Generative Domain Adaptation for Face Anti-Spoofing

Qianyu Zhou[1*†], Ke-Yue Zhang[2*], Taiping Yao[2], Ran Yi[1], Kekai Sheng[2], Shouhong Ding[2✉], and Lizhuang Ma[1✉]

[1] Shanghai Jiao Tong University, Shanghai, China
[2] Youtu Lab, Tencent, Shanghai, China
{zhouqianyu, ranyi}@sjtu.edu.cn, ma-lz@cs.sjtu.edu.cn
{zkyezhang,taipingyao,saulsheng,ericshding}@tencent.com

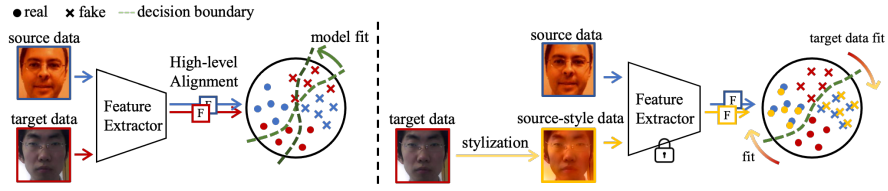
Abstract. Face anti-spoofing (FAS) approaches based on unsupervised domain adaption (UDA) have drawn growing attention due to promising performances for target scenarios. Most existing UDA FAS methods typically fit the trained models to the target domain via aligning the distribution of semantic high-level features. However, insufficient supervision of unlabeled target domains and neglect of low-level feature alignment degrade the performances of existing methods. To address these issues, we propose a novel perspective of UDA FAS that directly fits the target data to the models, *i.e.,* stylizes the target data to the source-domain style via image translation, and further feeds the stylized data into the well-trained source model for classification. The proposed Generative Domain Adaptation (GDA) framework combines two carefully designed consistency constraints: 1) Inter-domain neural statistic consistency guides the generator in narrowing the inter-domain gap. 2) Dual-level semantic consistency ensures the semantic quality of stylized images. Besides, we propose intra-domain spectrum mixup to further expand target data distributions to ensure generalization and reduce the intra-domain gap. Extensive experiments and visualizations demonstrate the effectiveness of our method against the state-of-the-art methods.

Keywords: Face anti-spoofing, unsupervised domain adaptation

1 Introduction

Face recognition (FR) techniques [12,35,71,114,45,80] have been widely utilized in identity authentication products, *e.g.,* smartphones login, access control, *etc.* Despite its gratifying progress in recent years, FR systems are vulnerable to face presentation attacks (PA), *e.g.,* printed photos, video replay, and *etc.* To protect such FR systems from various face presentation attacks, face anti-spoofing (FAS)

∗ Equal contributions.
† Work done during an internship at Youtu Lab, Tencent.
✉ Corresponding authors.



(a) Previous UDA in FAS: Model fit to Target data (b) Our UDA in FAS: Target data fit to Model

Fig. 1: Conventional UDA FAS methods typically force the model fit to the target data via performing the high-level feature alignment across domains. However, insufficient supervisions and neglect of low-level feature alignment inevitably affect the discrimination ability of FAS models. Instead, we aim to directly fit the target data to the source-trained models in a reverse manner via both the high-level and low-level alignment. By generating source-style images and predicting with a well-trained model, we address these issues without changing models

attracts great attention. Nowadays, based on hand-crafted features [2,55,16,37,88], and deeply-learned features [13,44,61,97,101,47,102], several methods achieve promising performance in intra-dataset scenarios. However, they all suffer from performance degradation when adapting to the target domains in real-world scenarios due to the domain gap across different domains.

To improve the generalization, FAS approaches based on domain generalization (DG) and unsupervised domain adaption (UDA) have been proposed on cross-domain scenarios. However, DG FAS approaches [68,31,69,49,48,9,111] only utilize the seen data in the training stage, which fail in utilizing the information of the target data, thus resulting in unsatisfactory performance on the target domain. Although UDA FAS methods [77,40,76,32,66,107,72] utilize both the labeled source domain and the unlabeled target domain to bridge the domain gap, most of them typically fit the models to the target domain via aligning the distribution of semantic high-level features, as shown in Fig. 1 (a), without considering the specificity of FAS task. On the one hand, since the insufficient supervision of the target domain, fitting to it may inevitably affect the discrimination ability of the source model. On the other hand, as pointed out in [34], low-level features are especially vital to the FAS task. Thus, the above towards-target distribution alignment based on only high-level features may not be the most suitable way for UDA FAS.

To address the above issues, we propose a novel perspective of unsupervised domain adaptation (UDA) for face anti-spoofing (FAS). Different from existing methods that fit the models to the target data, we aim to directly fit the target data to the well-trained models, keeping the source-trained models unchanged, as shown in Fig. 1 (b). To achieve such fitting, we reformulate the unsupervised domain adaptation (UDA) in FAS as a domain stylization problem to stylize the target data with the source-domain style, and the stylized data is further fed into the well-trained source model for classification. In this work, we pro-

pose Generative Domain Adaptation (GDA) framework combining two carefully designed consistency constraints. Specifically, we present inter-domain neural statistic consistency (NSC) to guide the generator toward producing the source-style images, which fully aligns the target feature statistics with the source ones in both high-levels and low-levels, and narrows the inter-domain gap efficiently. Besides, to maintain the semantic qualities and liveness information of the target data during the stylization procedure, we introduce a dual-level semantic consistency (DSC) on both image level and feature level. Moreover, intra-domain spectrum mixup (SpecMix) is presented to further expand the target data distribution to ensure that the generator could correctly translate the unseen target domain to the source-style domain. To the best of our knowledge, this is the first work that reveals the potential of image translation for UDA FAS.

Our main contributions can be summarized as follows:

• We propose a novel perspective of unsupervised domain adaptation for face anti-spoofing that directly fits the target data to the source model by stylizing the target data with the source-domain style via image translation.

• To ensure the stylization, we present a Generative Domain Adaptation framework combined with two carefully designed consistency constraints, inter-domain neural statistic consistency (NSC) and dual-level semantic consistency (DSC). And intra-domain spectrum mixup (SpecMix) is presented to further expand the target data distribution to ensure generalization.

• Extensive experiments and visualizations demonstrate the effectiveness of our proposed method against the state-of-the-art competitors.

2 Related Work

Face Anti-Spoofing. Face anti-spoofing (FAS) aims to detect a face image whether taken from a real person or various face presentation attacks [4,10,7,8]. Pioneer works utilize handcrafted features to tackle this problem, such as SIFT [62], LBP [2,55,16], and HOG [37,88]. Several methods utilize the information from different domains, such as HSV and YCrCb color spaces [2,3], temporal domains [70,1], and Fourier spectrum [43]. Recent approaches leverage CNN to model FAS with binary classification [13,44,61,100] or additional supervision, *e.g.,* depth map [97], reflection map [101] and r-ppg signal [47,27]. Other methods adopt disentanglement [102,52] and custom operators [98,96,6] to improve the performance. Despite good outcomes in the intra-dataset setting, their performances still drop significantly on target domains due to large domain shifts.

Cross-Domain Face Anti-Spoofing. To improve the performances under the cross-domain settings, domain generalization (DG) [41,39,106,57] is introduced into FAS tasks. Nevertheless, DG FAS methods [68,31,69,49,48,9,111] aim to map the samples into a common feature space and lack the specific information of the unseen domains, inevitably resulting in unsatisfactory results. Considering the availability of the unlabeled target data in real-world applications, several works tackle the above issue based on unsupervised domain adaptation (UDA)

methods. Recent studies of UDA FAS mainly rely on pseudo labeling [66,54], adversarial learning [40,77,76,32] or minimizing domain discrepancy [32,40] to narrow the domain shifts. However, they still suffer from insufficient supervision of the unlabeled target domains, which may cause the negative transfer to the source models. Besides, most works mainly focus on the alignment of high-level semantic features, overlooking the low-level features which are essential to the FAS tasks. In contrast, we aim to address these two issues for UDA FAS.

Unsupervised Domain Adaptation. Unsupervised domain adaptation (UDA) aims to bridge the domain shifts between the labeled source domain and unlabeled target domain. Recent methods focus on adversarial learning [17,73,63,56], self-training [115,116,14,84], consistency regularization [11,108,109,112], prototypical alignment [85,103,33], feature disentanglement [82,5,110] and image translation [24,21,94,93,53,30,105,20,26,25]. Despite its gratifying progress, such "model-fitting-to data" paradigm is not practical for FAS task due to plenty of different domains in real-world scenarios. Besides, the discrimination ability of the source model may also be affected during re-training. In contrast, we propose a new yet practical approach that adapts the target data to the source model, keeping the source model unchanged. To the best of our knowledge, this is the first work that reveals the potential of image translation for UDA FAS.

3 Methodology

3.1 Overview

In UDA FAS, we have access to the labeled source domain, denoted as $D_s = \{(x_s, y_s) \mid x_s \subset \mathbb{R}^{H\times W\times 3}, y_s \in [0,1]\}$, and the unlabeled target domain, denoted as $D_t = \{(x_t) \mid x_t \subset \mathbb{R}^{H\times W\times 3}\}$. Regarding that insufficient supervision and neglect of low-level feature alignment in previous UDA FAS approaches lead to inferior performances, we are motivated to perform both the high-level and low-level feature alignment and make the target data fit to the model in a reverse manner. Our training include two stages: the first phase using the source domains only for training the FAS models, including a feature extractor F, a classifier H, a depth estimator R; the second phase for domain adaptation. During the latter phase, only the image generator G is optimized, and other source models with an ImageNet pre-trained VGG module ϕ are fixed during the adaptation.

Fig. 2 shows the overall GDA framework. We aim to stylize the unlabeled target domain to the source-style domain, making the unlabeled target data fit to the source models, so that the well-trained models do not need to be changed. To mitigate the intra-domain gap, input images are firstly diversified in continuous frequency space via intra-domain spectrum mixup (SpecMix) to produce augmented images. Then, the generator translates both the original and the diversified target images into the source-style images. To extract the source style information to guide the image translation, we match the generated statistics of the source-style images with those stored source statistics in the pre-trained

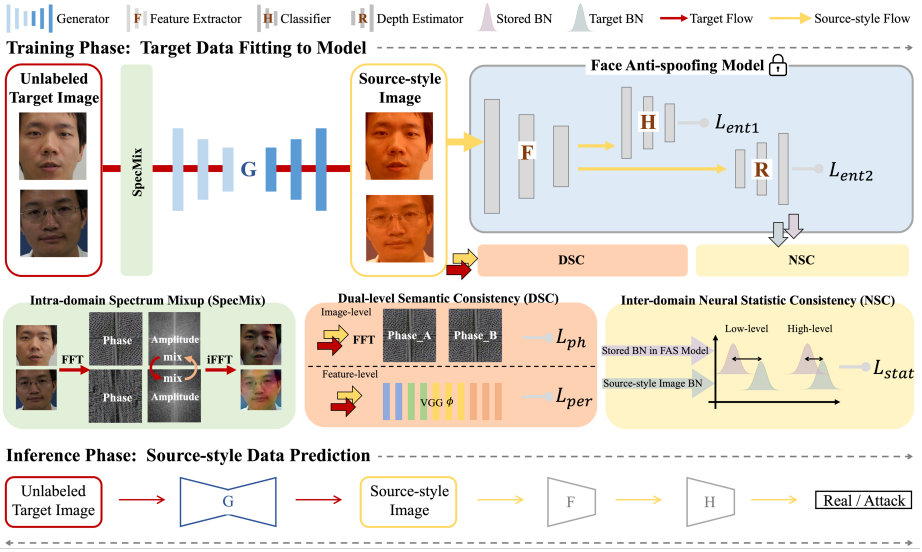


Fig. 2: Overview of Generative Domain Adaptation framework. The parameters of the source-trained models are fixed during adaptation. Given the unlabeled target data, we only optimize the parameters of the generator G. Firstly, we generate diversified target images via *intra-domain spectrum mixup* (SpecMix), thus enhancing the generalization abilities of the generator G in bridging the intra-domain gap. Then, *inter-domain neural statistic consistency* (NSC) fully matches generated feature statistics with the stored ones in high level and low levels, thus mitigating the inter-domain gap. Thus, the translated images can retain the source style. Furthermore, we introduce *dual-level semantic consistency* (DSC) to ensure content-preserving and prevent form semantic distortions

model via inter-domain neural statistic consistency (NSC), thus bridging the inter-domain gap. Finally, to preserve the target content and prevent semantic distortions during the generation, we propose a dual-level semantic consistency (DSC) on both the feature level and image level.

3.2 Generative Domain Adaptation

Inter-domain Neural Statistic Consistency. Batch normalization (BN) [29] normalizes each input feature within a mini-batch in a channel-wise manner so that the output has zero-mean and unit-variance. Let B and $\{z_i\}_{i=1}^{B}$ denote the mini-batch size and the input features to the batch normalization, respectively. The key to the BN layer is to compute the batch-wise statistics, *e.g.,* means μ and variances σ^2 of the features within the mini-batch:

$$\mu \leftarrow \frac{1}{B}\sum_{i=1}^{B} x_i, \;\; \sigma^2 \leftarrow \frac{1}{B}\sum_{i=1}^{B}\left(x_i - \mu\right)^2, \tag{1}$$

In the first phase of training FAS models, the source statistics $\bar{\mu}_s^{n+1}$ and $\bar{\sigma}_s^{n+1}$ at step $n+1$ are exponential moving average of that at step n with a ratio α:

$$\begin{aligned} \bar{\mu}_s^{n+1} &= (1-\alpha)\bar{\mu}_s^n + \alpha\mu_s^n \\ \left(\bar{\sigma}_s^{n+1}\right)^2 &= (1-\alpha)\left(\bar{\sigma}_s^n\right)^2 + \alpha\left(\sigma_s^n\right)^2 \end{aligned} \tag{2}$$

We observe that such neural statistics [67,29] of the source features stored in the well-trained FAS models provide sufficient supervisions for both the low-level and high-level features, which can represent domain-specific styles and could be fully used to aid the distribution alignment in UDA. However, the previous methods only use the output features of high-level layers for distribution alignment, and are unable to make full use of rich and discriminative liveness cues in low-level features, which is vital to FAS tasks. Thus, given those stored BN statistics, we can easily estimate the source-style distribution $D_{\tilde{s}}$, where $D_{\tilde{s}} = \left\{(x_{\tilde{s}}) \mid x_{\tilde{s}} = G(x_t) \subset \mathbb{R}^{H\times W\times 3}\right\}$.

Inspired by data-free knowledge distillation [95], we propose an *inter-domain neural statistic consistency loss* $\mathcal{L}_{\text{stat}}$ to match the feature statistics between the running mean $\bar{\mu}_{\tilde{s}}$, running variances $\bar{\sigma}_{\tilde{s}}$ of the source-style data $D_{\tilde{s}}$ and the stored statistics $\bar{\mu_s}$, $\bar{\sigma}_s$ of source models M, thus bridging the inter-domain gap:

$$\mathcal{L}_{\text{stat}} = \frac{1}{L}\sum_{l=1}^{L}\left\{\left\|\bar{\mu}_{\tilde{s}}^l - \bar{\mu_{\text{s}}}^l\right\|_2 + \left\|\bar{\sigma}_{\tilde{s}}^l - \bar{\sigma}_{\text{s}}^l\right\|_2\right\} \tag{3}$$

where $l \in \{1, 2, ..., L\}$ denotes the layer l in the source-trained models, including the feature extractor F, classifier H, and depth estimator R. Guided by loss $\mathcal{L}_{\text{stat}}$, we could approximate the source-style domain that has the similar style as the source domain. Different from [95] that generates image contents from an input random noise, our NSC uses BN statistics alignment as one constraint to stylize the input images without changing contents.

Dual-level Semantic Consistency. To preserve the semantic contents during the image translation, we propose a *dual-level semantic consistency* on both feature level and image level to constrain the contents.

On the feature level, given the generated source-style image $x_{\tilde{s}}$ and the original target image x_t as inputs, a perceptual loss $\mathcal{L}_{\text{per}}$ is imposed onto the latent features of the ImageNet pre-trained VGG module ϕ, thus narrowing the perceptual differences between them:

$$\mathcal{L}_{\text{per}}^{\phi}(x_{\tilde{s}}, x_s) = \frac{1}{CHW}\left\|\phi(x_{\tilde{s}}) - \phi(x_t)\right\|_2^2 \tag{4}$$

However, merely using this perceptual loss in the spatial space is not powerful enough to ensure semantic consistency. This is mainly because the latent features are deeply-encoded, and some important semantic cues may be lost. Many previous works [36,64,60,19,22,94,93] suggest that the Fourier transformations from one domain to another only affect the amplitude, but not the phase of their

spectrum, where the phase component retains most of the contents in the original signals, while the amplitude component mainly contains styles. And inspired by [93], we consider explicitly penalizing the semantic inconsistency by ensuring the phase is retained before and after the image translation. For a single-channel image x, its Fourier transformation $\mathcal{F}(x)$ is formulated:

$$\mathcal{F}(x)(u,v) = \sum_{h=0}^{H-1}\sum_{w=0}^{W-1} x(h,w) e^{-j2\pi\left(\frac{h}{H}u + \frac{w}{W}v\right)} \tag{5}$$

As such, we enforce phase consistency between the original target image x_t and the source-style image $x_{\tilde{s}}$ by minimizing the following loss $\mathcal{L}_{ph}$:

$$\mathcal{L}_{ph}(x_{\tilde{s}}, x_t) = -\sum_j \frac{\langle \mathcal{F}(x_t)_j, \mathcal{F}(x_{\tilde{s}})_j \rangle}{\|\mathcal{F}(x_t)_j\|_2 \cdot \|\mathcal{F}(x_{\tilde{s}})_j\|_2} \tag{6}$$

where $\langle , \rangle$ is the dot-product, and $\| \cdot \|_2$ is the L_2 norm. Note that Eq. (6) is the negative cosine distance between the original phases and the generated phases. Therefore, by minimizing $\mathcal{L}_{ph}$, we can directly minimize their image-level differences on the Fourier spectrum and keep the phase consistency.

Intra-domain Spectrum Mixup. Given the unlabeled target data, we observe that the generator cannot perform well due to the lack of consideration of intra-domain domain shifts across different target subsets. If training only on the seen training subsets of the target domain and testing on the unseen testing subsets of the target domain, image qualities of the source-style domain could be less-desired. As such, we wish to a learn more robust generator G under varying environmental changes, *e.g.,* illumination, color.

Since previous findings [36,64,60,19,22,94,93,86] reveal that phase tends to preserve most contents in the Fourier spectrum of signals, while the amplitude mainly contains domain-specific styles, we propose to generate diversified images that retain contents but with new styles in the continuous frequency space. Through the FFT algorithm [59], we can efficiently compute the Fourier transformation $\mathcal{F}(x_t)$ and its inverse transformation $\mathcal{F}^{-1}(x_t)$ of the target image $x_t \in D_t$ via Eq. 5. The amplitude and phase components are formulated as:

$$\begin{aligned} \mathcal{A}(x_t)(u,v) &= \left[R^2(x_t)(u,v) + I^2(x_t)(u,v)\right]^{1/2} \\ \mathcal{P}(x_t)(u,v) &= \arctan\left[\frac{I(x_t)(u,v)}{R(x_t)(u,v)}\right], \end{aligned} \tag{7}$$

where $R(x_t)$ and $I(x_t)$ denote the real and imaginary part of $\mathcal{F}(x_t)$, respectively. For RGB images, the Fourier transformation for each channel is computed independently to get the corresponding amplitude and phase components.

Inspired from [99,86], we introduce *intra-domain spectrum mixup* (SpecMix) by linearly interpolating between the amplitude spectrums of two arbitrary images x_t^k, $x_t^{k'}$ from the same unlabeled target domain D_t:

$$\hat{\mathcal{A}}(x_t^k) = (1-\lambda)\mathcal{A}(x_t^k) + \lambda\mathcal{A}(x_t^{k'}), \tag{8}$$

where $\lambda \sim U(0, \eta)$, and the hyper-parameter η controls the strength of the augmentation. The mixed amplitude spectrum is then combined with the original phase spectrum to reconstruct a new Fourier representation:

$$\mathcal{F}(\hat{x}_t^k)(u,v) = \hat{\mathcal{A}}(x_t^k)(u,v) * e^{-j*\mathcal{P}(x_t^k)(u,v)}, \tag{9}$$

which is then fed to inverse Fourier transformation to generate the interpolated image: $\hat{x}_t^k = \mathcal{F}^{-1}[\mathcal{F}(\hat{x}_t^k)(u,v)]$.

This proposed *intra-domain spectrum mixup* is illustrated in Fig 2. By conducting the aforementioned steps, we could generate unseen target samples with new style and the original content in continuous frequency space. Thus, by feeding forward those diversified images to the generator G, the generalization abilities across different subsets within the target domain could be further enhanced.

3.3 Overall Objective and Optimization

Entropy loss. Minimizing the Shannon entropy of the label probability distribution has been proved to be effective in normal UDA task [58,74,83,65,75]. In this paper, we compute entropy loss via the classifier and depth estimator, respectively. The total entropy loss are penelized with $\mathcal{L}_{ent} = \mathcal{L}_{ent1} + \mathcal{L}_{ent2}$.

$$\begin{aligned} \mathcal{L}_{ent1} &= \sum_{c=1}^{C} -\langle p_c(x_{\tilde{s}}) \cdot \log(p_c(x_{\tilde{s}}))\rangle \\ \mathcal{L}_{ent2} &= \sum_{c=1}^{C}\sum_{h=1}^{H}\sum_{w=1}^{W} -\langle r_c(x_{\tilde{s}})(h,w) \cdot \log(r_c(x_{\tilde{s}}(h,w)))\rangle \end{aligned} \tag{10}$$

Total loss. During the adaptation procedure, the parameters of the source model F, H R, and the VGG module ϕ are fixed, and we only optimize the parameters of the generator G. The total loss L_{total} is the weighted sum of the aforementioned loss functions:

$$\mathcal{L}_{total} = \mathcal{L}_{stat} + \mathcal{L}_{per} + \lambda_{ent}\mathcal{L}_{ent} + \lambda_{ph}\mathcal{L}_{ph}, \tag{11}$$

where λ_{ent}, λ_{ph}, are the weighting coefficients for the loss $\mathcal{L}_{ent}$, $\mathcal{L}_{ph}$ respectively.

4 Experiments

In this section, we first describe the experimental setup in Section 4.1, including the benchmark datasets and the implementation details. Then, in Section 4.2, we demonstrate the effectiveness of our proposed method compared to the state-of-the-art approaches and related works on multi-source scenarios and single-source scenarios. Next, in Section 4.3, we conduct ablation studies to investigate the role of each component in the method. Finally, we provide more visualization and analysis in Section 4.4 to reveal the insights of the proposed method.

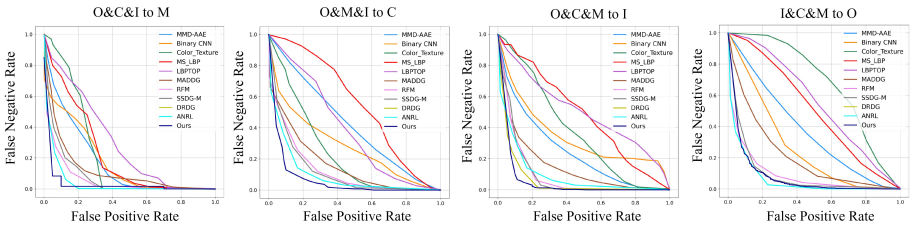


Fig. 3: ROC curves compared to the state-of-the-art FAS approaches

Table 1: Comparisons to the-state-of-art FAS methods on four testing domains

Methods	**O&C&I to M**		**O&M&I to C**		**O&C&M to I**		**I&C&M to O**	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
IDA [81]	66.6	27.8	55.1	39.0	28.3	78.2	54.2	44.6
LBPTOP [15]	36.9	70.8	42.6	61.5	49.5	49.5	53.1	44.0
MS_LBP [55]	29.7	78.5	54.2	44.9	50.3	51.6	50.2	49.3
ColorTexture [3]	28.0	78.4	30.5	76.8	40.4	62.7	63.5	32.7
Binary CNN [87]	29.2	82.8	34.8	71.9	34.4	65.8	29.6	77.5
Auxiliary (ALL) [51]	-	-	28.4	-	27.6	-	-	-
Auxiliary (Depth) [51]	22.7	85.8	33.5	73.1	29.1	71.6	30.1	77.6
MMD-AAE [42]	27.0	83.1	44.5	58.2	31.5	75.1	40.9	63.0
MADDG [68]	17.6	88.0	24.5	84.5	22.1	84.9	27.9	80.0
RFM [69]	13.8	93.9	20.2	88.1	17.3	90.4	16.4	91.1
SSDG-M [31]	16.7	90.5	23.1	85.5	18.2	94.6	25.2	81.8
DRDG [49]	12.4	95.8	19.1	88.8	15.6	91.8	15.6	91.8
ANRL [48]	10.8	96.8	17.9	89.3	16.0	91.0	15.7	91.9
SDA-FAS [78]	15.4	91.8	24.5	84.4	15.6	90.1	23.1	84.3
DIPE-FAS [15]	18.2	-	25.5	-	20.0	-	17.5	-
VLAD-VSA [79]	11.4	96.4	20.8	86.3	12.3	93.0	21.2	86.9
Ours	**9.2**	**98.0**	**12.2**	**93.0**	**10.0**	**96.0**	**14.4**	**92.6**

4.1 Experimental Setup

Datasets. We use four public datasets that are widely-used in FAS research to evaluate the effectiveness of our method: OULU-NPU [4] (denoted as O), CASIA-MFSD [104] (denoted as C), Idiap Replay-Attack [10] (denoted as I), and MSU-MFSD [81] (denoted as M). Strictly following the same protocols as previous UDA FAS methods [79,78,32,77,54], we use source domains to train the source model, the training set of the target domain for adaptation, and the testing set of the target domain for inference. Half Total Error Rate (HTER) and Area Under Curve (AUC) are used as the evaluation metrics [68].

Implementation Details. Our method is implemented via PyTorch on $24G$ NVIDIA 3090Ti GPU. We use the same backbone as existing works [49,48]. Note that we do not use any domain generalization techniques but just a binary classification loss and a depth loss during the first stage. We extract RGB channels of images, thus the input size is $256 \times 256 \times 3$. In the second stage, the coefficients λ_{ph} and λ_{ent} are set to 0.01 and 0.01 respectively. The generator G [113] is trained with the Adam optimizer with a learning rate of 1e-4.

Table 2: Comparisons (HTER) to unsupervised domain adaptation methods

Method	C $\longrightarrow$ I	C $\longrightarrow$ M	I $\longrightarrow$ C	I $\longrightarrow$ M	M $\longrightarrow$ C	M $\longrightarrow$ I	Average
ADDA [73]	41.8	36.6	49.8	35.1	39.0	35.2	39.6
DRCN [18]	44.4	27.6	48.9	42.0	28.9	36.8	38.1
Dup-GAN [28]	42.4	33.4	46.5	36.2	27.1	35.4	36.8
Auxliary [50]	27.6	–	28.4	–	–	–	–
De-spoof [34]	28.5	–	41.1	–	–	–	–
STASN [92]	31.5	–	30.9	–	–	–	–
Yang *et al.* [89]	49.2	18.1	39.6	36.7	49.6	49.6	40.5
KSA [40]	39.2	14.3	**26.3**	33.2	**10.1**	33.3	26.1
ADA [76]	17.5	9.3	41.6	30.5	17.7	5.1	20.3
DIPE-FAS [54]	18.9	–	30.1	–	–	–	–
DR-UDA [77]	15.6	9.0	34.2	29.0	16.8	3.0	17.9
USDAN-Un [32]	16.0	9.2	30.2	25.8	13.3	3.4	16.3
Ours	**15.1**	**5.8**	29.7	**20.8**	12.2	**2.5**	**14.4**

Table 3: Comparison to the source-free domain adaptation methods

Methods	**O&C&I to M**		**O&M&I to C**		**O&C&M to I**		**I&C&M to O**	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
AdaBN [38]	20.5	88.0	34.5	72.0	27.7	80.3	28.2	80.8
TENT [75]	20.1	88.0	35.0	71.2	27.2	79.6	28.3	80.7
SDAN [23]	17.7	90.0	25.9	81.3	28.2	84.2	32.9	75.0
SHOT [46]	15.0	87.6	20.1	84.3	40.2	57.8	25.3	78.2
G-SFDA [90]	37.5	67.8	38.9	67.2	32.6	73.6	40.4	63.7
NRC [91]	15.0	87.4	47.8	52.4	22.1	82.3	26.6	78.8
DIPE-FAS [54]	18.2	-	25.5	-	20.0	-	17.5	-
SDA-FAS [78]	15.4	91.8	24.5	84.4	16.4	92.0	23.1	84.3
Ours	**9.2**	**98.0**	**12.2**	**93.0**	**10.0**	**96.0**	**14.4**	**92.6**

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

To validate the generalization capability towards the target domain on the FAS task, we perform two experimental settings of UDA FAS, *i.e.,* multi-source domain adaptation and single-source domain adaptation, respectively.

Comparisons to FAS methods in multi-source adaptation. As shown in Table 1 and Fig. 3, our method outperforms all the state-of-the-art FAS methods under four challenging benchmarks, which demonstrates the effectiveness of our method. Conventional FAS approaches [81,15,55,3,87,51] do not consider learning the domain-invariant representations across domains and show less-desired performances. Besides, almost all DG FAS methods [42,68,31,69,48,49] lack a clear target direction for generalization, resulting in unsatisfactory performance in the target domain. Our method outperforms all the DG approaches by significant improvements in both HTER and AUC. A few DA approaches [79,78,15] conduct the experiments under this multi-source setting, while they all directly fit the model to the target domain with insufficient supervision and neglect the low-level features for adaptation, leading to undesirable outcomes. In contrast, our method is superior to them by a large margin in four challenging benchmarks.

Comparisons to FAS methods in single-source adaptation. To make a fair comparison to the normal UDA approaches in the FAS task, we also conduct

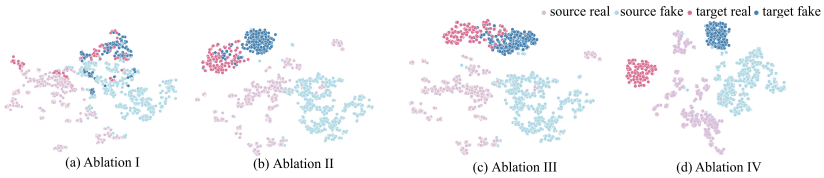


Fig. 4: The t-SNE visualization of features in different ablation studies

Table 4: Ablation of each component on four benchmarks

ID	Baseline	NSC	DSC	SpecMix	**O&C&I to M**		**O&M&I to C**		**O&C&M to I**		**I&C&M to O**	
					HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC
I	✓	-	-	-	29.2	77.8	32.7	76.4	19.4	85.8	27.1	80.9
II	✓	✓	-	-	20.0	89.0	28.7	79.7	17.3	88.4	21.1	84.9
III	✓	✓	✓	-	14.1	92.0	14.4	90.8	13.8	91.5	16.5	90.7
IV	✓	✓	✓	✓	**9.2**	**98.0**	**12.2**	**93.0**	**10.0**	**96.0**	**14.4**	**92.6**

experiments in single-source scenarios, where source models are pre-trained on the single-source domain. From Table 2, it is obvious to find that our proposed approach shows superiority under four of the six adaptation settings and achieves the best average HTER results. In some hard adaptation tasks, *e.g.,* I $\rightarrow$ C, and M $\rightarrow$ C, we can achieve the competitive results to the state-of-the-art methods. Interestingly, we find that some results in Table 2 are superior to results of Table 1. We guess the reason is that when training on multi-source domains, the style distribution is complicated, and it is hard to train a stable generator G, leading to inferior performances. Instead, training on only one source domain with simple style distribution is easier to obtain a better generator G.

Comparison to the related SFDA methods. Table 3 presents the comparison results with source-free domain adaptation (SFDA) approaches in four multi-source scenarios. As we can see, if we directly adapt the state-of-the-art SFDA approaches to the FAS task, the performances are less-desired. For example, some unsupervised/self-supervised techniques utilize pseudo labeling [54], neighborhood clustering [91,90], entropy minimization [46,75] and meta-learning [78] to reduce the domain gap between the source pre-trained model and the unlabeled target domain. The main reasons are two-fold. 1) Almost all of them do not fully utilize the source domain knowledge stored in the pre-trained model, which is not sufficient for feature alignment. 2) They largely neglect the intra-domain gap in the target domain itself, and do not consider learning a more robust domain-invariant representation under varying environmental changes within the target. In contrast, we address these two issues in an explicit manner, and show outstanding improvements on these challenging benchmarks.

4.3 Ablation Studies

In this section, we perform ablation experiments to investigate the effectiveness of each component, including inter-domain neural statistic consistency (NSC), intra-domain spectrum mixup (SpecMix), dual-level semantic consistency (DSC).

Table 5: Effect of hyper-parameter η of SpecMix

(a) Effect of SpecMix (η) on the training of our proposed model on O&C&I to M

η	0	0.1	0.2	0.3	0.4	0.5
HTER	14.1	9.2	10.0	10.0	10.0	10.0
AUC	92.0	98.0	97.9	97.8	97.3	97.2

(b) Effect of SpecMix (η) on the inference of a well-trained FAS model on Idiap (I)

η	0	0.1	0.2	0.3	0.4	0.5
HTER	0	0	0	0	0	0.4
AUC	100	100	100	100	100	99.9

Effectiveness of each component. Table 4 shows the ablation studies of each component in four different settings. The baseline means directly feeding forward the target image to the source-trained FAS model for prediction, and the results are with 77.8%, 76.4%, 85.8%, 80.9% AUC, respectively on O&C&I to M, O&M&I to C, O&C&M to I, and I&C&M to O, setting. By adding NSC, we boost the AUC performances to 89.0%, 79.7%, 88.4%, and 84.9%, respectively. Moreover, by adding DSC, we effectively achieve 92.0%, 90.8%, 91.5%, and 90.7%, respectively. Finally, our proposed SpecMix effectively increases the performance to 98.0%, 93.0%, 96.0%, 92.6% on four benchmarks, respectively. These improvements reveal the effectiveness and the complementarities of individual components of our proposed approach.

The t-SNE visualization of features. To understand how GDA framework aligns the feature representations, we utilize t-SNE to visualize the feature distributions of both the source and target datasets. As shown in Fig. 4 (a), we observe that the source data can be well discriminated by binary classification, however, the target data can not be well-classified between the real and fake faces without domain adaptation. From Fig. 4 (b), by adding NSC, the classification boundary becomes more clear but there are still some samples misclassified near the decision boundary, the main reason is that it lacks the constraints on the image contents during the generation. As such, by further adding DSC in Fig. 4 (c), the above issue is alleviated to some extent. By bridging both the inter-domain gap and intra-domain gap in Fig. 4 (d), our approach manages to learn a better decision boundary between these two categories, and meanwhile, our target features between different domains become more compact to align.

Discussions on SpecMix. Regarding that SpecMix could generate new styles in continuous frequency space, it it natural to ask several questions: *Will SpecMix change the category during the adaptation? Will mixing amplitude information affect the face liveness?* We conduct several experiments to answer the questions. Note that we use a hyper-parameter η to control the strength of augmentation in our SpecMix. Higher η leads to a larger upper bound ratio to mix another image's amplitude from the same batch with that of the current image. (1). In Table 5 (a), we investigate the effect of η on the training of our proposed GDA framework. During the adaptation, we find that if we set η=0, which means the intra-domain gaps are neglected, the performance is not perfect. If ranging η from 0.1 to 0.5, the performance changes are very slight compared to the best

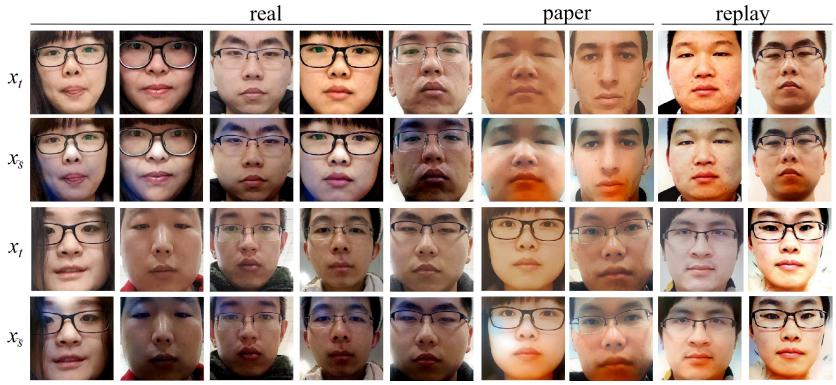


Fig. 5: Visualization of the target image x_t and the source-style image $x_{\tilde{s}}$

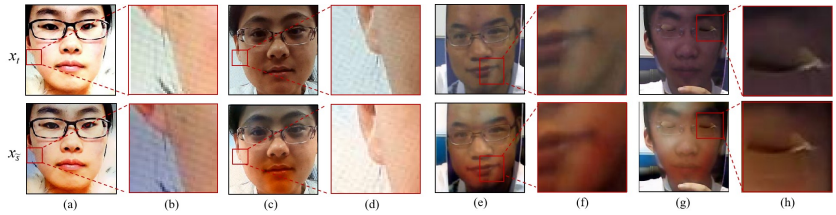


Fig. 6: Spoof details in the target images x_t and source-style images $x_{\tilde{s}}$

one when setting $\eta = 0.1$, but still achieves the state-of-the-art results. (2). As shown in Table 5 (b), we study the sensitivity of η during the inference of a well-trained FAS model on Idiap Replay-Attack dataset. If ranging η from 0.1 to 0.4, there are no performance changes, and when η=0.5, the effect is still slight. From these two aspects, we set $\eta = 0.1$ in all experiments, and under such cases, we argue that mixing the amplitude will not affect the face liveness.

4.4 Visualization and Analysis

Visualizations of generated images with source styles. To further explore whether the generator succeeds or not in stylizing the target images x_t to a generated images $x_{\tilde{s}}$ that preserves the target content with the source style, we visualize the adapted knowledge in the setting of I&C&M to O. As shown in Fig. 5, with the help of our proposed NSC, no matter what kind of faces they are, real faces or fake faces, the style differences between the source domain and the target domain are successfully captured by the generator, which illustrates the effectiveness of our proposed NSC. Besides, as shown in Fig. 6, the semantic consistency, especially the spoof details, *e.g.,* moire patterns, paper reflection, are well-maintained between the original target images and the pseudo source images, which demonstrates the effectiveness of our proposed DSC.

(a) Maximum Mean Discrepancy across domains (b) statistical discrepancy of mean in BN (c) statistical discrepancy of variance in BN

Fig. 7: Visualization of domain discrepancy (MMD and BN variance) of features

Visualizations of cross-domain discrepancy. As shown in Fig. 7 (a), we measure the maximum mean discrepancy (MMD) across domains. Source Only denotes directly forwarding the testing image to the source model without any domain adaptation techniques. Compared to Source Only, our model effectively reduce the MMD in both the shallow blocks and deep blocks, especially in the shallow blocks, which confirms that our framework successfully translate the target images to the source-style images. In Fig. 7 (b), we visualize the curve of mean discrepancy and variance discrepancy of each BN layer. As we can see, if directly feeding the target images to the source model in the test phase, the variation of the mean in BN between the source and the target (the red curve) is much larger than our method with domain adaptation methods (yellow curve). Our approach effectively prevents such feature misalignment. Besides, from Fig. 7 (c), we observe that the variation of variance in BN by feeding the source image to the source model (blue curve) is similar to that with our GDA (yellow curve), which means that our BN statistics effectively align with the source ones.

5 Conclusion

In this work, we reformulate UDA FAS as a domain stylization problem, aiming to fit the target data onto the well-trained models without changing the models. We propose Generative Domain Adaptation (GDA) framework with several carefully designed components. Firstly, we present an inter-domain neural statistic consistency (NSC) to guide the generator generating the source-style domain. Then, we introduce a dual-level semantic consistency (DSC) to prevent the generation from semantic distortions. Finally, we design an intra-domain spectrum mixup (SpecMix) to reduce the intra-domain gaps. Extensive experiments with analysis demonstrate the effectiveness of our proposed approach.

Acknowledgment: This work is supported by National Key Research and Development Program of China (2019YFC1521104), National Natural Science Foundation of China (72192821, 61972157), Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102), Shanghai Science and Technology Commission (21511101200, 22YF1420300), and Art major project of National Social Science Fund (I8ZD22).

References

1. Bao, W., Li, H., Li, N., Jiang, W.: A liveness detection method for face recognition based on optical flow field. In: International Conference on Image Analysis and Signal Processing. pp. 233–236. IEEE (2009)
2. Boulkenafet, Z., Komulainen, J., Hadid, A.: Face anti-spoofing based on color texture analysis. In: IEEE International Conference on Image Processing (ICIP). pp. 2636–2640. IEEE (2015)
3. Boulkenafet, Z., Komulainen, J., Hadid, A.: Face spoofing detection using colour texture analysis. IEEE Transactions on Information Forensics and Security (TIFS) **11**(8), 1818–1830 (2016)
4. Boulkenafet, Z., Komulainen, J., Li, L., Feng, X., Hadid, A.: Oulu-npu: A mobile face presentation attack database with real-world variations. In: 12th IEEE international conference on automatic face & gesture recognition (FG 2017). pp. 612–618. IEEE (2017)
5. Chang, W.L., Wang, H.P., Peng, W.H., Chiu, W.C.: All about structure: Adapting structural information across domains for boosting semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1900–1909 (2019)
6. Chen, S., Yao, T., Zhang, K.Y., Chen, Y., Sun, K., Ding, S., Li, J., Huang, F., Ji, R.: A dual-stream framework for 3d mask face presentation attack detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 834–841 (2021)
7. Chen, Z., Li, B., Wu, S., Xu, J., Ding, S., Zhang, W.: Shape matters: Deformable patch attack. In: European conference on computer vision (ECCV) (2022)
8. Chen, Z., Li, B., Xu, J., Wu, S., Ding, S., Zhang, W.: Towards practical certifiable patch defense with vision transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15148–15158 (2022)
9. Chen, Z., Yao, T., Sheng, K., Ding, S., Tai, Y., Li, J., Huang, F., Jin, X.: Generalizable representation learning for mixture domain face anti-spoofing. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). vol. 35, pp. 1132–1139 (2021)
10. Chingovska, I., Anjos, A., Marcel, S.: On the effectiveness of local binary patterns in face anti-spoofing. In: International Conference of Biometrics Special Interest Group. pp. 1–7. IEEE (2012)
11. Choi, J., Kim, T., Kim, C.: Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6830–6840 (2019)
12. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). pp. 4690–4699 (2019)
13. Feng, L., Po, L.M., Li, Y., Xu, X., Yuan, F., Cheung, T.C.H., Cheung, K.W.: Integration of image quality and motion cues for face anti-spoofing: A neural network approach. Journal of Visual Communication and Image Representation (JVCIR) (2016)
14. Feng, Z., Zhou, Q., Gu, Q., Tan, X., Cheng, G., Lu, X., Shi, J., Ma, L.: Dmt: Dynamic mutual training for semi-supervised learning. Pattern Recognition (PR) p. 108777 (2022)

15. Freitas Pereira, T., Komulainen, J., Anjos, A., De Martino, J., Hadid, A., Pietikäinen, M., Marcel, S.: Face liveness detection using dynamic texture. Eurasip Journal on Image and Video Processing **2014**(1), 1–15 (2014)
16. Freitas Pereira, T.d., Anjos, A., Martino, J.M.D., Marcel, S.: Lbp- top based countermeasure against face spoofing attacks. In: Asian Conference on Computer Vision (ACCV). pp. 121–132. Springer (2012)
17. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International Conference on Machine Learning (ICML). pp. 1180–1189. PMLR (2015)
18. Ghifary, M., Kleijn, W.B., Zhang, M., Balduzzi, D., Li, W.: Deep reconstruction-classification networks for unsupervised domain adaptation. In: European conference on computer vision (ECCV). pp. 597–613. Springer (2016)
19. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS) (2005)
20. Gu, Q., Zhou, Q., Xu, M., Feng, Z., Cheng, G., Lu, X., Shi, J., Ma, L.: Pit: Position-invariant transform for cross-fov domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 8761–8770 (2021)
21. Guo, S., Zhou, Q., Zhou, Y., Gu, Q., Tang, J., Feng, Z., Ma, L.: Label-free regional consistency for image-to-image translation. In: IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE (2021)
22. Hansen, B.C., Hess, R.F.: Structural sparseness and spatial phase alignment in natural scenes. JOSA A **24**(7), 1873–1885 (2007)
23. He, Y., Carass, A., Zuo, L., Dewey, B.E., Prince, J.L.: Self domain adapted network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (2020)
24. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: International conference on machine learning (ICML). pp. 1989–1998. PMLR (2018)
25. Hou, Y., Zheng, L.: Source free domain adaptation with image translation. arXiv preprint arXiv:2008.07514 (2020)
26. Hou, Y., Zheng, L.: Visualizing adapted knowledge in domain transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13824–13833 (2021)
27. Hu, C., Zhang, K.Y., Yao, T., Ding, S., Li, J., Huang, F., Ma, L.: An end-to-end efficient framework for remote physiological signal sensing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 2378–2384 (2021)
28. Hu, L., Kan, M., Shan, S., Chen, X.: Duplex generative adversarial network for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1498–1507 (2018)
29. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning (ICML). pp. 448–456 (2015)
30. Isobe, T., Jia, X., Chen, S., He, J., Shi, Y., Liu, J., Lu, H., Wang, S.: Multi-target domain adaptation with collaborative consistency learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8187–8196 (2021)

31. Jia, Y., Zhang, J., Shan, S., Chen, X.: Single-side domain generalization for face anti-spoofing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
32. Jia, Y., Zhang, J., Shan, S., Chen, X.: Unified unsupervised and semi-supervised domain adaptation network for cross-scenario face anti-spoofing. Pattern Recognition (PR) **115**, 107888 (2021)
33. Jiang, Z., Li, Y., Yang, C., Gao, P., Wang, Y., Tai, Y., Wang, C.: Prototypical contrast adaptation for domain adaptive segmentation. In: European Conference on Computer Vision (ECCV) (2022)
34. Jourabloo, A., Liu, Y., Liu, X.: Face de-spoofing: Anti-spoofing via noise modeling. In: Proceedings of the European conference on computer vision (ECCV). pp. 290–306 (2018)
35. Kemelmacher-Shlizerman, I., Seitz, S.M., Miller, D., Brossard, E.: The megaface benchmark: 1 million faces for recognition at scale. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp. 4873–4882 (2016)
36. Kermisch, D.: Image reconstruction from phase information only. JOSA **60**(1), 15–17 (1970)
37. Komulainen, J., Hadid, A., Pietikäinen, M.: Context based face anti-spoofing. In: 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS). pp. 1–8. IEEE
38. Li, Yanghao, Wang, Naiyan, Shi, Jianping, Hou, Xiaodi, Liu, Jiaying: Adaptive batch normalization for oractical domain adaptation. Pattern Recognition (PR) (2018)
39. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.: Learning to generalize: Meta-learning for domain generalization. In: Proceedings of the AAAI conference on artificial intelligence (AAAI). vol. 32 (2018)
40. Li, H., Li, W., Cao, H., Wang, S., Huang, F., Kot, A.C.: Unsupervised domain adaptation for face anti-spoofing. IEEE Transactions on Information Forensics and Security (TIFS) **13**(7), 1794–1809 (2018)
41. Li, H., Pan, S.J., Wang, S., Kot, A.C.: Domain generalization with adversarial feature learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp. 5400–5409 (2018)
42. Li, H., Pan, S.J., Wang, S., Kot, A.C.: Domain generalization with adversarial feature learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5400–5409 (2018)
43. Li, J., Wang, Y., Tan, T., Jain, A.K.: Live face detection based on the analysis of fourier spectra. In: Biometric technology for human identification. vol. 5404, pp. 296–303. SPIE (2004)
44. Li, L., Feng, X., Boulkenafet, Z., Xia, Z., Li, M., Hadid, A.: An original face anti-spoofing approach using partial convolutional neural network. In: International Conference on Image Processing Theory, Tools and Applications (IPTA) (2016)
45. Li, S., Xu, J., Xu, X., Shen, P., Li, S., Hooi, B.: Spherical confidence learning for face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15629–15637 (2021)
46. Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: International Conference on Machine Learning (ICML). pp. 6028–6039. PMLR (2020)
47. Lin, B., Li, X., Yu, Z., Zhao, G.: Face liveness detection by rppg features and contextual patch-based cnn. In: International Conference on Biometric Engineering and Applications (ICBEA) (2019)

48. Liu, S., Zhang, K.Y., Yao, T., Bi, M., Ding, S., Li, J., Huang, F., Ma, L.: Adaptive normalized representation learning for generalizable face anti-spoofing pp. 1469–1477 (2021)
49. Liu, S., Zhang, K.Y., Yao, T., Sheng, K., Ding, S., Tai, Y., Li, J., Xie, Y., Ma, L.: Dual reweighting domain generalization for face presentation attack detection. International Joint Conference on Artificial Intelligence (IJCAI) (2021)
50. Liu, S., Lan, X., Yuen, P.C.: Remote photoplethysmography correspondence feature for 3d mask face presentation attack detection. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
51. Liu, Y., Jourabloo, A., Liu, X.: Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). pp. 389–398 (2018)
52. Liu, Y., Stehouwer, J., Liu, X.: On disentangling spoof trace for generic face anti-spoofing. In: European Conference on Computer Vision (ECCV). pp. 406–422. Springer (2020)
53. Liu, Y., Zhang, W., Wang, J.: Source-free domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1215–1224 (2021)
54. Lv, L., Xiang, Y., Li, X., Huang, H., Ruan, R., Xu, X., Fu, Y.: Combining dynamic image and prediction ensemble for cross-domain face anti-spoofing. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 2550–2554 (2021)
55. Maatta, J., Hadid, A., Pietikainen, M.: Face spoofing detection from single images using micro-texture analysis. In: Proceedings of the IEEE International Joint Conference on Biometrics (IJCB) (2011)
56. Meng, R., Chen, W., Yang, S., Song, J., Lin, L., Xie, D., Pu, S., Wang, X., Song, M., Zhuang, Y.: Slimmable domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7141–7150 (2022)
57. Meng, R., Li, X., Chen, W., Yang, S., Song, J., Wang, X., Zhang, L., Song, M., Xie, D., Pu, S.: Attention diversification for domain generalization. In: European Conference on Computer Vision (ECCV) (2022)
58. Morerio, P., Cavazza, J., Murino, V.: Minimal-entropy correlation alignment for unsupervised deep domain adaptation. arXiv preprint arXiv:1711.10288 (2017)
59. Nussbaumer, H.J.: The fast fourier transform. In: Fast Fourier Transform and Convolution Algorithms, pp. 80–111. Springer (1981)
60. Oppenheim, A.V., Lim, J.S.: The importance of phase in signals. Proceedings of the IEEE **69**(5), 529–541 (1981)
61. Patel, K., Han, H., Jain, A.K.: Cross-database face antispoofing with robust feature representation. In: Chinese Conference on Biometric Recognition. pp. 611–619. Springer (2016)
62. Patel, K., Han, H., Jain, A.K.: Secure face unlock: Spoof detection on smartphones. IEEE Transactions on Information Forensics and Security (TIFS) **11**(10), 2268–2283 (2016)
63. Pei, Z., Cao, Z., Long, M., Wang, J.: Multi-adversarial domain adaptation. In: Thirty-second AAAI conference on artificial intelligence (AAAI) (2018)
64. Piotrowski, L.N., Campbell, F.W.: A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. Perception **11**(3), 337–346 (1982)

65. Prabhu, V., Khare, S., Kartik, D., Hoffman, J.: Sentry: Selective entropy optimization via committee consistency for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 8558–8567 (2021)
66. Quan, R., Wu, Y., Yu, X., Yang, Y.: Progressive transfer learning for face anti-spoofing. IEEE Transactions on Image Processing (TIP) **30**, 3946–3955 (2021)
67. Santurkar, S., Tsipras, D., Ilyas, A., Madry, A.: How does batch normalization help optimization? Advances in neural information processing systems **31** (2018)
68. Shao, R., Lan, X., Li, J., Yuen, P.C.: Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
69. Shao, R., Lan, X., Yuen, P.C.: Regularized fine-grained meta face anti-spoofing. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) (2020)
70. Siddiqui, T.A., Bharadwaj, S., Dhamecha, T.I., Agarwal, A., Vatsa, M., Singh, R., Ratha, N.: Face anti-spoofing with multifeature videolet aggregation. In: 2016 23rd International Conference on Pattern Recognition (ICPR). pp. 1035–1040. IEEE (2016)
71. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: Closing the gap to human-level performance in face verification. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp. 1701–1708 (2014)
72. Tu, X., Zhang, H., Xie, M., Luo, Y., Zhang, Y., Ma, Z.: Deep transfer across domains for face antispoofing. Journal of Electronic Imaging **28**(4), 043001 (2019)
73. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). pp. 7167–7176 (2017)
74. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2517–2526 (2019)
75. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Fully test-time adaptation by entropy minimization. In: International Conference on Learning Representations (ICLR) (2021)
76. Wang, G., Han, H., Shan, S., Chen, X.: Improving cross-database face presentation attack detection via adversarial domain adaptation. In: Proceedings of the IEEE International Conference on Biometrics (ICB) (2019)
77. Wang, G., Han, H., Shan, S., Chen, X.: Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection. IEEE Transactions on Information Forensics and Security (TIFS) **16**, 56–69 (2021)
78. Wang, J., Zhang, J., Bian, Y., Cai, Y., Wang, C., Pu, S.: Self-domain adaptation for face anti-spoofing. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI). vol. 35, pp. 2746–2754 (2021)
79. Wang, J., Zhao, Z., Jin, W., Duan, X., Lei, Z., Huai, B., Wu, Y., He, X.: Vladvsa: Cross-domain face presentation attack detection with vocabulary separation and adaptation. In: Proceedings of the 29th ACM International Conference on Multimedia (ACM MM). pp. 1497–1506 (2021)
80. Wang, J., Liu, Y., Hu, Y., Shi, H., Mei, T.: Facex-zoo: A pytorch toolbox for face recognition. In: Proceedings of the 29th ACM International Conference on Multimedia (ACM MM). pp. 3779–3782 (2021)
81. Wen, D., Han, H., Jain, A.K.: Face spoof detection with image distortion analysis. IEEE Transactions on Information Forensics and Securityn (TIFS) **10**(4), 746–761 (2015)

82. Wu, A., Han, Y., Zhu, L., Yang, Y.: Instance-invariant domain adaptive object detection via progressive disentanglement. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) pp. 1–1 (2021). https://doi.org/10.1109/TPAMI.2021.3060446
83. Wu, X., Zhang, S., Zhou, Q., Yang, Z., Zhao, C., Latecki, L.J.: Entropy minimization versus diversity maximization for domain adaptation. IEEE Transactions on Neural Networks and Learning Systems (TNNLS) (2021)
84. Xu, H., Liu, F., Zhou, Q., Hao, J., Cao, Z., Feng, Z., Ma, L.: Semi-supervised 3d object detection via adaptive pseudo-labeling. In: 2021 IEEE International Conference on Image Processing (ICIP). pp. 3183–3187. IEEE (2021)
85. Xu, M., Wang, H., Ni, B., Tian, Q., Zhang, W.: Cross-domain detection via graph-induced prototype alignment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12355–12364 (2020)
86. Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q.: A fourier-based framework for domain generalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14383–14392 (2021)
87. Yang, J., Lei, Z., Li, S.Z.: Learn convolutional neural network for face anti-spoofing. In: arXiv preprint arXiv:1408.5601 (2014)
88. Yang, J., Lei, Z., Liao, S., Li, S.Z.: Face liveness detection with component dependent descriptor. In: 2013 International Conference on Biometrics (ICB). pp. 1–6. IEEE (2013)
89. Yang, J., Lei, Z., Yi, D., Li, S.Z.: Person-specific face antispoofing with subject domain adaptation. IEEE Transactions on Information Forensics and Security (TIFS) **10**(4), 797–809 (2015)
90. Yang, S., Wang, Y., van de Weijer, J., Herranz, L., Jui, S.: Generalized source-free domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 8978–8987 (2021)
91. Yang, S., van de Weijer, J., Herranz, L., Jui, S., et al.: Exploiting the intrinsic neighborhood structure for source-free domain adaptation. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 34, pp. 29393–29405 (2021)
92. Yang, X., Luo, W., Bao, L., Gao, Y., Gong, D., Zheng, S., Li, Z., Liu, W.: Face anti-spoofing: Model matters, so does data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3507–3516 (2019)
93. Yang, Y., Lao, D., Sundaramoorthi, G., Soatto, S.: Phase consistent ecological domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9011–9020 (2020)
94. Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4085–4095 (2020)
95. Yin, H., Molchanov, P., Alvarez, J.M., Li, Z., Mallya, A., Hoiem, D., Jha, N.K., Kautz, J.: Dreaming to distill: Data-free knowledge transfer via deepinversion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8715–8724 (2020)
96. Yu, Z., Li, X., Niu, X., Shi, J., Zhao, G.: Face anti-spoofing with human material perception. In: European Conference on Computer Vision (ECCV). pp. 557–575. Springer (2020)
97. Yu, Z., Li, X., Shi, J., Xia, Z., Zhao, G.: Revisiting pixel-wise supervision for face anti-spoofing. IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM) **3**(3), 285–295 (2021)

98. Yu, Z., Zhao, C., Wang, Z., Qin, Y., Su, Z., Li, X., Zhou, F., Zhao, G.: Searching central difference convolutional networks for face anti-spoofing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5295–5305 (2020)
99. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
100. Zhang, J., Tai, Y., Yao, T., Meng, J., Ding, S., Wang, C., Li, J., Huang, F., Ji, R.: Aurora guard: Reliable face anti-spoofing via mobile lighting system. arXiv preprint arXiv:2102.00713 (2021)
101. Zhang, K.Y., Yao, T., Zhang, J., Liu, S., Yin, B., Ding, S., Li, J.: Structure destruction and content combination for face anti-spoofing. In: 2021 IEEE International Joint Conference on Biometrics (IJCB). pp. 1–6. IEEE (2021)
102. Zhang, K.Y., Yao, T., Zhang, J., Tai, Y., Ding, S., Li, J., Huang, F., Song, H., Ma, L.: Face anti-spoofing via disentangled representation learning. In: European Conference on Computer Vision (ECCV). pp. 641–657. Springer (2020)
103. Zhang, P., Zhang, B., Zhang, T., Chen, D., Wang, Y., Wen, F.: Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). pp. 12414–12424 (2021)
104. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: A face antispoofing database with diverse attacks. In: 2012 5th IAPR international conference on Biometrics (ICB). pp. 26–31. IEEE (2012)
105. Zhao, Y., Zhong, Z., Luo, Z., Lee, G.H., Sebe, N.: Source-free open compound domain adaptation in semantic segmentation. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) (2022)
106. Zhao, Y., Zhong, Z., Yang, F., Luo, Z., Lin, Y., Li, S., Sebe, N.: Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6277–6286 (2021)
107. Zhou, F., Gao, C., Chen, F., Li, C., Li, X., Yang, F., Zhao, Y.: Face anti-spoofing based on multi-layer domain adaptation. In: 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). pp. 192–197. IEEE (2019)
108. Zhou, Q., Feng, Z., Gu, Q., Cheng, G., Lu, X., Shi, J., Ma, L.: Uncertainty-aware consistency regularization for cross-domain semantic segmentation. Computer Vision and Image Understanding (CVIU) p. 103448 (2022)
109. Zhou, Q., Feng, Z., Gu, Q., Pang, J., Cheng, G., Lu, X., Shi, J., Ma, L.: Context-aware mixup for domain adaptive semantic segmentation. arXiv preprint arXiv:2108.03557 (2021)
110. Zhou, Q., Gu, Q., Pang, J., Feng, Z., Cheng, G., Lu, X., Shi, J., Ma, L.: Self-adversarial disentangling for specific domain adaptation. arXiv preprint arXiv:2108.03553 (2021)
111. Zhou, Q., Zhang, K.Y., Yao, T., Yi, R., Ding, S., Ma, L.: Adaptive mixture of experts learning for generalizable face anti-spoofing. In: Proceedings of the 30th ACM International Conference on Multimedia (ACM MM) (2022)
112. Zhou, Q., Zhuang, C., Lu, X., Ma, L.: Domain adaptive semantic segmentation with regional contrastive consistency regularization. In: 2022 IEEE International Conference on Multimedia and Expo (ICME). IEEE (2022)
113. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision (ICCV). pp. 2223–2232 (2017)

114. Zhu, W., Wang, C.Y., Tseng, K.L., Lai, S.H., Wang, B.: Local-adaptive face recognition via graph-based meta-clustering and regularized adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20301–20310 (2022)
115. Zou, Y., Yu, Z., Kumar, B., Wang, J.: Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In: Proceedings of the European conference on computer vision (ECCV). pp. 289–305 (2018)
116. Zou, Y., Yu, Z., Liu, X., Kumar, B., Wang, J.: Confidence regularized self-training. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5982–5991 (2019)