDCN, the metrics that are reduced by 1.2%, 0.0%, and 0.6% for APCER, BPCER, and ACER respectively. The source of the advantage is the Sup-APD module. It pushes the same spoofing sample into a variety of unknown domains.

Results on SiW. Tab. 2 lists the results of different methods on SiW dataset. Whether the defense system is Auxiliary [33] or CDCN [56], our approach obtains the best performance in Protocol 1 with 0% in all three metrics including APCER, BPCER, and ACER.

Specifically, the proposed method 'Ours(Aux.)' outperforms Auxiliary [33] with a significant margin, *e.g.*, 3.58% of all three metrics on Protocol 1. The proposed method 'Ours(CDCN)' also has non-negligible advantages for defense system CDCN [56], *e.g.*, 0.12% for ACER on Protocol 1. Those improvements mainly benefit from the Adv-APG module that can generate a lot of spoofing samples of being aligned with live faces, which prompts the depth estimator $\mathcal{D}_{ep}$ to focus on attack clues in the face region rather than changes in facial posture and expression.

Table 2. Evaluation results on three protocols of SiW dataset.

P.	Method	APCER(%)	BPCER(%)	ACER(%)
1	STASN [51]	-	-	1.00
	MetaFAS-DR [38]	0.52	0.50	0.51
	STDN [52]	0.00	0.00	**0.00**
	NAS-FAS [55]	0.07	0.17	0.12
	Aux. [33]	3.58	3.58	3.58
	Ours(Aux.)	0.00	0.00	**0.00**
	CDCN [56]	0.07	0.17	0.12
	Ours(CDCN)	0.00	0.00	**0.00**
2	MetaFAS-DR [38]	0.25±0.32	0.33±0.27	0.29±0.28
	STASN [51]	-	-	0.28±0.05
	STDN [52]	0.00±0.00	0.00±0.00	**0.00±0.00**
	NAS-FAS [55]	0.00±0.00	0.09±0.10	0.04±0.05
	Aux. [33]	0.57±0.69	0.57±0.69	0.57±0.69
	Ours(Aux.)	0.09±0.17	0.21±0.25	0.15±0.11
	CDCN [56]	0.00±0.00	0.13±0.09	0.06±0.04
	Ours(CDCN)	0.00±0.00	0.09±0.04	0.05±0.06
3	STASN [51]	-	-	12.10±1.50
	STDN [52]	8.30±3.30	7.50±3.30	7.90±3.30
	MetaFAS-DR [38]	7.98±4.98	7.35±5.67	7.66±5.32
	NAS-FAS [55]	1.58±0.23	1.46±0.08	**1.52±0.13**
	Aux. [33]	8.31±3.81	8.31±3.81	8.31±3.81
	Ours(Aux.)	3.74±2.15	7.10±1.56	5.42±2.13
	CDCN [56]	1.67±0.11	1.76±0.12	1.71±0.11
	Ours(CDCN)	1.39±0.16	1.65±0.14	**1.52±0.17**

4.4. Results on Unknown Attack.

Results on OULU-NPU. The results of Protocol 2 and 4 on OULU-NPU are reported in Tab. 1. For the Auxiliary defense system, significant improvements are achieved in both two protocols when equipped with our attack system, *i.e.*, about 1.6% and 6.5% improvements for ACER on Protocol 2 and Protocol 4. While for the defense system CDCN, the ACER of 0.3% and 1.6% are reduced on Protocol 2 and Protocol 4, respectively.

It shows that our attack system, mainly the Sup-APD module, can not only effectively simulate the domain styles of multiple acquisition devices, but also simulate the sensor noises introduced by various spoofing mediums for the same attack. In addition, compared to STDN [52], we can observe that our approach 'Ours(Aux.)' reduces the error rate in all metrics for the most challenging Protocol 4. It further demonstrates that our method improves the detection ability for unseen attacks by synthesizing various effective spoofing samples is better than that of disentangling spoof traces from the spoofing samples [52]. This is due to the lack of ground truth for spoof traces, and results in the subtle spoofing clues are extremely difficult to be separated from the spoofing samples.

Results on SiW. It is worth noting that on Protocol 3 of SiW (see Tab. 2), our approach 'Ours(Aux.)' outperforms previous methods, *e.g.*, Auxiliary [33] and MetaFAS-DR [38], with achieving 3.74%, 7.10% and 5.42% on APCER, BPCER, and ACER, respectively. When the defense system is replaced by CDCN [56], our approach

Table 3. Quantitative ablation study of each component.

Method	SiW(Protocol 1)			OULU-NPU(Protocol 2)		
	APCER	BPCER	ACER	APCER	BPCER	ACER
Aux. [33]	3.58	3.58	3.58	2.7	2.7	2.7
Aux.+APG	0.48	1.01	0.75	4.4	**0.6**	2.5
Aux.+APD	2.96	3.85	3.41	1.1	1.7	1.4
Ours	**0.00**	**0.00**	**0.00**	**1.1**	1.1	**1.1**

'Ours(CDCN)' achieves the best performance in most metrics, such as 1.39% and 1.52% on APCER and ACER, respectively.

4.5. Ablation Study

To evaluate the contribution of each component in our framework, we perform an ablation study and introduce two variations according to the improvements, *i.e.*, the Auxiliary [33] with Adv-APG (denoted as Aux.+APG), and with Sup-APD (denoted as Aux.+APD).

Effect of the Adv-APG. We verify the effectiveness of our improvements on both the testing protocols of known and unknown attack clues. For the former, we conduct experiments on Protocol 1 of SiW with four methods, *i.e.*, Aux., Aux.+APG, Aux.+APD, and our approach (means Aux.+APG+APD), respectively. From Tab. 3, we can obverse that the most significant contribution comes from the term of APG since the ACER drops sharply compared with APD. In addition, the performance is further improved by combining with APG and APD, *e.g.*, 0% on APCER, BPCER, and ACER of Protocol 1.

Effect of the Sup-APD. In addition, we set up a series of comparative experiments on Protocol 2 of OULU-NPU to measure the performance of our works in unknown attack clues. Tab. 3 shows the comparison results of four methods. The baseline method Aux. achieves decent performance on three metrics, such as all 2.7% on APCER, BPCER, and ACER. Adding the APG and APD to the baseline can reduce the ACER from 2.7% to 2.5% and 1.4%, respectively. It indicates that the Sup-APD plays a more important role in mitigating the impact of different attack mediums, such as unseen printers or displays. Furthermore, in comparison to baseline Aux. which is found to be vulnerable to unseen attacks, our approach by combining APG and APD reduces the ACER from 2.7% to 1.1%. It further demonstrates that we unify the attack and defense system in the way of adversarial training for face anti-spoofing demonstrates better robustness.

4.6. Visualization Analysis

To visually demonstrate the generation effect of our attack system, we randomly select a testing sample from the OULU-NPU dataset, which contains a live face and 4 fake faces with different attack types (*i.e.*, Print1, Print2, Replay1, and Replay2), and then generate spoofing samples corresponding to each face that are aligned with the live face using the trained generator $\mathcal{G}_{en}$.

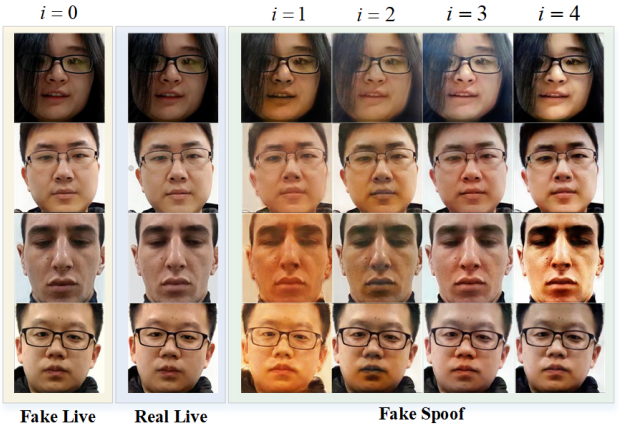

Figure 4. Display of generated samples in two opposite mappings. Note that the $i = 0$, $i = 1$, $i = 2$, $i = 3$, and $i = 4$ are the labels of Live, Print1, Print2, Replay1, and Replay2, respectively.

We show some generated samples in Fig. 4. For Mapping1, our model learns to translate the real live face (the second column) into the fake spoofing sample of Print1, Print2, Replay1, and Replay2 by specifying the attack label as 1, 2, 3, 4, respectively. For Mapping2, our model transforms four types of real spoofing samples into the fake live sample (the first column) by specifying the label as 0. Whether the fake live samples or fake spoofing samples, it is difficult to distinguish from the corresponding real samples.

5. Conclusion

This work proposes an attack-agnostic framework for tackling face anti-spoofing by unifying the attack and defense system. Two modules are introduced in the attack system: the Adv-APG module generates a series of spoofing samples, and the Sup-APD module pulls the spoofing samples to an unknown domain. Finally, extensive experiments demonstrate the performance of the proposed approach.

Acknowledgments

This work was supported by the National Key Research and Development Plan under Grant 2021YFF0602103, the External cooperation key project of Chinese Academy Sciences 173211KYSB20200002, the Chinese National Natural Science Foundation Projects 61876179 and 61961160704, the Science and Technology Development Fund of Macau (No. 0010/2019/AFJ, 0008/2019/A1, 0025/2019/AKP, 0019/2018/ASC, 0004/2020/A1, 0070/2020/AMJ).

References

[1] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face spoofing detection using colour texture analysis. *TIFS*, 2016. 3

[2] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face antispoofing using speeded-up robust features and fisher vector encoding. *SPL*, 2017. 3

[3] Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. Oulu-npu: A mobile face presentation attack database with real-world variations. In *FG*, 2017. 1, 2, 3, 4, 6

[4] Zhihong Chen, Taiping Yao, Kekai Sheng, Shouhong Ding, Ying Tai, Jilin Li, Feiyue Huang, and Xinyu Jin. Generalizable representation learning for mixture domain face anti-spoofing. *arXiv preprint arXiv:2105.02453*, 2021. 1

[5] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *BIOSIG*, 2012. 1, 2, 3, 6

[6] Yunjey Choi, Minje Choi, Munyoung Kim, Jung Woo Ha, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *CVPR*. 4, 5, 6

[7] Artur Costa-Pazo, Sushil Bhattacharjee, Esteban Vazquez-Fernandez, and Sebastien Marcel. The replay-mobile face presentation-attack database. In *BIOSIG*, 2016. 2

[8] Nesli Erdogmus and Sebastien Marcel. Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect. In *BTAS*, 2014. 2

[9] Hao Fang, Ajian Liu, Jun Wan, Sergio Escalera, Chenxu Zhao, Xu Zhang, Stan Z Li, and Zhen Lei. Surveillance face anti-spoofing. *arXiv preprint arXiv:2301.00975*, 2023. 2

[10] Jon Gauthier. Conditional generative adversarial nets for convolutional face generation. *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester*, 2014(5):2, 2014. 4

[11] Anjith George and Sébastien Marcel. Cross modal focal loss for rgbd face anti-spoofing. In *CVPR*, pages 7882–7891, 2021. 2

[12] Anjith George and Sebastien Marcel. On the effectiveness of vision transformers for zero-shot face anti-spoofing, 2021. 3

[13] Anjith George, Zohreh Mostaani, David Geissenbuhler, Olegs Nikisins, André Anjos, and Sébastien Marcel. Biometric face presentation attack detection with multi-channel convolutional neural network. *TIFS*, 15:42–55, 2019. 2

[14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, X. Bing, and Y. Bengio. Generative adversarial nets. *MIT Press*, 2014. 4

[15] Jianzhu Guo, Xiangyu Zhu, Yang Yang, Fan Yang, Zhen Lei, and Stan Z Li. Towards fast, accurate and stable 3d dense face alignment. In *ECCV*, 2020. 5

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, June 2016. 3, 6

[17] Hsin-Ping Huang, Deqing Sun, Yaojie Liu, Wen-Sheng Chu, Taihong Xiao, Jinwei Yuan, Hartwig Adam, and Ming-Hsuan Yang. Adaptive transformers for robust few-shot cross-domain face anti-spoofing. *arXiv preprint arXiv:2203.12175*, 2022. 3

[18] international organization for standardization. Iso/iec jtc 1/sc 37 biometrics: Information technology biometric presentation attack detection part 1: Framework. In *https://www.iso.org/obp/ui/iso*, 2016. 6

[19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017. 6

[20] Amin Jourabloo, Yaojie Liu, and Xiaoming Liu. Face de-spoofing: Anti-spoofing via noise modeling. *arXiv*, 2018. 3

[21] Jukka Komulainen, Abdenour Hadid, and Matti Pietikainen. Context based face anti-spoofing. In *BTAS*, 2013. 1, 2

[22] Xuan Li, Jun Wan, Yi Jin, Ajian Liu, Guodong Guo, and Stan Z Li. 3dpc-net: 3d point cloud network for face anti-spoofing. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2020. 1

[23] Ajian Liu, Xuan Li, Jun Wan, Yanyan Liang, Sergio Escalera, Hugo Jair Escalante, Meysam Madadi, Yi Jin, Zhuoyuan Wu, Xiaogang Yu, et al. Cross-ethnicity face anti-spoofing recognition challenge: A review. *IET Biometrics*, 10(1):24–43, 2021. 1

[24] Ajian Liu and Yanyan Liang. Ma-vit: Modality-agnostic vision transformers for face anti-spoofing. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 1180–1186, 2022. 3

[25] Ajian Liu, Zichang Tan, Xuan Li, Jun Wan, Sergio Escalera, Guodong Guo, and Stan Z Li. Static and dynamic fusion for multi-modal cross-ethnicity face anti-spoofing. *arXiv preprint arXiv:1912.02340*, 2019. 1

[26] Ajian Liu, Zichang Tan, Jun Wan, Sergio Escalera, Guodong Guo, and Stan Z Li. Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1179–1187, 2021. 1, 2

[27] Ajian Liu, Zichang Tan, Jun Wan, Yanyan Liang, Zhen Lei, Guodong Guo, and Stan Z Li. Face anti-spoofing via adversarial cross-modality translation. *IEEE Transactions on Information Forensics and Security*, 16:2759–2772, 2021. 1

[28] Ajian Liu, Jun Wan, Sergio Escalera, Hugo Jair Escalante, Zichang Tan, Qi Yuan, Kai Wang, Chi Lin, Guodong Guo, Isabelle Guyon, et al. Multi-modal face anti-spoofing attack detection challenge at cvpr2019. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–10, 2019. 1

[29] Ajian Liu, Jun Wan, Ning Jiang, Hongbin Wang, and Yanyan Liang. Disentangling facial pose and appearance information for face anti-spoofing. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 4537–4543. IEEE, 2022. 3

[30] Ajian Liu, Chenxu Zhao, Zitong Yu, Anyang Su, Xing Liu, Zijian Kong, Jun Wan, Sergio Escalera, Hugo Jair Escalante, Zhen Lei, et al. 3d high-fidelity mask face presentation attack detection challenge. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 814–823, 2021. 1

[31] Ajian Liu, Chenxu Zhao, Zitong Yu, Jun Wan, Anyang Su, Xing Liu, Zichang Tan, Sergio Escalera, Junliang Xing,

Yanyan Liang, et al. Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 17:2497–2507, 2022. 2

[32] Si-Qi Liu, Xiangyuan Lan, and Pong C Yuen. Remote photoplethysmography correspondence feature for 3d mask face presentation attack detection. In *ECCV*, 2018. 2

[33] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *CVPR*, 2018. 1, 2, 3, 6, 7, 8

[34] Yaojie Liu, Joel Stehouwer, and Xiaoming Liu. On disentangling spoof trace for generic face anti-spoofing. In *ECCV*, pages 406–422, 2020. 1, 3

[35] Xiyang Luo, Ruohan Zhan, Huiwen Chang, Feng Yang, and Peyman Milanfar. Distortion agnostic deep watermarking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13548–13557, 2020. 6

[36] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. 2017. 4

[37] Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *ICCV*, 2007. 3

[38] Yunxiao Qin, Chenxu Zhao, Xiangyu Zhu, Zezheng Wang, Zitong Yu, Tianyu Fu, Feng Zhou, Jingping Shi, and Zhen Lei. Learning meta model for zero- and few-shot face anti-spoofing, 2019. 7

[39] William Robson Schwartz, Anderson Rocha, and Helio Pedrini. Face spoofing detection through partial least squares and low-level descriptors. In *IJCB*, 2011. 1, 3

[40] 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, 2019. 3

[41] Lavanya Sharan, Ce Liu, Ruth Rosenholtz, and Edward H Adelson. Recognizing materials using perceptually inspired features. *IJCV*, 103(3):348–371, 2013. 3

[42] Joel Stehouwer, Amin Jourabloo, Yaojie Liu, and Xiaoming Liu. Noise modeling, synthesis and classification for generic object anti-spoofing. In *CVPR*, pages 7294–7303, 2020. 3

[43] Joel Stehouwer, Amin Jourabloo, Yaojie Liu, and Xiaoming Liu. Noise modeling, synthesis and classification for generic object anti-spoofing. In *CVPR*, 2020. 4, 6

[44] Holger Steiner, Andreas Kolb, and Norbert Jung. Reliable face anti-spoofing using multispectral swir imaging. In *ICB*. IEEE, 2016. 2

[45] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 6

[46] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Cross-domain face presentation attack detection via multi-domain disentangled representation learning. In *CVPR*, pages 6678–6687, 2020. 3

[47] Liting Wang, Xiaoqing Ding, and Chi Fang. Face live detection method based on physiological motion analysis. *TST*, 2009. 3

[48] Zezheng Wang, Zitong Yu, Chenxu Zhao, Xiangyu Zhu, Yunxiao Qin, Qiusheng Zhou, Feng Zhou, and Zhen Lei. Deep spatial gradient and temporal depth learning for face anti-spoofing. *CVPR*, 2020. 3

[49] Jianwei Yang, Zhen Lei, and Stan Z Li. Learn convolutional neural network for face anti-spoofing. *arXiv*, 2014. 2, 3

[50] Jianwei Yang, Zhen Lei, Shengcai Liao, and Stan Z Li. Face liveness detection with component dependent descriptor. In *ICB*, 2013. 1

[51] Xiao Yang, Wenhan Luo, Linchao Bao, Yuan Gao, Dihong Gong, Shibao Zheng, Zhifeng Li, and Wei Liu. Face anti-spoofing: Model matters, so does data. In *CVPR*, pages 3507–3516, 2019. 4, 6, 7

[52] Xiaoming Liu Yaojie Liu, Joel Stehouwer. On disentangling spoof trace for generic face anti-spoofing. In *CVPR*, 2020. 2, 6, 7

[53] Zitong Yu, Xiaobai Li, Xuesong Niu, Jingang Shi, and Guoying Zhao. Face anti-spoofing with human material perception. In *ECCV*, pages 557–575, 2020. 3

[54] Zitong Yu, Xiaobai Li, Pichao Wang, and Guoying Zhao. Transrppg: Remote photoplethysmography transformer for 3d mask face presentation attack detection. *IEEE Signal Processing Letters*, 2021. 3

[55] Zitong Yu, Jun Wan, Yunxiao Qin, Xiaobai Li, Stan Z. Li, and Guoying Zhao. Nas-fas: Static-dynamic central difference network search for face anti-spoofing. In *TPAMI*, 2020. 7

[56] Zitong Yu, Chenxu Zhao, Zezheng Wang, Yunxiao Qin, Zhuo Su, Xiaobai Li, Feng Zhou, and Guoying Zhao. Searching central difference convolutional networks for face anti-spoofing. *CVPR*, 2020. 1, 2, 3, 6, 7

[57] Ke-Yue Zhang, Taiping Yao, Jian Zhang, Ying Tai, Shouhong Ding, Jilin Li, Feiyue Huang, Haichuan Song, and Lizhuang Ma. Face anti-spoofing via disentangled representation learning. In *ECCV*, pages 641–657, 2020. 3

[58] Shifeng Zhang, Ajian Liu, Jun Wan, Yanyan Liang, Guodong Guo, Sergio Escalera, Hugo Jair Escalante, and Stan Z Li. Casia-surf: A large-scale multi-modal benchmark for face anti-spoofing. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(2):182–193, 2020. 1, 2

[59] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 919–928, 2019. 1

[60] 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*, 2020. 1, 2

[61] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z Li. A face antispoofing database with diverse attacks. In *ICB*, 2012. 2, 6

[62] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *ICCV*, 2017. 5, 6