Test-Time Domain Generalization for Face Anti-Spoofing

Qianyu Zhou[1]*, Ke-Yue Zhang[2]*, Taiping Yao[2], Xuequan Lu[3], Shouhong Ding[2]†, Lizhuang Ma[1]†
[1]Shanghai Jiao Tong University; [2]Youtu Lab, Tencent; [3] La Trobe University.
[1]zhouqianyu@sjtu.edu.cn, [1]ma-lz@cs.sjtu.edu.cn,
[2]{zkyezhang,taipingyao,ericshding}@tencent.com, [3]b.lu@latrobe.edu.au

Abstract

Face Anti-Spoofing (FAS) is pivotal in safeguarding facial recognition systems against presentation attacks. While domain generalization (DG) methods have been developed to enhance FAS performance, they predominantly focus on learning domain-invariant features during training, which may not guarantee generalizability to unseen data that differs largely from the source distributions. Our insight is that testing data can serve as a valuable resource to enhance the generalizability beyond mere evaluation for DG FAS. In this paper, we introduce a novel Test-Time Domain Generalization (TTDG) framework for FAS, which leverages the testing data to boost the model's generalizability. Our method, consisting of Test-Time Style Projection (TTSP) and Diverse Style Shifts Simulation (DSSS), effectively projects the unseen data to the seen domain space. In particular, we first introduce the innovative TTSP to project the styles of the arbitrarily unseen samples of the testing distribution to the known source space of the training distributions. We then design the efficient DSSS to synthesize diverse style shifts via learnable style bases with two specifically designed losses in a hyperspherical feature space. Our method eliminates the need for model updates at the test time and can be seamlessly integrated into not only the CNN but also ViT backbones. Comprehensive experiments on widely used cross-domain FAS benchmarks demonstrate our method's state-of-the-art performance and effectiveness.

1. Introduction

Face anti-spoofing (FAS) is critical in safeguarding face recognition [11, 35, 46, 88] systems against different types of presentation attacks, such as printed photos or replaying videos. To address these presentation attacks, researchers have developed a range of FAS approaches, including those based on handcrafted features [5, 17, 18, 39, 67, 98, 102], as well as methods relying on deep learning for feature extraction [7, 14, 48, 51, 95, 97, 107, 109, 110]. While these techniques have shown promising results within specific datasets, they often struggle to perform well when confronted with unseen domains due to the distribution shifts.

*Equal contribution.
†Corresponding author.

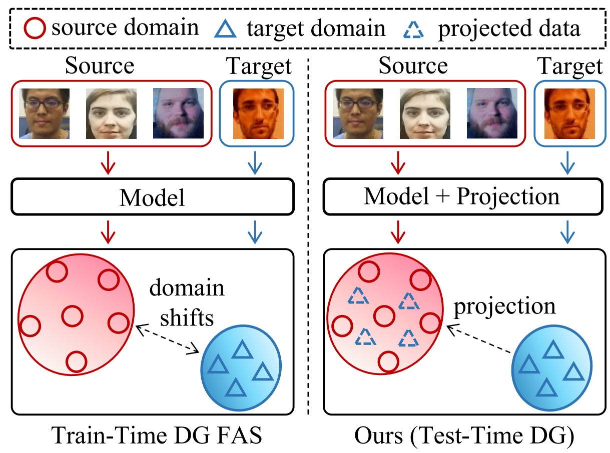


Figure 1. Conventional DG FAS approaches typically learn domain-invariant features at train time but cannot guarantee generalizability to unseen data that largely differ from source domains. In contrast, we propose test-time DG for FAS that projects the unseen testing data to the seen space, thus enhancing the generalizability of FAS model without any model updates at test time.

To improve the performance in unseen environments, recent research has introduced domain generalization (DG) techniques into FAS tasks. Some adversarial learning [30, 76, 96] techniques tend to align the domain distributions via mini-maxing the domain discriminator, and meta-learning [9, 13, 31, 54, 55, 116] methods tend to simulate the unseen domain from the source domains. Other methods, *e.g.,* instance whitening [121] and contrastive learning [96], align various instances in a self-supervised manner. However, all these methods focus on learning domain-invariant features during training to enhance generalization. As a result, they may still encounter performance degradation when dealing with unseen domains that have a significantly large discrepancy with the source domains.

To mitigate this issue, some domain adaptation [37, 38, 89, 117] based FAS approaches aim to directly leverage the target data to align the testing distributions with the training ones. Nevertheless, such methods suffer from two limitations in utilizing the target data. Firstly, they necessitate updating the models for the target domain, which imposes a significant computational burden and in turn, severely impacts the performance in the source domains. Secondly, they require a large amount of testing data for adaptation, which is not always available in realistic scenarios.

A natural question is can we leverage the target data in a more effective manner to enhance the generalizability of the FAS model? To address this question, we propose a novel Test-Time Domain Generalization (TTDG) framework for DG FAS. Unlike traditional DG FAS methods which solely focus on training data, our insight is that testing data can serve as a valuable resource to enhance the generalizability beyond mere prediction before classification. Our TTDG framework elegantly utilizes the testing data to improve the performance of the FAS model without any model updates at test time, as shown in Figure 1. Firstly, an innovative Test-Time Style Projection (TTSP) is introduced to dynamically leverage unseen samples by projecting them to the known source space based on the similarity between the unseen samples and the training distributions. Specifically, to accurately model the training distributions, we are motivated to design a series of style bases to handle the various domain shifts in the training data, *e.g.,* illumination and color, *etc.* However, manually selecting style bases for source domains is cumbersome and time-expensive, and there is no guarantee that selected bases will fully capture the domain shifts in the training data, and accurately project the unseen samples into the correct position during the test time. As such, we design the efficient Diverse Style Shifts Simulation (DSSS) with two new losses to model diverse style shifts via learnable style bases in a hyperspherical feature space. The first loss is a style diversity loss that encourages each learnable style basis to be orthogonal in the hyperspherical space, thus increasing the diversity in the style bases. The second one is a content consistency loss that ensures each projected feature is closely aligned with its corresponding content feature, preventing content distortion. Our TTDG is model-agnostic and can be seamlessly integrated into not only CNN but also ViT backbones.

Our contributions are three-fold:

• We offer a new perspective for DG FAS that leverages the testing data to enhance the generalizability beyond evaluation and propose a novel Test-Time Domain Generalization (TTDG) framework for FAS. To the best of our knowledge, this is the first work that studies test-time DG for FAS.

• We present Test-Time Style Projection (TTSP) to project the unseen samples to the seen source distributions via the aggregation of a set of style bases. Besides, we design Diverse Style Shifts Simulation (DSSS) with two new losses to synthesize diverse distribution shifts via learnable style bases in a hyperspherical feature space.

• We conduct extensive experiments that demonstrate the state-of-the-art performance and effectiveness of our TTDG on widely used cross-domain FAS benchmarks.

2. Related Work

Face Anti-Spoofing. Face anti-spoofing (FAS) aims to determine whether an image captures a genuine human face or a presentation attack, such as a printed photo or video replay. Early FAS research relied on hand-crafted features [4, 18, 39, 67, 72, 102] to detect spoof patterns. With the rise of deep learning, various techniques, *e.g.,* classification-based methods [14, 24, 45, 59, 73, 85, 103], regression-based methods [2, 36, 58, 75, 93, 100, 104–107, 109], and generative models [34, 51, 57, 60, 99] have been explored to enhance FAS performance. Recently, vision Transformer [12, 68, 81] has shown promising potential in FAS [20, 23, 28, 47, 50, 52, 94, 95]. Despite their gratifying progress in the intra-dataset settings, their performances degrade significantly when applied to different target domains. To mitigate this challenge, domain adaptation techniques [16, 19, 21, 22, 53, 78, 79, 82, 101, 115, 118–120] have been recently integrated into FAS [32, 43, 61, 64, 70, 87, 90, 92], but the target data is not always accessible in real scenarios and might fail these methods. Thus, domain generalization techniques [42, 44, 65, 66] have been introduced to improve the performance on unseen domains via adversarial learning [30, 40, 76, 96], meta-learning [9, 13, 31, 54, 55, 116], instance whitening [121] and *etc* [25, 26, 62, 63, 114]. Nevertheless, almost all of them merely focus on learning domain-invariant features at train time and may fail in real-world scenarios that differ significantly from the source domains. Besides, they overlook the role of testing data beyond just evaluation. In this work, we propose a novel perspective that leverages the testing data to boost the generalizability of FAS models.

Test-Time Domain Adaptation and Domain Generalization. Test-time adaptation (TTA) [8, 69, 86, 91, 111] has been studied to enhance the model's transferability to the target domain, where the classifier is updated partially or fully using incoming batches of test samples. Kim *et al.* [37, 38] proposed a style selective normalization for test-time adaptive FAS. Similarly, [117] and [89] aim to align the target domain with the source ones in a reverse manner. Nevertheless, they need batches of the target data for gradient updates [37, 38, 89] or an additional model for fine-tuning [117], which requires considerable computation burden during the test time. In addition, acquiring such a large number of testing data is impractical in realistic scenarios. In contrast to these methods, test-time domain generalization (TTDG) is more challenging since it does not require

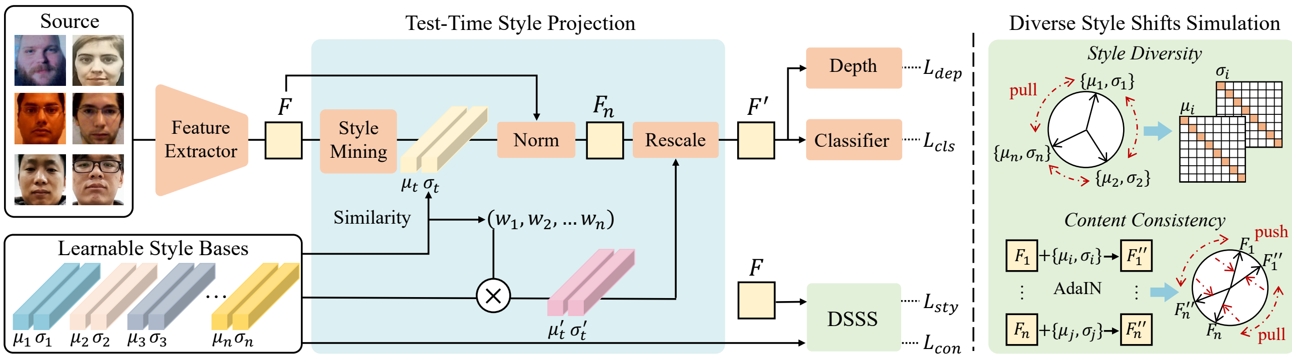


Figure 2. Overview of the proposed Test-Time Domain Generalization (TTDG) framework for DG FAS. In particular, we first introduce Test-Time Style Projection (TTSP) to project arbitrarily unseen samples to the known source space based on the similarity between the unseen sample and the style bases. We then design Diverse Style Shifts Simulation (DSSS) to synthesize diverse style shifts via learnable style bases. $\mathcal{L}_{\text{sty}}$ and $\mathcal{L}_{\text{con}}$ are two new losses for maximizing the style diversity and content consistency in a hyperspherical feature space. Our TTDG eliminates the need for model updates at test time and can be seamlessly integrated into the CNN and ViT backbones.

any model updates at test time. Park *et al.* introduced test-time style shifting [71], which shifts the style of the test sample to the nearest source domain before making the prediction. Besides, [29, 33, 113] pulled the target closer to the source distributions via Fourier transformation and normalization, *etc.* Despite their encouraging advancements, they suffer from two limitations. Firstly, style bases in [29, 71, 113] are roughly defined and inadaptive, either one for each domain or one for all domains, which is inapplicable to FAS tasks that typically have a mixture of domains. In contrast, we do not require any domain label and design a set of learnable style bases that are more fine-grained to automatically capture style shifts. Secondly, their projections lack a clear objective for optimization, making projection less reliable for FAS. Conversely, we propose explicit optimization goals to ensure the reliability of projection.

3. Methodology

Figure 2 shows the overview of the proposed Test-Time Domain Generalization (TTDG) framework, which aims to leverage the testing data to improve the generalizability of FAS models. Our TTDG framework consists of two key components. Firstly, we present Test-Time Style Projection (TTSS) to project the styles of the unseen samples to the style representation space built on style bases, according to the similarity between the unseen style and style bases. In addition, we design Diverse Style Shifts Simulation (DSSS) with two new losses to synthesize diverse distribution shifts via learnable style bases in a hyperspherical feature space.

3.1. Theoretical Analysis

To perform domain generalization, we need to first understand the distribution shifts measured by $\mathcal{H}$-divergence [3]:

$$d_{\mathcal{H}}(\mathcal{D}_s, \mathcal{D}_t) = 2 \sup_{h \in \mathcal{H}} \left|\Pr_{x \sim \mathcal{D}_s}[h(x) = 1] - \Pr_{x \sim \mathcal{D}_t}[h(x) = 1]\right| \tag{1}$$

where classifier $h : \mathcal{X} \to \{0, 1\}$. Then, [1] defines the convex hull Λ_s of $\mathcal{D}_s$ that is a set of mixture of source domains:

$$\Lambda_s = \left\{ \sum_{i=1}^{K} \eta_i \mathcal{D}_s^i \mid \eta \in \Delta_{K-1} \right\}, \tag{2}$$

where η denotes non-negative coefficient in the $(K-1)$-dimensional simplex Δ_{K-1}. Next, an ideal case $\bar{\mathcal{D}}_t \in \Lambda_s$ is assumed that the ideal target domain $\bar{\mathcal{D}}_t$ lies in the source domain convex hull Λ_s. Under this assumption, the risk $\epsilon_t(h)$ on the target domain $\mathcal{D}_t$ is upper-bounded [1] by:

$$\epsilon_t(h) \leq \sum_{i=1}^{K} \eta_i \epsilon_s^i(h) + \gamma + \zeta, \tag{3}$$

where, on the right side, the first term represents the risks over all source domains, and the second term $\gamma = d_{\mathcal{H}\Delta\mathcal{H}}(\bar{\mathcal{D}}_t, D_t)$ denotes the $\mathcal{H}$-divergence between the ideal target $\bar{\mathcal{D}}_t$ and the real target domain D_t, and the third term $\zeta = \sup_{D_s', D_s'' \in \Lambda_s} d_{\mathcal{H}\Delta\mathcal{H}}(D_s', D_s'')$ is the largest $\mathcal{H}$-divergence between any pair of source domains. $\mathcal{H}\Delta\mathcal{H}$ corresponds to $\{h(x) \oplus h'(x) \mid h, h' \in \mathcal{H}\}$. The first term can be minimized by empirical risk minimization (ERM), and the second term is hard to minimize due to no access to the target domain at training, and the third term can be minimized by removing the source domain-specific information which is the style information in the context of this work.

Almost all previous DG FAS approaches [9, 13, 30, 31, 41, 49, 54, 55, 76, 96, 116] assume that the ideal target domain $\bar{\mathcal{D}}_t$ is covered by the source convex hull Λ_s, and thus a model can achieve acceptable performance on the target domain by just minimizing the source divergence (Eq. (3)).

However, this assumption typically does not hold in reality since the realistic target data may differ significantly from the source domains. Recent works [96, 121] rely on data augmentation to generate the data outside the source distributions, possibly extending the source convex hull Λ_s (*i.e.,* $\gamma \to 0$). Nevertheless, the augmented source domain might not fully overlap with the target domain, leading to the failure of generalization of existing models on unseen domains.

The above analysis motivates us to re-think the domain generalization for FAS. Our core idea is to leverage the testing data as a valuable resource to enhance the generalizability. Our TTSP (Section 3.2) and our DSSS (Section 3.3) aim to pull the target data closer to the source convex hull Λ_s, thus reducing the difference between D_t and $\bar{D}_t$ ($\gamma \to 0$).

3.2. Test-Time Style Projection

Our test-time style projection (TTSP) strategy aims to project the styles of the unseen test samples to the known space to handle arbitrary unseen domains during the testing phase. To achieve this goal, there are two key questions that we need to explore for DG FAS. Firstly, how to represent the known style space to the utmost extent? Secondly, how to effectively shift or project the unseen sample to the known domains? We address these questions below.

Regarding the first question, our idea is to build a robust style representation space that can be defined by a series of style bases since various presentation attacks primarily vary in terms of styles, such as illumination, color, *etc.*, and such style differences are the main factor in leading to domain shifts. Therefore, the key to improving the generalizability of the FAS model lies in narrowing the style gaps. Previous FAS studies [37, 38, 96, 117, 121] have demonstrated that the statistics of the latent features of FAS models can reflect the style information of the input image x_t, and most of them commonly employ the channel-wise mean and variance of these features to represent the style distribution of x_t. Following them, $F_t \in \mathbb{R}^{C\times H\times W}$ is denoted as the feature of x_t from the feature extractor, where C denotes the number of channels. The channel-wise mean $\mu_t(F_t) \in \mathbb{R}^C$ and variance $\sigma_t(F_t) \in \mathbb{R}^C$ of the feature F_t can be calculated as follows (the Style Mining part in Figure 2):

$$\mu_t = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} F_t, \sigma_t = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}(F_t-\mu_t)^2}, \quad (4)$$

We design a series of style bases $B_{sty} = \{(\mu_b^n, \sigma_b^n)\}_{n=1}^N$ to preserve the style information of source domains, where N denotes the number of style bases. Then, we build a style representation space based on these style bases B_{sty} for realizing the test-time projection of the unseen style. The way of properly selecting these style bases for the DG FAS task will be discussed in Section 3.3.

As for the second question, we aim to project the style of unseen faces into the style representation space as a weighted combination of style bases. To achieve this goal, firstly, we calculate the cosine distance to estimate the style distribution discrepancy d_n between the current image x_t and the n-th style basis (μ_b^n, σ_b^n), defined as follows:

$$d_n = \frac{\mu_t \cdot \mu_b^n}{\|\mu_t\| \cdot \|\mu_b^n\|} + \frac{\sigma_t \cdot \mu_b^n}{\|\sigma_t\| \cdot \|\mu_b^n\|}, w_n = \frac{e^{d_n}}{\sum_{n=1}^{N} e^{d_n}}, \quad (5)$$

where w_n denotes the estimated weighting factor calculated by the Softmax operation such that the sum of $w = \{w_n \mid n = 1, 2, \ldots, N\}$ is equal to 1.

Next, we can obtain the projected style (μ_t', σ_t') by the weighted combination of style bases as follows:

$$\mu_t' = \sum_{n=1}^{N} w_n \cdot \mu_b^n, \qquad \sigma_t' = \sum_{n=1}^{N} w_n \cdot \sigma_b^n, \quad (6)$$

With the projected styles (μ_t', σ_t') and input feature F_t of the t-th sample, the style projected feature F_t' is defined as:

$$F_t' = \sigma_t'(F_t) \cdot \left(\frac{F_t - \mu_t(F_t)}{\sigma_t(F_t)}\right) + \mu_t'(F_t), \quad (7)$$

As such, each test sample that has a style gap with the source domains will be projected to the source domains via the aggregation of the set of style bases. For example, when the unseen faces have a large discrepancy with the style representation space, the nearest style basis that the model is familiar with will have a large contribution to the projection, and the farthest style basis will have less contribution.

3.3. Diverse Style Shifts Simulation

Although building a style representation space is promising for test-time DG, manually selecting style bases of source domains is cumbersome and time-expensive, especially when the style space is continually changing during the model updating, and it requires to re-select the bases from all the source domains in every epoch. This manner will complicate the procedure and largely reduce the model's efficiency. Besides, there is no guarantee that manually selected style bases will represent the style representation space to the utmost extent. For example, those selected bases might merely include dominant styles that have high frequency and ignore the rare styles with low frequency in the source domains. As a result, these selected style bases may steer the model in the incorrect direction at test time.

To address these issues, instead of utilizing the manually selected style bases, we propose a novel strategy that uses learnable style bases in hyperspherical feature space to synthesize diverse style shifts for FAS. Our method, namely Diverse Style Shifts Simulation (DSSS), is more efficient and effective. In addition, two new loss functions are specifically introduced to guide the learning of learnable style bases. We will describe them in detail below.

Style Diversity Loss. To maximize the diversity of N style bases in a hyperspherical feature space, we present a style diversity loss such that the i-th style basis $B^i_{sty} = \left(\mu^i_b, \sigma^i_b\right)$ is orthogonal to other ones $B^k_{sty} \in \{(\mu^n_b, \sigma^n_b)\}^N_{k=1,k!=i}$. Regarding this, the style diversity loss $\mathcal{L}_{\text{style}}$ for learning the i-th style basis is computed by:

$$\mathcal{L}_{\text{sty}} = \sum_{\substack{k=1\\k\neq i}}^{N} \left| \frac{\mu^i_b}{\|\mu^i_b\|} \cdot \frac{\mu^k_b}{\|\mu^k_b\|} \right| + \sum_{\substack{k=1\\k\neq i}}^{N} \left| \frac{\sigma^i_b}{\|\sigma^i_b\|} \cdot \frac{\sigma^k_b}{\|\sigma^k_b\|} \right|, \tag{8}$$

The objective of the style loss $\mathcal{L}_{\text{style}}$ is to minimize the absolute value of the cosine similarity between the i-th style basis and every other existing style basis. When this loss value reaches zero, it signifies that the i-th style basis has achieved orthogonality with respect to all the other ones.

Content Consistency Loss. Merely using style diversity loss to learn style bases might potentially result in a less desirable outcome because learnable style bases could substantially distort the content information when used to generate a style-content reassembled feature. Thus, for each basis, we encourage the style-content feature to exhibit the highest consistency with its corresponding content feature.

Specifically, for each input feature F_t, we randomly select a style basis B^i_{sty} from the style bases set, and reassemble a style-content feature F''_t using Eq. (7). Then, we devise a content consistency loss $\mathcal{L}_{\text{content}}$ that maximizes the cosine similarity scores between F''_t and F_t as follows:

$$z_{mt} = \frac{F''_t}{\|F''_t\|_2} \cdot \frac{F_m}{\|F_m\|_2}, \tag{9}$$

$$\mathcal{L}_{\text{con}} = -\frac{1}{M} \sum_{t=1}^{M} \log \left(\frac{\exp(z_{tt})}{\sum_{m=1}^{M} \exp(z_{mt})} \right), \tag{10}$$

where M denotes the batch size and z_{mt} is the cosine similarity score between the style-content reassembled feature F''_t and the content feature F_m of the m-th sample. This content loss $\mathcal{L}_{\text{content}}$ encourages each style-content feature to be closer to its corresponding original feature. This way forces each i-th style basis B^i_{sty} to preserve content information when used to synthesize style-content features.

3.4. Training and Inference

Training. To ensure that the feature extractor captures task-relevant features F_t of each sample X_t for good classification, we introduce a binary classification loss $\mathcal{L}_{\text{cls}}$:

$$\mathcal{L}_{\text{cls}} = -\sum_{(X_t, Y^{cls}_t)} Y^{cls}_t \log(\text{Cls}(F_t)), \tag{11}$$

In this equation, Cls represents the binary classifier responsible for distinguishing genuine faces from face presentation attacks. Here, X_t corresponds to the input image, and Y^{cls}_t is the classification label, as illustrated in Figure 2.

Previous research [54, 55, 58] has demonstrated the usefulness of depth information for guiding FAS at the pixel level. We follow them by utilizing a depth estimator (Dep), which estimates depth maps for live faces and zero maps for spoof faces. With the guidance of depth label Y^{dep}_t, we introduce the depth loss $\mathcal{L}_{Dep}$, defined as follows:

$$\mathcal{L}_{\text{dep}} = \sum_{(X_t, Y^{dep}_t)} \left| \text{Dep}(F_t) - Y^{dep}_t \right|^2_2, \tag{12}$$

Additionally, to ensure that our FAS model could well project the unseen sample to the known space during the test time, we simulate this projection process during the training, and the total training loss is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \lambda_d \mathcal{L}_{\text{dep}} + \mathcal{L}_{\text{sty}} + \lambda_c \mathcal{L}_{\text{con}}, \tag{13}$$

Instead of manually re-selecting the style bases in each epoch, we jointly train the whole model with learnable style bases in every iteration, which is a more efficient manner.

Inference. During the test phase, unseen faces are fed into the feature extractor and then projected into the style representation space via our TTSP. The outputs are next fed into the classifier for making the final prediction. Note that different from existing TTA FAS methods [37, 38, 89, 117], *our model does not require any parameter update at the test time, which is more flexible in real-world scenarios*. Moreover, *our TTDG method can be seamlessly integrated into not only the CNN backbone but also the ViT backbone*.

4. Experiments

4.1. Experimental Setup

Datasets and Protocols. We conducted experiments on four public FAS datasets, namely CASIA-MFSD (C) [112], Idiap Replay-Attack (I) [10], MSU-MFSD (M) [98], OULU-NPU (O) [6] to verify the efficacy of our approach. These datasets include print, paper cut, and replay attacks, and were gathered using different capturing devices, including diverse illumination conditions, various background scenes, and racial demographics. Thus, there exist substantial domain shifts among these datasets. For all experiments, we strictly follow the same experimental protocols as previous DG FAS methods [30, 54, 55, 76, 77, 80, 116].

Implementation Details: Following the precedent setting of previous DG FAS methods [30, 54, 55, 76, 77], we employ the same CNN [30, 108] and ViT-Base [12] backbone to ensure fair comparisons. During training, the hyperparameter λ_c is empirically set to 0.4, and N is set to 64 for all experiments. Following prior works [54, 55, 116, 121], we utilize pseudo-depth maps generated by PRNet [15] for depth supervision and set $\lambda_d = 0.1$ when training with the CNN-based backbone. As for training with ViT [12], we follow [20, 47, 95] and do not use the depth estimator ($\lambda_d = 0$). The Half Total Error Rate (HTER) and the Area Under Curve (AUC) are used as evaluation metrics. The lower HTER and higher AUC indicate better performance.

Methods	I&C&M to O		O&C&M to I		O&C&I to M		O&M&I to C	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
MADDG [76]	27.98	80.02	22.19	84.99	17.69	88.06	24.50	84.51
D^2AM [9]	15.27	90.87	15.43	91.22	12.70	95.66	20.98	85.58
SSDG [30]	25.17	81.83	18.21	94.61	16.67	90.47	23.11	85.45
RFM [77]	16.45	91.16	17.30	90.48	13.89	93.98	20.27	88.16
DRDG [55]	15.63	91.75	15.56	91.79	12.43	95.81	19.05	88.79
ANRL [54]	15.67	91.90	16.03	91.04	10.83	96.75	17.85	89.26
FGHV [56]	13.58	93.55	16.29	90.11	9.17	96.92	12.47	93.47
SSAN [96]	19.51	88.17	14.00	94.58	10.42	94.76	16.47	90.81
AMEL [116]	11.31	93.96	18.60	88.79	10.23	96.62	11.88	94.39
EBDG [13]	15.66	92.02	18.69	92.28	9.56	**97.17**	18.34	90.01
IADG [121]	11.45	94.50	11.04	93.15	8.45	96.99	12.74	94.00
Ours (TTDG)	**10.00**	**95.70**	**6.50**	**97.98**	**7.91**	96.83	**8.14**	**96.49**
ViTranZFAS [20]	15.67	89.59	16.64	85.07	10.95	95.05	14.33	92.10
TTN-S [95]	12.64	94.20	14.15	94.06	9.58	95.79	9.81	95.07
DiVT-V [47]	18.06	90.21	**5.71**	97.73	10.00	96.64	14.67	93.08
Ours (TTDG-V)	**10.00**	**96.15**	9.62	**98.18**	**4.16**	**98.48**	**7.59**	**98.18**

Table 1. Comparison with the state-of-art FAS methods on four testing domains. TTDG-V denotes TTDG with ViT-Base [12] backbone.

Methods	I&C&M to O		O&C&M to I		O&C&I to M		O&M&I to C	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
DCN [33]	15.52	90.44	18.75	87.23	14.16	95.19	15.74	91.51
TF-Cal [113]	13.29	93.71	19.75	90.35	12.08	95.58	14.26	92.10
Sty.-Pro [29]	13.19	93.69	14.25	91.63	14.58	92.60	14.81	92.63
Ours (TTDG)	**10.00**	**95.70**	**6.50**	**97.98**	**7.91**	**96.83**	**8.14**	**96.49**

Table 2. Comparison with test-time domain generalization methods. The bold numbers indicate the best performance.

Methods	M&I to C		M&I to O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)
MADDG [76]	41.02	64.33	39.35	65.10
SSDG [30]	31.89	71.29	36.01	66.88
D^2AM [9]	32.65	72.04	27.70	75.36
DRDG [55]	31.28	71.50	33.35	69.14
ANRL [54]	31.06	72.12	30.73	74.10
SSAN [96]	30.00	76.20	29.44	76.62
EBDG [13]	27.97	75.84	25.94	78.28
AMEL [116]	24.52	82.12	19.68	87.01
IADG [121]	23.51	84.20	22.70	84.28
Ours (TTDG)	**17.77**	**86.69**	**17.70**	**90.09**

Table 3. Comparison results on limited source domains.

4.2. Comparisons to the State-of-the-art Methods

Comparison Results on Leave-One-Out Settings. As shown in Table 1, Table 2, we verify our proposed method in four standard leave-one-out settings. Note that in each experiment, all methods are compared using the same backbone to ensure fair comparisons. In all experiments, IADG [121] is implemented by using the same backbone that removes the DKG module. From the tables, we draw the following observations. (1) Our proposed TTDG method consistently outperforms the majority of state-of-the-art DG FAS methods [13, 30, 47, 54, 55, 76, 77, 116] under five testing settings. This superiority can be attributed to the fact that most existing approaches tend to overlook the role that testing data plays in enhancing the generalizability of the FAS model beyond mere evaluation, resulting in sub-optimal performances. In contrast, we introduce test-time DG for FAS that leads to substantial improvements. (2) Existing test-time DG approaches [27, 33, 113] exhibit less-desired performances in these benchmark settings. The reasons lie in two aspects. Firstly, they tend to roughly define style bases, which is inapplicable to FAS tasks that typically have a mixture of domains. In contrast, we design N=64 learnable style bases to automatically capture style shifts in a fine-grained manner. Besides, their projections lack a clear direction for optimization, making projection less reliable for FAS. Conversely, we propose explicit optimization goals ($\mathcal{L}_{\text{sty}}$&$\mathcal{L}_{\text{con}}$) to facilitate the projection process.

Comparison Results on Limited Source Domains. Following previous works [13, 30, 54, 55, 76], we also evaluate our method on limited source domains. Table 3 shows that our method outperforms state-of-the-art approaches by

Methods	I&C&M to O		O&C&M to I		O&C&I to M		O&M&I to C	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
TF-Cal [113]	12.74	93.50	14.00	93.18	13.75	91.06	13.33	93.36
TTSS [71]	15.10	93.30	12.50	92.91	9.58	95.88	13.14	93.65
Sty.-Pro [29]	13.05	93.66	11.75	92.51	11.25	94.98	13.51	93.27
Ours (TTSP)	**10.00**	**95.70**	**6.50**	**97.98**	**7.91**	**96.83**	**8.14**	**96.49**

Table 4. Ablation studies on different test-time style shifting strategies on four testing domains.

Methods	I&C&M to O		O&C&M to I		O&C&I to M		O&M&I to C	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
Random Selection	13.40	92.81	15.12	89.56	13.75	93.50	13.51	92.72
FPS [74] Selection	11.35	95.14	10.12	95.42	10.00	94.64	12.40	94.55
Learnable (w/o DSSS)	13.64	93.18	13.75	93.68	13.33	95.02	16.48	92.67
Learnable (w DSSS)	**10.00**	**95.70**	**6.50**	**97.98**	**7.91**	**96.83**	**8.14**	**96.49**

Table 5. Ablation studies on different selection strategies of style bases on four testing domains.

Loss				I&C&M to O		O&C&M to I		O&C&I to M		O&M&I to C	
$\mathcal{L}_{\text{cls}}$	$\mathcal{L}_{\text{dep}}$	$\mathcal{L}_{\text{sty}}$	$\mathcal{L}_{\text{con}}$	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
✓	–	–	–	15.96	90.77	16.40	91.65	16.16	92.44	18.53	89.77
✓	✓	–	–	13.64	93.18	13.75	93.68	13.33	95.02	16.48	92.67
✓	✓	✓	–	13.05	93.90	12.62	94.42	10.41	94.35	16.29	93.72
✓	✓	–	✓	11.63	94.32	10.25	95.78	9.10	95.06	12.03	93.95
✓	✓	✓	✓	**10.00**	**95.70**	**6.50**	**97.98**	**7.91**	**96.83**	**8.14**	**96.49**

Table 6. Ablation studies on each loss (with TTSP) on four testing domains.

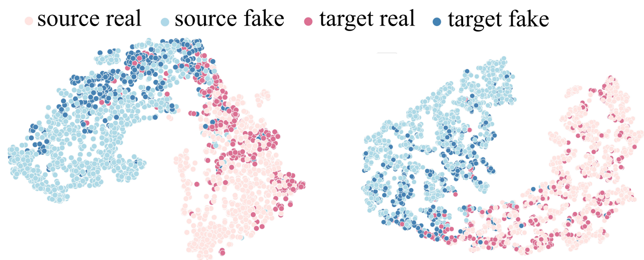
a significant margin (5% $\sim$ 6% in HTER) when dealing with extremely limited source domains. This also demonstrates that our TTDG remains effective when applied to unseen domains, regardless of the number of source domains. In contrast to previous methods, TTDG does not require any domain labels and is more flexible in realistic scenarios.

4.3. Ablation Studies

Effects of Various Test-Time Style Shifting Schemes. Table 4 shows the effect of different test-time shifting schemes while preserving the DSSS unchanged. TF-Cal [113] directly shifts the amplitude of the Fourier representation to the source prototype in a simple manner. Similarly, TTSS [71] shifts the style statistics of the test sample to the nearest source domain before making the prediction. However, they neglect the contribution of other similar source domains and achieve less desirable outcomes. Sty.-Pro [29] directly projects the unseen samples to bases via the Wasserstein distance [83], which cannot be directly applied to the hyper-spherical feature space we constructed and shows less-desired results in DG FAS (Table 4). In contrast, we introduce a cosine distance-based similarity (Eq. (5)) and perform the projection into the same space, which is more suitable for FAS task. Thus, our TTSP outperforms various test-time style shifting strategies by a large margin.

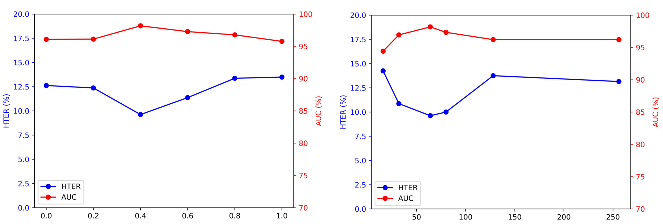
Impacts of Different Style Bases Selection Strategies. Table 5 illustrates the impact of various style bases selection strategies while keeping TTSP consistent. Random selection means randomly selecting N style bases from all source domains. It shows inferior results in four domains since it cannot fully represent the style representation space. FPS [74] selects the bases according to the farthest point sampling strategy and achieves better results. However, both of them are time-expensive since they need to re-select the bases from all the source domains in each epoch. TTDG (w/o DSSS) means style bases are learned with task-related losses ($\mathcal{L}_{\text{cls}}$ & $\mathcal{L}_{\text{dep}}$) only and such randomly learnable bases are diverse to some extent. In contrast, our learnable bases under two proposed losses are more effective, further improving the performance by a large margin.

Contribution of Each Loss. Table 6 demonstrates the contribution of each proposed loss with TTSP. When we train style bases using $\mathcal{L}_{\text{style}}$ but without $\mathcal{L}_{\text{content}}$, we observe limited performance improvements compared to the baseline. This is because the style-content features obtained become more diverse within the same class but lack content consistency. Conversely, merely using $\mathcal{L}_{\text{content}}$ without $\mathcal{L}_{\text{style}}$ achieves a certain performance boost since it encourages the style-content features to be more consistent with the content features but lacks style diversity. Finally, only by jointly incorporating both losses will we achieve the best results. This shows that our TTDG needs to be trained under the guidance of both loss functions ($\mathcal{L}_{\text{style}}$ & $\mathcal{L}_{\text{content}}$).



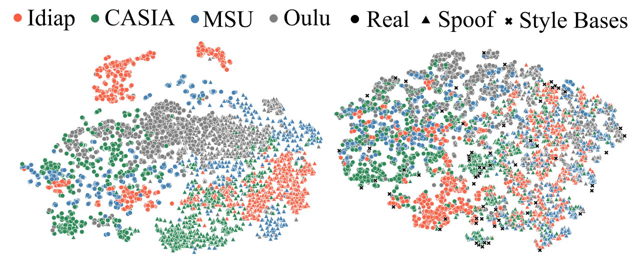
(a) Train-Time DG (b) Test-Time DG

Figure 3. Comparison results of t-SNE [84] feature visualization for train-time DG and our test-time DG method.



(a) Effect of the weights λ_c of $L_{content}$ (b) Effect on the number N of style bases.

Figure 4. Hyper-parameter analyses on the O&C&M to I setting.



(a) Before Projection (b) After Projection

Figure 5. T-SNE [84] visualization of features for different domains before (a) and after test-time style projection (b).

4.4. Visualization and Analysis

T-SNE Visualization of Feature Distributions. To reveal how testing data leads to the generalizability boost, we employ the t-SNE [84] visualization tool on the I&C&M to O setting to analyze the effectiveness of our proposed method.

Figure 3 shows the feature distributions between train-time DG [116] and our test-time DG method. We have two observations: (1) In Figure 3 (a), testing samples near the decision boundary are almost misclassified, where the previous method is ineffective, while points away from the decision boundary are well-classified. (2) Although source samples in Figure 3 (a) are well-separated, many target samples are misclassified, while our method in Figure 3 (b) has much fewer misclassified samples, which verifies ours superiority. (3) In Figure 3 (b), our source and target domains exhibit better alignment, indicating better generalizability. This is because the target domain varies across samples, making the weights of the selected bases different in TTDG.

Figure 5 illustrates the variations of style distributions between different domains before and after style projection. We have three observations as follows: (1) Before style projection (Figure 5 (a)), it is evident that the style distribution of distinct domains is separated. After style projection, the style distribution of the unseen domain is approximately situated within the style bases. (2) Furthermore, the unseen domain aligns more closely with the source domains (Figure 5 (b)), demonstrating that TTSP successfully projects unseen styles into the seen space. (3) Finally, the learnable style bases are diverse enough to represent the whole space, and most of them lie in the outlier of the style representation space. When TTDG encounters an unseen sample (O), they often relate it to a previously perceived similar one (C), and thus some of the bases are near the domain C.

Hyper-parameter Analysis. During the optimization, it is essential to balance the weight between different losses. We study the impact of λ_c on TTDG-V. As shown in Figure 4 (a), reducing λ_c may not significantly facilitate the training process, while increasing it too much can result in the propagation of incorrect gradients throughout the network. Based on our empirical findings, we set λ_c to 0.4 for all experiments. Next, we analyzed the impact on the number N of style bases of TTDG-V in Figure 4 (b). A smaller value of N is insufficient for representing the source style space, causing the model to become overly specific and resulting in poor generalization. Conversely, a larger value of N introduces redundant bases, leading to less desirable outcomes. Thus, we set N to a default value of 64 in experiments.

5. Conclusion

In this paper, we present a new perspective for DG FAS that leverages testing data to enhance the generalizability beyond mere evaluation. We propose a novel Test-Time Domain Generalization (TTDG) framework for FAS, which is the first work that studies test-time DG for FAS. Specifically, we introduce Test-Time Style Projection (TTSP) to project the styles of the unseen samples to the source domains via the aggregation of a set of style bases. In addition, we design Diverse Style Shifts Simulation (DSSS) to synthesize diverse distribution shifts via learnable style bases in a hyperspherical feature space, thereby promoting the test-time DG. Extensive experiments demonstrate the state-of-the-art performance and the effectiveness of our TTDG on widely used cross-domain FAS benchmarks.

6. Acknowledgement

The work is supported by Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102), Shanghai Science and Technology Commission (21511101200), National Natural Science Foundation of China (No. 72192821), YuCaiKe (No. 14105167-2023).

References

[1] Isabela Albuquerque, João Monteiro, Mohammad Darvishi, Tiago H Falk, and Ioannis Mitliagkas. Generalizing to unseen domains via distribution matching. *arXiv preprint arXiv:1911.00804*, 2019. 3

[2] Yousef Atoum, Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Face anti-spoofing using patch and depth-based cnns. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 319–328, 2017. 2

[3] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning (ML)*, 79:151–175, 2010. 3

[4] Zinelabidine Boulkenafet, Jukka Komulainen, Abdenour Hadid, et al. Face anti-spoofing based on color texture analysis. In *IEEE International Conference on Image Processing (ICIP)*, pages 2636–2640, 2015. 2

[5] Zinelabidine Boulkenafet, Jukka Komulainen, Abdenour Hadid, et al. Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security (TIFS)*, 11(8):1818–1830, 2017. 1

[6] Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. Oulu-npu: A mobile face presentation attack database with real-world variations. In *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 612–618, 2017. 5

[7] Rizhao Cai, Zhi Li, Renjie Wan, Haoliang Li, Yongjian Hu, and Alex C Kot. Learning meta pattern for face anti-spoofing. *IEEE Transactions on Information Forensics and Security (TIFS)*, 17:1201–1213, 2022. 1

[8] Dian Chen, Dequan Wang, Trevor Darrell, and Sayna Ebrahimi. Contrastive test-time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 295–305, 2022. 2

[9] 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. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 1132–1139, 2021. 1, 2, 3, 6

[10] Ivana Chingovska, André Anjos, Sébastien Marcel, et al. On the effectiveness of local binary patterns in face anti-spoofing. In *International Conference of Biometrics Special Interest Group (BIOSIG)*, pages 1–7, 2012. 5

[11] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4690–4699, 2019. 1

[12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2020. 2, 5, 6

[13] Zhekai Du, Jingjing Li, Lin Zuo, Lei Zhu, and Ke Lu. Energy-based domain generalization for face anti-spoofing. In *Proceedings of the 30th ACM International Conference on Multimedia (ACM MM)*, pages 1749–1757, 2022. 1, 2, 3, 6

[14] Litong Feng, Lai-Man Po, Yuming Li, Xuyuan Xu, Fang Yuan, Terence Chun-Ho Cheung, and Kwok-Wai Cheung. Integration of image quality and motion cues for face anti-spoofing: A neural network approach. *Journal of Visual Communication and Image Representation (JVCIR)*, 38: 451–460, 2016. 1, 2

[15] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. Joint 3d face reconstruction and dense alignment with position map regression network. In *European Conference on Computer Vision (ECCV)*, pages 534–551, 2018. 5

[16] Zhengyang Feng, Qianyu Zhou, Qiqi Gu, Xin Tan, Guangliang Cheng, Xuequan Lu, Jianping Shi, and Lizhuang Ma. Dmt: Dynamic mutual training for semi-supervised learning. *Pattern Recognition (PR)*, page 108777, 2022. 2

[17] Tiago Freitas Pereira, Jukka Komulainen, André Anjos, José De Martino, Abdenour Hadid, Matti Pietikäinen, and Sébastien Marcel. Face liveness detection using dynamic texture. *Eurasip Journal on Image and Video Processing (JIVP)*, 2014(1):1–15, 2014. 1

[18] Tiago de Freitas Pereira, André Anjos, José Mario De Martino, and Sébastien Marcel. Lbp- top based countermeasure against face spoofing attacks. In *Asian Conference on Computer Vision (ACCV)*, pages 121–132, 2012. 1, 2

[19] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning (ICML)*, pages 1180–1189, 2015. 2

[20] Anjith George and Sébastien Marcel. On the effectiveness of vision transformers for zero-shot face anti-spoofing. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8, 2021. 2, 5, 6

[21] Qiqi Gu, Qianyu Zhou, Minghao Xu, Zhengyang Feng, Guangliang Cheng, Xuequan Lu, Jianping Shi, and Lizhuang Ma. Pit: Position-invariant transform for cross-fov domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8761–8770, 2021. 2

[22] Shaohua Guo, Qianyu Zhou, Ye Zhou, Qiqi Gu, Junshu Tang, Zhengyang Feng, and Lizhuang Ma. Label-free regional consistency for image-to-image translation. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2021. 2

[23] Zong-Wei Hong, Yu-Chen Lin, Hsuan-Tung Liu, Yi-Ren Yeh, and Chu-Song Chen. Domain-generalized face anti-spoofing with unknown attacks. In *IEEE International Conference on Image Processing (ICIP)*, pages 820–824, 2023. 2

[24] Chengyang Hu, Junyi Cao, Ke-Yue Zhang, Taiping Yao, Shouhong Ding, and Lizhuang Ma. Structure destruction and content combination for generalizable anti-spoofing. *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 4(4):508–521, 2022. 2

[25] Chengyang Hu, Ke-Yue Zhang, Taiping Yao, Shouhong Ding, and Lizhuang Ma. Rethinking generalizable face anti-spoofing via hierarchical prototype-guided distribution refinement in hyperbolic space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2

[26] Chengyang Hu, Ke-Yue Zhang, Taiping Yao, Shice Liu, Shouhong Ding, Xin Tan, and Lizhuang Ma. Domain-hallucinated updating for multi-domain face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 2193–2201, 2024. 2

[27] 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. In *European Conference on Computer Vision (ECCV)*, pages 37–54, 2022. 6

[28] Pei-Kai Huang, Cheng-Hsuan Chiang, Jun-Xiong Chong, Tzu-Hsien Chen, Hui-Yu Ni, and Chiou-Ting Hsu. Ldc-former: Incorporating learnable descriptive convolution to vision transformer for face anti-spoofing. In *IEEE International Conference on Image Processing (ICIP)*, pages 121–125, 2023. 2

[29] Wei Huang, Chang Chen, Yong Li, Jiacheng Li, Cheng Li, Fenglong Song, Youliang Yan, and Zhiwei Xiong. Style projected clustering for domain generalized semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3061–3071, 2023. 3, 6, 7

[30] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Single-side domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8484–8493, 2020. 1, 2, 3, 5, 6

[31] Yunpei Jia, Jie Zhang, and Shiguang Shan. Dual-branch meta-learning network with distribution alignment for face anti-spoofing. *IEEE Transactions on Information Forensics and Security (TIFS)*, 17:138–151, 2021. 1, 2, 3

[32] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Unified unsupervised and semi-supervised domain adaptation network for cross-scenario face anti-spoofing. *Pattern Recognition (PR)*, 115:107888, 2021. 2

[33] Yuxuan Jiang, Yanfeng Wang, Ruipeng Zhang, Qinwei Xu, Ya Zhang, Xin Chen, and Qi Tian. Domain-conditioned normalization for test-time domain generalization. In *European Conference on Computer Vision (ECCV)*, pages 291–307, 2022. 3, 6

[34] Amin Jourabloo, Yaojie Liu, and Xiaoming Liu. Face de-spoofing: Anti-spoofing via noise modeling. In *European Conference on Computer Vision (ECCV)*, pages 290–306, 2018. 2

[35] Ira Kemelmacher-Shlizerman, Steven M Seitz, Daniel Miller, and Evan Brossard. The megaface benchmark: 1 million faces for recognition at scale. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4873–4882, 2016. 1

[36] Taewook Kim, YongHyun Kim, Inhan Kim, and Daijin Kim. Basn: Enriching feature representation using bipartite auxiliary supervisions for face anti-spoofing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 494–503, 2019. 2

[37] Young-Eun Kim, Woo-Jeoung Nam, Kyungseo Min, and Seong-Whan Lee. Style-guided domain adaptation for face presentation attack detection. *arXiv preprint arXiv:2203.14565*, 2022. 2, 4, 5

[38] Young-Eun Kim, Woo-Jeoung Nam, Kyungseo Min, and Seong-Whan Lee. Style selective normalization with meta learning for test-time adaptive face anti-spoofing. *Expert Systems with Applications (ESWA)*, 214:119106, 2023. 2, 4, 5

[39] Jukka Komulainen, Abdenour Hadid, Matti Pietikäinen, et al. Context based face anti-spoofing. In *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–8, 2013. 1, 2

[40] Youngjun Kwak, Minyoung Jung, Hunjae Yoo, JinHo Shin, and Changick Kim. Liveness score-based regression neural networks for face anti-spoofing. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023. 2

[41] Binh M Le and Simon S Woo. Gradient alignment for cross-domain face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 3

[42] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta-learning for domain generalization. In *Proceedings of the AAAI conference on artificial intelligence (AAAI)*, pages 3490–3497, 2018. 2

[43] Haoliang Li, Wen Li, Hong Cao, Shiqi Wang, Feiyue Huang, and Alex C. Kot. Unsupervised domain adaptation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security (TIFS)*, 13(7):1794–1809, 2018. 2

[44] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5400–5409, 2018. 2

[45] Lei Li, Xiaoyi Feng, Zinelabidine Boulkenafet, Zhaoqiang Xia, Mingming Li, and Abdenour Hadid. An original face anti-spoofing approach using partial convolutional neural network. In *International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6, 2016. 2

[46] Shen Li, Jianqing Xu, Xiaqing Xu, Pengcheng Shen, Shaoxin Li, and Bryan Hooi. Spherical confidence learning for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15629–15637, 2021. 1

[47] Chen-Hao Liao, Wen-Cheng Chen, Hsuan-Tung Liu, Yi-Ren Yeh, Min-Chun Hu, and Chu-Song Chen. Domain invariant vision transformer learning for face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6098–6107, 2023. 2, 5, 6

[48] Bofan Lin, Xiaobai Li, Zitong Yu, and Guoying Zhao. Face liveness detection by rppg features and contextual patch-based cnn. In *Proceedings of the 3rd International Confer-

ence on Biometric Engineering and Applications (ICBEA), pages 61–68, 2019. 1

[49] Xun Lin, Shuai Wang, Rizhao Cai, Yizhong Liu, Ying Fu, Zitong Yu, Wenzhong Tang, and Alex Kot. Suppress and rebalance: Towards generalized multi-modal face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 3

[50] 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)*, pages 1180–1186, 2022. 2

[51] 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 (TIFS)*, 16:2759–2772, 2021. 1, 2

[52] Ajian Liu, Zichang Tan, Zitong Yu, Chenxu Zhao, Jun Wan, Yanyan Liang Zhen Lei, Du Zhang, Stan Z Li, and Guodong Guo. Fm-vit: Flexible modal vision transformers for face anti-spoofing. *IEEE Transactions on Information Forensics and Security (TIFS)*, 18:4775–4786, 2023. 2

[53] Fengqi Liu, Jingyu Gong, Qianyu Zhou, Xuequan Lu, Ran Yi, Yuan Xie, and Lizhuang Ma. Cloudmix: Dual mixup consistency for unpaired point cloud completion. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2024. 2

[54] Shubao Liu, Ke-Yue Zhang, Taiping Yao, Mingwei Bi, Shouhong Ding, Jilin Li, Feiyue Huang, and Lizhuang Ma. Adaptive normalized representation learning for generalizable face anti-spoofing. In *Proceedings of the 29th ACM International Conference on Multimedia (ACM MM)*, pages 1469–1477, 2021. 1, 2, 3, 5, 6

[55] Shubao Liu, Ke-Yue Zhang, Taiping Yao, Kekai Sheng, Shouhong Ding, Ying Tai, Jilin Li, Yuan Xie, and Lizhuang Ma. Dual reweighting domain generalization for face presentation attack detection. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 867–873, 2021. 1, 2, 3, 5, 6

[56] Shice Liu, Shitao Lu, Hongyi Xu, Jing Yang, Shouhong Ding, and Lizhuang Ma. Feature generation and hypothesis verification for reliable face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 1782–1791, 2022. 6

[57] Yaojie Liu and Xiaoming Liu. Spoof trace disentanglement for generic face anti-spoofing. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 45(3): 3813–3830, 2022. 2

[58] Yaojie Liu, Amin Jourabloo, Xiaoming Liu, et al. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 389–398, 2018. 2, 5

[59] Yaojie Liu, Joel Stehouwer, Amin Jourabloo, and Xiaoming Liu. Deep tree learning for zero-shot face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4680–4689, 2019. 2

[60] Yaojie Liu, Joel Stehouwer, Xiaoming Liu, et al. On disentangling spoof trace for generic face anti-spoofing. In *European Conference on Computer Vision (ECCV)*, pages 406–422, 2020. 2

[61] Yuchen Liu, Yabo Chen, Wenrui Dai, Mengran Gou, Chun-Ting Huang, and Hongkai Xiong. Source-free domain adaptation with contrastive domain alignment and self-supervised exploration for face anti-spoofing. In *European Conference on Computer Vision (ECCV)*, pages 511–528, 2022. 2

[62] Yuchen Liu, Yabo Chen, Wenrui Dai, Chenglin Li, Junni Zou, and Hongkai Xiong. Causal intervention for generalizable face anti-spoofing. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 01–06, 2022. 2

[63] Yuchen Liu, Yabo Chen, Mengran Gou, Chun-Ting Huang, Yaoming Wang, Wenrui Dai, and Hongkai Xiong. Towards unsupervised domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 20654–20664, 2023. 2

[64] Yuchen Liu, Yabo Chen, Wenrui Dai, Mengran Gou, Chun-Ting Huang, and Hongkai Xiong. Source-free domain adaptation with domain generalized pretraining for face anti-spoofing. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2024. 2

[65] Shaocong Long, Qianyu Zhou, Chenhao Ying, Lizhuang Ma, and Yuan Luo. Diverse target and contribution scheduling for domain generalization. *arXiv preprint arXiv:2309.16460*, 2023. 2

[66] Shaocong Long, Qianyu Zhou, Chenhao Ying, Lizhuang Ma, and Yuan Luo. Rethinking domain generalization: Discriminability and generalizability. *arXiv preprint arXiv:2309.16483*, 2023. 2

[67] Jukka Määttä, Abdenour Hadid, and Matti Pietikäinen. Face spoofing detection from single images using micro-texture analysis. In *International Joint Conference on Biometrics (IJCB)*, pages 1–7, 2011. 1, 2

[68] Sachin Mehta and Mohammad Rastegari. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*, 2021. 2

[69] Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *International Conference on Machine Learning (ICML)*, pages 16888–16905, 2022. 2

[70] Ankush Panwar, Pratyush Singh, Suman Saha, Danda Pani Paudel, and Luc Van Gool. Unsupervised compound domain adaptation for face anti-spoofing. In *16th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–8, 2021. 2

[71] Jungwuk Park, Dong-Jun Han, Soyeong Kim, and Jaekyun Moon. Test-time style shifting: Handling arbitrary styles in domain generalization. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, pages 27114–27131, 2023. 3, 7

[72] Keyurkumar Patel, Hu Han, Anil K Jain, and otehrs. Secure face unlock: Spoof detection on smartphones. *IEEE Transactions on Information Forensics and Security (TIFS)*, 11 (10):2268–2283, 2016. 2

[73] Keyurkumar Patel, Hu Han, Anil K Jain, et al. Cross-database face antispoofing with robust feature representation. In *Chinese Conference on Biometric Recognition (CCBR)*, pages 611–619, 2016. 2

[74] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems (NeurIPS)*, 30, 2017. 7

[75] Yunxiao Qin, Zitong Yu, Longbin Yan, Zezheng Wang, Chenxu Zhao, and Zhen Lei. Meta-teacher for face anti-spoofing. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 44(10):6311–6326, 2021. 2

[76] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C Yuen. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10023–10031, 2019. 1, 2, 3, 5, 6

[77] Rui Shao, Xiangyuan Lan, and Pong C Yuen. Regularized fine-grained meta face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 11974–11981, 2020. 5, 6

[78] Yiran Song, Qianyu Zhou, Xiangtai Li, Deng-Ping Fan, Xuequan Lu, and Lizhuang Ma. Ba-sam: Scalable bias-mode attention mask for segment anything model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2

[79] Yiran Song, Qianyu Zhou, Xuequan Lu, Zhiwen Shao, and Lizhuang Ma. Simada: A simple unified framework for adapting segment anything model in underperformed scenes. *arXiv preprint arXiv:2401.17803*, 2024. 2

[80] Yiyou Sun, Yaojie Liu, Xiaoming Liu, Yixuan Li, and Wen-Sheng Chu. Rethinking domain generalization for face anti-spoofing: Separability and alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 24563–24574, 2023. 5

[81] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International Conference on Machine Learning (ICML)*, pages 10347–10357, 2021. 2

[82] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7167–7176, 2017. 2

[83] SS Vallender. Calculation of the wasserstein distance between probability distributions on the line. *Theory of Probability & Its Applications*, 18(4):784–786, 1974. 7

[84] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research (JMLR)*, 9(11), 2008. 8

[85] Chien-Yi Wang, Yu-Ding Lu, Shang-Ta Yang, and Shang-Hong Lai. Patchnet: A simple face anti-spoofing framework via fine-grained patch recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20281–20290, 2022. 2

[86] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020. 2

[87] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection. *IEEE Transactions on Information Forensics and Security (TIFS)*, 16:56–69, 2021. 2

[88] Jun Wang, Yinglu Liu, Yibo Hu, Hailin Shi, and Tao Mei. Facex-zoo: A pytorch toolbox for face recognition. In *Proceedings of the 29th ACM International Conference on Multimedia (ACM MM)*, pages 3779–3782, 2021. 1

[89] Jingjing Wang, Jingyi Zhang, Ying Bian, Youyi Cai, Chunmao Wang, and Shiliang Pu. Self-domain adaptation for face anti-spoofing. In *Proceedings of the AAAI conference on artificial intelligence (AAAI)*, pages 2746–2754, 2021. 2, 5

[90] Jiong Wang, Zhou Zhao, Weike Jin, Xinyu Duan, Zhen Lei, Baoxing Huai, Yiling Wu, and Xiaofei He. Vlad-vsa: Cross-domain face presentation attack detection with vocabulary separation and adaptation. In *Proceedings of the 29th ACM International Conference on Multimedia (ACM MM)*, pages 1497–1506, 2021. 2

[91] Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7201–7211, 2022. 2

[92] Yahang Wang, Xiaoning Song, Tianyang Xu, Zhenhua Feng, and Xiao-Jun Wu. From rgb to depth: domain transfer network for face anti-spoofing. *IEEE Transactions on Information Forensics and Security (TIFS)*, 16:4280–4290, 2021. 2

[93] 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. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5042–5051, 2020. 2

[94] Zhuo Wang, Qiangchang Wang, Weihong Deng, and Guodong Guo. Face anti-spoofing using transformers with relation-aware mechanism. *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 4(3):439–450, 2022. 2

[95] Zhuo Wang, Qiangchang Wang, Weihong Deng, and Guodong Guo. Learning multi-granularity temporal characteristics for face anti-spoofing. *IEEE Transactions on Information Forensics and Security (TIFS)*, 17:1254–1269, 2022. 1, 2, 5, 6

[96] Zhuo Wang, Zezheng Wang, Zitong Yu, Weihong Deng, Jiahong Li, Tingting Gao, and Zhongyuan Wang. Domain generalization via shuffled style assembly for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4123–4133, 2022. 1, 2, 3, 4, 6

[97] Zezheng Wang, Zitong Yu, Xun Wang, Yunxiao Qin, Jiahong Li, Chenxu Zhao, Xin Liu, and Zhen Lei. Consistency regularization for deep face anti-spoofing. *IEEE Transactions on Information Forensics and Security (TIFS)*, 18:1127–1140, 2023. 1

[98] Di Wen, Hu Han, Anil K Jain, et al. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Securityn (TIFS)*, 10(4):746–761, 2015. 1, 5

[99] Hangtong Wu, Dan Zeng, Yibo Hu, Hailin Shi, and Tao Mei. Dual spoof disentanglement generation for face anti-spoofing with depth uncertainty learning. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 32(7):4626–4638, 2021. 2

[100] Xiaojun Wu, Jinghui Zhou, Jun Liu, Fangyi Ni, and Haoqiang Fan. Single-shot face anti-spoofing for dual pixel camera. *IEEE Transactions on Information Forensics and Security (TIFS)*, 16:1440–1451, 2020. 2

[101] Hongyi Xu, Fengqi Liu, Qianyu Zhou, Jinkun Hao, Zhijie Cao, Zhengyang Feng, and Lizhuang Ma. Semi-supervised 3d object detection via adaptive pseudo-labeling. In *IEEE International Conference on Image Processing (ICIP)*, pages 3183–3187, 2021. 2

[102] Jianwei Yang, Zhen Lei, Shengcai Liao, and Stan Z Li. Face liveness detection with component dependent descriptor. In *IEEE International Conference on Biometrics (ICB)*, pages 1–6, 2013. 1, 2

[103] Jianwei Yang, Zhen Lei, Stan Z Li, et al. Learn convolutional neural network for face anti-spoofing. In *arXiv preprint arXiv:1408.5601*, 2014. 2

[104] Zitong Yu, Xiaobai Li, Xuesong Niu, Jingang Shi, and Guoying Zhao. Face anti-spoofing with human material perception. In *European Conference on Computer Vision (ECCV)*, pages 557–575, 2020. 2

[105] 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. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 43(9):3005–3023, 2020.

[106] 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. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5295–5305, 2020.

[107] Zitong Yu, Xiaobai Li, Jingang Shi, Zhaoqiang Xia, and Guoying Zhao. Revisiting pixel-wise supervision for face anti-spoofing. *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, 3(3):285–295, 2021. 1, 2

[108] Zitong Yu, Yunxiao Qin, Hengshuang Zhao, Xiaobai Li, and Guoying Zhao. Dual-cross central difference network for face anti-spoofing. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1281–1287, 2021. 5

[109] 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 *European Conference on Computer Vision (ECCV)*, pages 641–657, 2020. 1, 2

[110] Ke-Yue Zhang, Taiping Yao, Jian Zhang, Shice Liu, Bangjie Yin, Shouhong Ding, and Jilin Li. Structure destruction and content combination for face anti-spoofing. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–6, 2021. 1

[111] Marvin Zhang, Sergey Levine, and Chelsea Finn. Memo: Test time robustness via adaptation and augmentation. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:38629–38642, 2022. 2

[112] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z Li. A face antispoofing database with diverse attacks. In *5th IAPR International Conference on Biometrics (ICB)*, pages 26–31, 2012. 5

[113] Xingchen Zhao, Chang Liu, Anthony Sicilia, Seong Jae Hwang, and Yun Fu. Test-time fourier style calibration for domain generalization. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1721–1727, 2022. 3, 6, 7

[114] Guanghao Zheng, Yuchen Liu, Wenrui Dai, Chenglin Li, Junni Zou, and Hongkai Xiong. Learning causal representations for generalizable face anti spoofing. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023. 2

[115] Qianyu Zhou, Zhengyang Feng, Qiqi Gu, Guangliang Cheng, Xuequan Lu, Jianping Shi, and Lizhuang Ma. Uncertainty-aware consistency regularization for cross-domain semantic segmentation. *Computer Vision and Image Understanding (CVIU)*, page 103448, 2022. 2

[116] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Ran Yi, Shouhong Ding, and Lizhuang Ma. Adaptive mixture of experts learning for generalizable face anti-spoofing. In *Proceedings of the 30th ACM International Conference on Multimedia (ACM MM)*, pages 6009–6018, 2022. 1, 2, 3, 5, 6, 8

[117] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Ran Yi, Kekai Sheng, Shouhong Ding, and Lizhuang Ma. Generative domain adaptation for face anti-spoofing. In *European Conference on Computer Vision (ECCV)*, pages 335–356, 2022. 2, 4, 5

[118] Qianyu Zhou, Chuyun Zhuang, Xuequan Lu, and Lizhuang Ma. Domain adaptive semantic segmentation with regional contrastive consistency regularization. In *IEEE International Conference on Multimedia and Expo (ICME)*, 2022. 2

[119] Qianyu Zhou, Zhengyang Feng, Qiqi Gu, Jiangmiao Pang, Guangliang Cheng, Xuequan Lu, Jianping Shi, and Lizhuang Ma. Context-aware mixup for domain adaptive semantic segmentation. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 33(2): 804–817, 2023.

[120] Qianyu Zhou, Qiqi Gu, Jiangmiao Pang, Xuequan Lu, and Lizhuang Ma. Self-adversarial disentangling for specific domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 45(7):8954–8968, 2023. 2

[121] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Ran Yi, Shouhong Ding, and Lizhuang Ma. Instance-aware domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20453–20463, 2023. 1, 2, 4, 5, 6