Source-Free Domain Adaptation with Contrastive Domain Alignment and Self-supervised Exploration for Face Anti-spoofing

Yuchen Liu[1], Yabo Chen[2], Wenrui Dai[2(✉)], Mengran Gou[3], Chun-Ting Huang[3], and Hongkai Xiong[1]

[1] Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
{liuyuchen6666,xionghongkai}@sjtu.edu.cn
[2] Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{chenyabo,daiwenrui}@sjtu.edu.cn
[3] Qualcomm AI Research, Shanghai, China
{mgou,chunting}@qti.qualcomm.com

Abstract. Despite promising success in intra-dataset tests, existing face anti-spoofing (FAS) methods suffer from poor generalization ability under domain shift. This problem can be solved by aligning source and target data. However, due to privacy and security concerns of human faces, source data are usually inaccessible during adaptation for practical deployment, where only a pre-trained source model and unlabeled target data are available. In this paper, we propose a novel Source-free Domain Adaptation framework for Face Anti-Spoofing, namely SDA-FAS, that addresses the problems of source knowledge adaptation and target data exploration under the source-free setting. For source knowledge adaptation, we present novel strategies to realize self-training and domain alignment. We develop a contrastive domain alignment module to align conditional distribution across different domains by aggregating the features of fake and real faces separately. We demonstrate in theory that the pre-trained source model is equivalent to the source data as source prototypes for supervised contrastive learning in domain alignment. The source-oriented regularization is also introduced into self-training to alleviate the self-biasing problem. For target data exploration, self-supervised learning is employed with specified patch shuffle data augmentation to explore intrinsic spoofing features for unseen attack types. To our best knowledge, SDA-FAS is the first attempt that jointly

Y. Liu, Y. Chen—Equal contribution. Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc. Datasets were downloaded and evaluated by Shanghai Jiao Tong University researchers.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-19775-8_30.

S. Avidan et al. (Eds.): ECCV 2022, LNCS 13672, pp. 511–528, 2022.
https://doi.org/10.1007/978-3-031-19775-8_30

optimizes the source-adapted knowledge and target self-supervised exploration for FAS. Extensive experiments on thirteen cross-dataset testing scenarios show that the proposed framework outperforms the state-of-the-art methods by a large margin.

Keywords: Face anti-spoofing · Source-free domain adaptation

1 Introduction

Face recognition (FR) systems are widely employed for human-computer interaction in our daily life. Face anti-spoofing (FAS) is crucial to protect FR systems from presentation attacks, e.g., print attack, video attack and 3D mask attack. Traditional FAS methods extract texture patterns with hand-crafted descriptors [8,15,24].

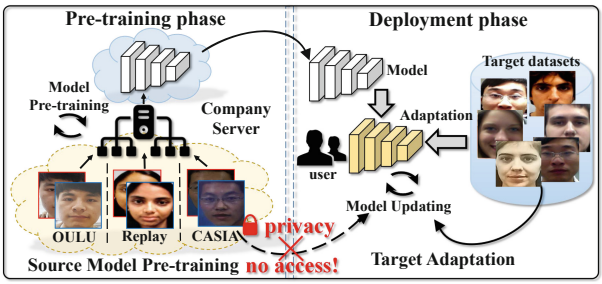
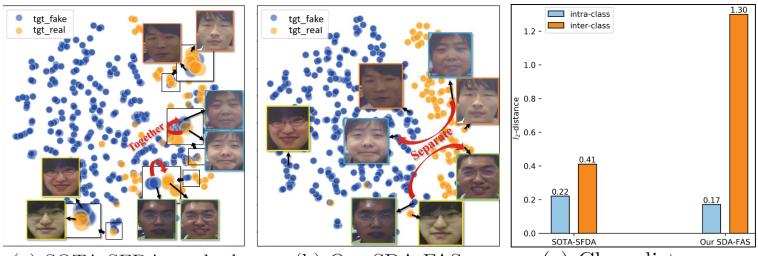


Fig. 1. A practical application scenario for face anti-spoofing. In the pre-training phase, the company builds a model based on the collected face data. When deployed on the user side, few collected unlabeled data can improve the performance through adaptation, but has distribution discrepancies with source knowledge. Moreover, due to privacy and security concerns of face data, users have no access to any source data of the company but the trained model.

With the rise of deep learning, convolutional neural networks (CNNs) have been adopted to extract deep semantic features [40,45,46]. Despite promising success in intra-dataset tests, these methods are dramatically degraded in cross-dataset tests where training data are from the source domain and test data are from the target domain with different distributions. The distribution discrepancies in illumination, background and resolution undermine the performance and an adaptation process is required to mitigate domain shift.

Domain adaptation (DA) based methods leverage maximum mean discrepancy (MMD) loss [16,32] and adversarial training [12,35,36] to align the source and target domains, which need to access source data. Unfortunately, they might be infeasible for sensitive facial images due to the restriction by institutional policies, legal issues and privacy concerns. For example, according to the General Data Protection Regulation (GDPR) [30], institutions in the European Union



(a) SOTA-SFDA method (b) Our SDA-FAS (c) Class distance

Fig. 2. The t-SNE visualization of extracted features and corresponding faces under O & M & I→C. Same border color for faces with the same identity. **(a)** SOTA-SFDA method SHOT [17] achieves marginal distribution alignment, which is prone to map the features of real and fake faces together. **(b)** Our method with conditional distribution alignment separates them well and increases the discrimination ability. **(c)** Intra-class and inter-class distance of extracted features for SHOT and our SDA-FAS.

are regulated to protect the privacy of their data. Figure 1 illustrates a practical application scenario of *source-free domain adaptation* for FAS. A model is first pre-trained based on the (large-scale) source data and is released for deployment. In the deployment phase, the source data cannot be shared for adapting the pre-trained model to the target data, as they contain sensitive biometrics information. Besides, face images acquired under different illumination, background, resolution or using cameras with different parameters will lead to distribution discrepancies between source and target data. These distribution discrepancies have to be overcome using only the pre-trained source model and unlabeled target data. Domain generalization (DG) methods [11,26,27] learn a robust source model without exploiting the target data and achieve limited performance in practice. Consequently, Source-Free Domain Adaptation (SFDA) for face anti-spoofing is an important yet challenging problem remained to be solved.

Recently, SFDA has been considered to tackle a similar issue on image classification [1,17,41,42]. In image classification, label consistency among data with high local affinity is encouraged [41,42] or marginal distribution of source and target domains is implicitly aligned [1,17] to harmonize the clustered features in the feature space. Different from image classification, in FAS, fake faces of the same identity have similar facial features, whereas real faces of different identities differ. The intra-class distance between real faces of different identities probably exceeds the inter-class distance between real and fake faces of the same identity [11,27]. Clusters of features do not exist in FAS and SFDA models for image classification inevitably lead to degraded performance. Table 1 and Fig. 2 provide empirical results as supporting evidence, where SHOT [17], the state-of-the-art SFDA method, tends to cluster the features of real and fake faces together and obscures the discrimination ability. These problems urge a SFDA method designed specifically for FAS to achieve promising performance.

Lv et al. [20] accommodate to source-free setting for FAS by directly applying self-training but lack specific design for sufficiently exploring FAS tasks. The performance gain by adaptation is trivial (i.e., 1.9% HTER reduction on average),

as shown in Table 1. To summarize, challenges to *source-free domain adaptation* for FAS include source knowledge adaptation and target data exploration.

- **Source knowledge adaptation.** Existing self-training and marginal domain alignment cannot adapt source knowledge well in FAS, especially when source data are unavailable. The target pseudo labels generated by the source model are noisy, especially under domain shift, leading to the accumulated error of self-training. Marginal distribution alignment is prone to cluster the features of real and fake faces and greatly degrades the discrimination ability for FAS.
- **Target data exploration.** Unseen attack types in the target data lead to enormous domain discrepancies where source knowledge is inapplicable and biased. It is indispensable to explore target data by itself to boost generalization ability. However, target data exploration is ignored in existing methods.

To address these issues, we propose a novel Source-free Domain Adaptation for Face Anti-Spoofing, namely SDA-FAS. Regarding source knowledge adaptation, we design novel strategies for self-training and domain alignment. We develop a contrastive domain alignment module for mitigating feature distribution discrepancies under a source-free setting. The pre-trained classifier weight is employed as the source prototypes with a theoretical guarantee of equivalence in training. We also introduce the source-oriented regularization into self-training to alleviate a self-biasing problem. For target data exploration, self-supervised learning is implemented with specified patch shuffle data augmentation to mine the intrinsic spoofing features of the target data, which also mitigates the reliance on pseudo-labels and boosts the tolerance to interfering knowledge transferred from the source domain. Contributions of this paper are summarized as below:

- We propose a novel contrastive domain alignment module to align the features of target data with the source prototypes of the same category for mitigating distribution discrepancies with theoretical support.
- We implement self-supervised learning with specified patch shuffle data augmentation to explore the target data for robust features in the case where unseen attack types emerge and source knowledge is unreliable.
- Our method is evaluated extensively on thirteen cross-dataset testing benchmarks and outperforms the state-of-the-art methods by a large margin.

To our best knowledge, SDA-FAS is the first attempt that unifies the transfer of pre-trained source knowledge and the self-exploration of unlabeled target data for FAS under a practical yet challenging source-free setting.

2 Related Work

Face Anti-spoofing. Existing face anti-spoofing (FAS) methods can be classified into three categories, i.e., handcrafted, deep learning, and DG/DA methods. Handcrafted methods extract the frame-level features using handcrafted descriptors such as LBP [8], HOG [15] and SIFT [24]. Deep learning methods boost the discrimination ability of extracted features. Yang et al. [40] first introduce CNNs into FAS, and Xu et al. [39] design a CNN-LSTM architecture to extract temporal

features. Intrinsic spoofing patterns are further explored with pixel-wise supervision [44], e.g., depth maps [18], reflection maps [43] and binary masks [19]. These methods achieve remarkable performance in intra-dataset tests but degrade significantly in cross-dataset tests due to distribution discrepancies.

DG and DA have been leveraged to mitigate domain shift in cross-dataset tests. DG methods focus on extracting domain invariant features without target data. MADDG [27] learns a shared feature space with multi-adversarial learning. SSDG [11] develops a single-side DG framework by only aggregating real faces from different source domains. DA methods achieve the domain alignment using source data and unlabeled target data. Maximum mean discrepancy (MMD) loss [16,32] and adversarial training [12,35,36] are leveraged to align the feature space between the source and target domains. Quan et al. [25] present a transfer learning framework to progressively make use of unlabeled target data with reliable pseudo labels for training. However, these methods fail to work or suffer from poor performance in a practical yet challenging source-free setting, which considers the privacy and security issues of sensitive face images.

Source-Free Domain Adaptation. Domain adaptation aims at transferring knowledge from source domain to target domain. Recently, source-free domain adaptation (SFDA) has been considered to address privacy issues. PrDA [14] progressively updates the model in a self-learning manner with filtered pseudo labels. Based on the source hypothesis, SHOT [17] aligns the marginal distribution of source and target domains via information maximization. DECISION [1] further extends SHOT to a multi-source setting. TENT [34] adapts batch normalization's affine parameters with an entropy penalty. NRC [41] exploits the intrinsic cluster structure to encourage label consistency among data with high local affinity. However, existing works cannot be easily employed in FAS due to the different nature of tasks. Recently, Lv et al. [20] realize SFDA for FAS by directly using the pseudo labels for self-training, but suffer from trivial performance gain after adaptation due to the accumulated training error brought by noisy pseudo labels, especially under domain shift.

Contrastive Learning. Contrastive learning is popular for self-supervised representation learning. To obtain the best feature representations, the InfoNCE loss [22] is introduced to pull together an anchor and one positive sample (constructed by augmenting the anchor), and push apart the anchor from many negative samples. Besides, self-supervised features can be learned by only matching the similarity between the anchor and the positive sample [3,5]. Contrastive learning is also introduced into image classification in a supervised manner [13], where categorical labels are used to build positive and negative samples.

3 Proposed Method

3.1 Overview

We consider the practical source-free domain adaptation setting for face anti-spoofing, in which only a trained source model and unlabeled target domain

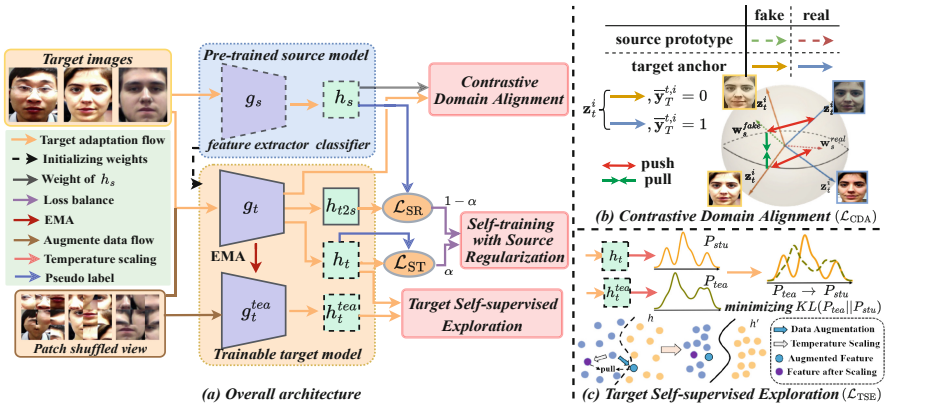


Fig. 3. **(a)** The overall architecture contains a pre-trained source model (in blue) and a trainable target model with three modules (in orange). For self-training with source regularization, pseudo labels $\overline{\mathbf{y}}_T^t$ and $\overline{\mathbf{y}}_T^s$ generated by target and source model supervise the outputs $\tilde{\mathbf{y}}_t$ and $\tilde{\mathbf{y}}_{t2s}$, respectively. **(b)** Contrastive domain alignment. The features of target data are pulled with the source prototypes of the same category (in green arrow) and pushed away from different categories (in red arrow) for conditional domain alignment. **(c)** Target self-supervised exploration. The original image and its patch shuffled view are sent to the student and teacher network. The output distributions are matched by minimizing the KL divergence, i.e., augmented features are pulled with features after scaling, which facilitates the learning of a compact feature space

data are available for adaptation. To recover the knowledge in the pre-trained source model, we leverage a self-training way to generate pseudo labels for target supervision. To alleviate the self-biasing problem caused by vanilla self-training, we introduce the source-oriented pseudo labels as regularization in Sect. 3.2. Considering that general SFDA methods align the marginal distribution, they could fail in adapting the source knowledge and mitigating domain shift in FAS where intra-class distances are prone to being larger than inter-class distances. Therefore, we propose a novel contrastive domain alignment module tailored for FAS that aligns target features to source prototypes for conditional distribution alignment with theoretical insights in Sect. 3.3. For unseen attack types not covered by the source knowledge, we introduce a target self-supervised exploration module with patch shuffle data augmentation to get rid of the facial structure and mine the intrinsic spoofing features in Sect. 3.4.

Figure 3 illustrates the overall architecture of our proposed framework that consists of a pre-trained source model and a trainable target model. The pre-trained source model consists of a feature extractor and a one-layer linear classifier, the parameters of which are fixed during adaptation. The feature extractor consists of a transformer encoder for feature encoding and a convolution layer for feature embedding. The target model consists of a student network and a teacher network. The student network consists of a feature extractor with multi-branch classifiers. The parameters of each target module are initialized by the parameters of the pre-trained source model.

3.2 Self-training with Source Regularization

Self-training Baseline (ST). Given the target domain data $\mathcal{D}_T = \{\mathbf{x}_T\}$ and the student network of the target model $f_t = h_t \circ g_t$ (initialized by f_s), the network output is $\tilde{\mathbf{y}}_t = h_t(g_t(\mathbf{x}_T))$ and the self-training loss is

$$\mathcal{L}_{\text{ST}} = \mathbb{1}\left(\max\left(\mathbf{c}_T^t\right) \geq \gamma\right) \mathcal{L}_{ce}(\tilde{\mathbf{y}}_t, \overline{\mathbf{y}}_T^t), \tag{1}$$

where $\mathbf{c}_T^t = \sigma(h_t(g_t(\mathbf{x}_T)))$ is the prediction confidence, $\overline{\mathbf{y}}_T^t = \text{argmax}(h_t(g_t(\mathbf{x}_T)))$ is the generated pseudo label, and $\mathbb{1} \in \{0, 1\}$ is an indicator function that values 1 only when the input condition holds. γ is the confidence threshold to select out more reliable pseudo-labels.

Though self-training is effective in exploring unlabeled data [29], due to domain shift, it leads to the accumulated error and results in a self-biasing problem caused by noisy pseudo labels. As shown in Fig. 5, the accuracy of pseudo labels for ST gradually drops to about 50%, which is no better than a random guess for binary classification. Therefore, we introduce the regularization of source-oriented knowledge to alleviate the self-biasing problem.

Source-Oriented Regularization (SR). The target data $\mathcal{D}_T = \{\mathbf{x}_T\}$ are fed into the fixed pre-trained source model $f_s = h_s \circ g_s$ to obtain the source-oriented pseudo labels $\overline{\mathbf{y}}_T^s = \text{argmax}(h_s(g_s(\mathbf{x}_T)))$ and prediction confidence $\mathbf{c}_T^s = \sigma(h_s(g_s(\mathbf{x}_T)))$. The cross-entropy loss for SR compares the output $\tilde{\mathbf{y}}_{t2s} = h_{t2s}(g_t(\mathbf{x}_T))$ of h_{t2s} with $\overline{\mathbf{y}}_T^s$ as

$$\mathcal{L}_{\text{SR}} = \mathbb{1}\left(\max\left(\mathbf{c}_T^s\right) \geq \gamma\right) \mathcal{L}_{ce}(\tilde{\mathbf{y}}_{t2s}, \overline{\mathbf{y}}_T^s) \tag{2}$$

Then, ST and SR are dynamically adjusted during training. Due to domain shift, the target model produces many noisy pseudo labels in the early stage of training and generates more reliable pseudo labels as the training proceeds. Thus, we assign higher importance to SR at first and gradually increase the importance of ST. The overall loss is formulated as

$$\mathcal{L}_{\text{SSR}} = \alpha \cdot \mathcal{L}_{\text{ST}} + (1 - \alpha) \cdot \mathcal{L}_{\text{SR}}. \tag{3}$$

Here, the hyperparameter α gradually increases from 0 to 1.

3.3 Contrastive Domain Alignment

As discussed in Sect. 1, in real applications, faces are captured by various cameras under different environments, leading to distribution discrepancies in illumination, background and resolution. To mitigate the distribution discrepancies between source and target domains, DA methods employ MMD loss or adversarial learning, which requires full access to the source data. In the source-free setting, based on the source hypothesis, existing SFDA methods [1,17] align the marginal distributions of the source and target domains, i.e., $P(g_t(\mathbf{x}_S)) = P(g_t(\mathbf{x}_T))$.

However, such a marginal distribution alignment regardless of the categories suffers degraded performance in FAS. Since the intra-class distance tends to

exceed the inter-class distance in FAS, features of different categories exhibit close proximity. For example, given a real subject, the corresponding fake faces with the same identity have similar facial features, while the real faces with different identities have different facial features. As shown in Fig. 2, such a marginal distribution alignment [17] may align the features of real faces with those of fake ones, which implies the different conditional distribution $P(g_t(\mathbf{x}_S)|\mathbf{y}_S) \neq P(g_t(\mathbf{x}_T)|\mathbf{y}_T)$ and affects the discrimination ability.

Thus, as shown in Fig. 3(b), we propose a contrastive domain alignment module to align the conditional distribution between the source and target domains. Due to the inaccessibility of source data, we propose to use the weights of pre-trained classifier h_s as the feature embeddings of the source prototype to compute the supervised contrastive loss.

Proposition 1. *Given a trained model $f_s = h_s \circ g_s$, where g_s is the feature extractor and h_s is the one-layer linear classifier, the ℓ_2-normalized weight vectors $\{\mathbf{w}_s^{real}, \mathbf{w}_s^{fake}\}$ of the classifier are the equivalent representation of the feature embeddings $\{\mathbf{z}_s^{real}, \mathbf{z}_s^{fake}\}$ of the source prototypes for calculating the supervised contrastive loss.*

Proof. Please refer to the supplementary material.

With the generated pseudo labels denoting the category of the feature embeddings of the target data anchor, we have the supervised contrastive loss as

$$\mathcal{L}_{\text{CDA}} = -\sum_{i=1}^{N_t}\sum_{m=1}^{M}\left[\mathbb{1}(\max(\mathbf{c}_T^{t,i}) \geq \gamma, \overline{\mathbf{y}}_T^{t,i} = m)\cdot\log\frac{\exp(\langle\mathbf{z}_t^i, \mathbf{w}_s^m\rangle/\tau)}{\sum_{j=1}^{M}\exp(\langle\mathbf{z}_t^i, \mathbf{w}_s^j\rangle/\tau)}\right], \quad (4)$$

where $\mathbf{c}_T^{t,i} = \sigma(h_t(g_t(\mathbf{x}_T^i)))$, $\overline{\mathbf{y}}_T^{t,i} = \text{argmax}(h_t(g_t(\mathbf{x}_T^i)))$, $\mathbf{z}_t^i = g_t(\mathbf{x}_T^i)$, $\langle\cdot,\cdot\rangle$ denotes the inner product, τ is the temperature parameter, and M is the number of total categories. The contrastive domain alignment module has two properties: (1) pull together the feature embeddings of real (fake) faces in the target domain and those of the same category in the source domain to align the conditional distribution (green arrow in Fig. 3 (b)); (2) push apart the feature embeddings of real (fake) faces in the target domain from those of different categories in the source domain to enhance the discrimination ability (red arrow in Fig. 3 (b)).

3.4 Target Self-supervised Exploration

For FAS applications, novel fake faces are continuously evolved and it is likely to encounter diverse attack types or collecting ways unseen in the source data. For example, spoofing features of 2D attacks and 3D mask attacks are quite different. For the cases where distribution discrepancies are enormous and source knowledge fails to apply, the generalization ability will decrease. Thus, we introduce a target self-supervised exploration (TSE) module to mine the valuable information from the target domain. However, traditional data augmentation fails to fit with the spirit of FAS to capture detailed features. Taking the whole image as input will inevitably introduce global facial information. Thus, to suppress

facial structure information as the biased source knowledge that leads to larger intra-class distances than inter-class distances, patch shuffle [47] is leveraged as a data augmentation strategy to destroy the face structure and learn a more compact feature space. Moreover, TSE is naturally independent of pseudo labels and can boost the tolerance to the wrongly transferred source supervision. The difference between our method and self-supervised methods [3,5] lies in the fact that we utilize the patch shuffle augmentation specifically for FAS and the target model is initialized by the pre-trained source model.

Specifically, a Siamese-like architecture is implemented to maximize the similarity of two views from one image [4], which consists of a student network (i.e., $g_t^{stu} \triangleq g_t, h_t^{stu} \triangleq h_t, f_t^{stu} \triangleq f_t$) and a teacher network $f_t^{tea} = g_t^{tea} \circ h_t^{tea}$. The student network is optimized by gradient descent, whereas the teacher network is updated with an exponential moving average (EMA). Given a target data $\mathbf{x}_T$, a patch-disordered view $\mathbf{x}_{T'}$ is obtained by splitting and splicing. We firstly divide the image into several patches and then randomly permute the image patches to form a new image as a jigsaw. The original view $\mathbf{x}_T$ and patch-permuted view $\mathbf{x}'_T$ are alternatively fed into the student and teacher networks to obtain two pairs of output probability distributions $\{P_{stu}, P'_{tea}\}$ and $\{P'_{stu}, P_{tea}\}$. Since the two views contain the same detailed real/fake features, the output should be consistent, which is matched by minimizing the Kullback-Leibler (KL) divergence.

$$\mathcal{L}_{\text{TSE}} = D_{KL}(P'_{tea} \| P_{stu}) + D_{KL}(P_{tea} \| P'_{stu}) \tag{5}$$

After updating θ_t with Eq. (5) by gradient descent, the parameters θ_t^{tea} of the teacher network are updated with an EMA as $\theta_t^{tea} \leftarrow l\theta_t^{tea} + (1-l)\theta_t$. l is the rate parameter.

The proposed framework for FAS is trained in an end-to-end manner as

$$\mathcal{L} = \mathcal{L}_{\text{SSR}} + \lambda_1 \cdot \mathcal{L}_{\text{CDA}} + \lambda_2 \cdot \mathcal{L}_{\text{TSE}}, \tag{6}$$

where λ_1 and λ_2 are hyper-parameters to balance the losses.

4 Experiments

4.1 Experimental Settings

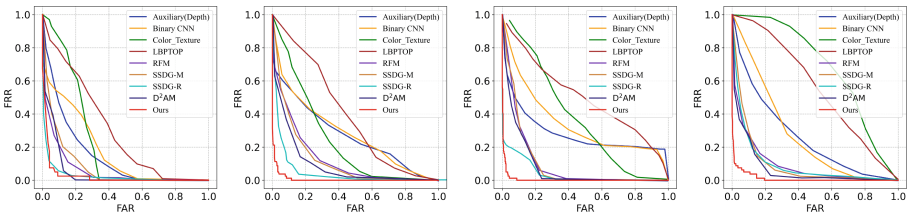
Datasets. Evaluations are made on five public datasets: Idiap Replay-Attack [7] (denoted as I), OULU-NPU [2] (denoted as O), CASIA-MFSD [50] (denoted as C), MSU-MFSD [38] (denoted as M) and CelebA-Spoof [49] (denoted as CA). CA is significantly largest with huge diversity.

Testing Scenarios. Following [27], one dataset is treated as one domain. For simplicity, we use A & B→C for the scenario that trains on the source domains A and B, and tests on the target domain C. There are thirteen scenarios in total:

- **Multi-source Domains Cross-dataset Test:** O & C & I→M, O & M & I→C, O & C & M→I, and I & C & M→O.

Table 1. HTER and AUC for multi-source domains cross-dataset test. From top to bottom, compared methods are state-of-the-art deep learning FAS (DL-FAS), DG based FAS (DG-FAS), DA based FAS (DA-FAS), SFDA based FAS (SFDA-FAS) and state-of-the-art general SFDA methods (SOTA-SFDA). SourceOnly is our pre-trained source model and (best) is the target model after adaptation. Our average result is based on 3 independent runs with different seeds to report the mean value with standard deviation. Lv et al.(base) is the pre-trained source model and (SE) is the target model after adaptation. † indicates our reproduced results with the released code.

	Methods	O & C & I →M		O & M & I→C		O & C & M→I		I & C & M→O	
		HTER(%)↓	AUC(%)↑	HTER(%)↓	AUC(%)↑	HTER(%)↓	AUC(%)↑	HTER(%)↓	AUC(%)↑
DL-FAS	Binary CNN [40]	29.25	82.87	34.88	71.94	34.47	65.88	29.61	77.54
	Auxiliary [18]	22.72	85.88	33.52	73.15	29.14	71.69	30.17	77.61
DG-FAS	RFM [28]	17.30	90.48	13.89	93.98	20.27	88.16	16.45	91.16
	SSDG-R [11]	7.38	97.17	10.44	95.94	11.71	96.59	15.61	91.54
	D^2AM [6]	15.43	91.22	12.70	95.66	20.98	85.58	15.27	90.87
DA-FAS	SDA [37]	15.4	91.8	24.5	84.4	15.6	90.1	23.1	84.3
	ADA [35]	16.9	-	24.2	-	23.1	-	25.6	-
	Wang *et al.* [36]	16.1	-	22.2	-	22.7	-	24.7	
	Quan *et al.* [25]	7.82±1.21	**97.67±1.09**	4.01±0.81	98.96±0.77	10.36±1.86	97.16±1.04	14.23±0.98	93.66±0.75
SFDA-FAS	Lv et al. [20](base)	19.28	-	27.77	-	23.58	-	18.22	-
	Lv et al. [20](SE)	18.17	-	25.51	-	20.04	-	17.5	-
SOTA-SFDA	TENT† [34]	9.58	96.18	16.67	93.12	11.25	95.63	14.13	93.20
	SHOT† [17]	8.33	95.45	17.96	91.67	9.75	96.64	13.33	93.77
Ours	SourceOnly	12.50	93.71	20.00	90.53	16.25	90.99	17.26	91.80
	SDA-FAS (best)	**5.00**	**97.96**	**2.40**	**99.72**	**2.62**	**99.48**	**5.07**	**99.01**
	SDA-FAS (avg.)	**5.97±1.19**	97.38±0.54	**3.08±0.24**	**99.54±0.19**	**3.54±0.46**	**99.11±0.41**	**6.52±1.26**	**98.37±0.25**



(a) O&C&I→M (b) O&M&I→C (c) O&C&M→I (d) I&C&M→O

Fig. 4. ROC curves for multi-source domains cross-dataset test on O, C, I and M.

- **Limited Source Domains Cross-dataset Test:** M & I→C and M & I→O.
- **Cross-dataset Test on Large-scale CA:** M & C & O→CA.
- **Single Source Domain Cross-dataset Test:** C→I, C→M, I→C, I→M, M→C, and M→I.

Evaluation Metrics. Following [11,27], Half Total Error Rate (HTER) (half of the summation of false acceptance rate and false rejection rate) and the Area Under the Curve (AUC) are used as the evaluation metrics.

Implementation Details. Following [11], MTCNN [48] is adopted for face detection. The detected faces are normalized to $256 \times 256 \times 3$ as inputs. DeiT-S [31] pre-trained on ImageNet is used as the transformer encoder. For pre-training on the source data, we randomly specify a 0.9/0.1 train-validation split and get the optimal model based on the HTER of the validation split. For adaptation,

the model is finetuned on the train set of target data and test on the test set, ensuring the test set is unseen in the whole procedure. The source code is released at https://github.com/YuchenLiu98/ECCV2022-SDA-FAS.

4.2 Experimental Results

Multi-source Domains Cross-dataset Test. Table 1 shows our SDA-FAS improves conventional deep learning FAS methods a lot by mitigating distribution discrepancies across different datasets. Besides, SDA-FAS performs better than DG based methods by exploiting unlabeled target data, as shown in Fig. 4. Moreover, SDA-FAS even outperforms the state-of-the-art DA method Quan et al. under a more challenging source-free setting, i.e., 7.71% HTER reduction and 4.71% AUC gain (lower HTER and higher AUC for better performance) for I & C & M→O that tests on the largest O dataset (among I, C, M and O). Furthermore, compared with SFDA based FAS method Lv et al. (SE), we greatly improve the performance, i.e., 3.77% vs. 20.30% HTER on average. Based on the pre-trained source model, our SDA-FAS achieves a large performance gain after adaptation with 12.7% HTER reduction on average, while Lv et al. only achieve 1.9%, validating the effectiveness of our adaptation framework. Finally, our SDA-

Table 2. HTER and AUC for test on O and C with limited source domain datasets.

Methods	M & I→C		M & I→O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)
LBPTOP [9]	45.27	54.88	47.26	50.21
SSDG-M [11]	31.89	71.29	36.01	66.88
RFM [28]	36.34	67.52	29.12	72.61
D^2AM [6]	32.65	72.04	27.70	75.36
SourceOnly	31.11	77.10	35.14	70.73
SDA-FAS	**15.37**	**91.35**	**22.53**	**83.54**

Table 3. HTER and AUC for test on large-scale CA.

Methods	M & C & O→CA	
	HTER(%)	AUC(%)
GRL Layer [10]	29.1	76.4
Domain-confusion [33]	33.7	70.3
Saha et al. [26]	27.1	79.2
Panwar et al. [23]	26.1	80.0
SourceOnly	29.7	77.5
SDA-FAS	**18.9**	**90.9**

Table 4. HTER(%) for single source domain cross-dataset test on C, I, and M datasets.

Methods	C→I	C→M	I→C	I→M	M→C	M→I	avg
Auxiliary [18]	27.6	-	28.4	-	-	-	-
Li et al. [16]	39.2	14.3	26.3	33.2	10.1	33.3	26.1
ADA [35]	17.5	9.3	41.6	30.5	17.7	5.1	20.3
Wang et al. [36]	15.6	**9.0**	34.2	29.0	16.8	**3.0**	17.9
USDAN-Un [12]	16.0	9.2	30.2	25.8	13.3	3.4	16.3
Lv et al. [20] (base)	21.1	-	34.4	-	-	-	-
Lv et al. [20] (SE)	18.9	-	30.1	-	-	-	-
SourceOnly	37.1	27.1	34.6	27.5	27.6	17.9	28.6
SDA-FAS	**11.5**	10.4	**19.6**	**24.1**	**10.0**	3.7	**13.2**

FAS outperforms the state-of-the-art general SFDA methods by proposing an adaptation framework specifically designed for FAS.

Limited Source Domains Cross-dataset Test. Compared with state-of-the-art DG method D^{2}AM, SDA-FAS improves the performance a lot by effectively using available unlabeled target data, i.e., 17.28% HTER reduction and 19.31% AUC gain for M & I→C, as shown in Table 2.

Cross-dataset Test on Large-Scale CA. For the most challenging test M & C & O→CA, where CA is much larger with unseen spoofing types (3D mask attacks), our SDA-FAS reduces HTER by 7.2% and increases AUC by 10.9% in comparison to the state-of-the-art DA method Panwar et al., as shown in Table 3. The promising results under a more practical source-free setting demonstrate that our method is effective and trustworthy for complex real-world scenarios.

Single Source Domain Cross-dataset Test. Table 4 shows that under a more difficult source-free setting, SDA-FAS outperforms all DA methods under four of the six tests and achieves the best average result (13.2% HTER). Besides, compared with the SFDA method Lv et al., SDA-FAS achieves a much larger performance gain after adaptation, 15.0% vs. 4.3% HTER reduction for I→C.

Table 5. Ablation studies on different components of our proposed SDA-FAS.

ST	SR	CDA	TSE	O & C & I→M		O & M & I→C		O & C & M→I		I & C & M→O	
				HTER (%)	AUC (%)	HTER (%)	AUC (%)	HTER (%)	AUC (%)	HTER (%)	AUC (%)
✓	✗	✗	✗	8.33	95.02	8.89	97.12	8.50	96.33	14.68	93.13
✓	✓	✗	✗	7.08	96.42	6.67	97.97	6.25	98.49	9.44	96.76
✓	✓	✓	✗	5.42	97.35	4.44	98.85	4.37	98.96	7.50	97.72
✓	✓	✓	✓	**5.00**	**97.96**	**2.40**	**99.72**	**2.62**	**99.48**	**5.07**	**99.01**

Table 6. HTER and AUC for unseen 3D mask attack type test on part of CA.

Methods	M & C & O→CA(3D mask)	
	HTER(%)	AUC(%)
Ours w/o TSE	20.52	89.91
Ours	**11.27**	**97.06**

Table 7. AUC(%) of the cross attack type test on C, I and M. Two attack types of unlabeled target data are used for training and tested on unseen attack type.

Methods	CASIA-MFSD (C)			Replay-Attack (I)			MSU (M)		
	Video	Cut	Warped	Video	Digital	Printed	Printed	HR	Mobile
DTN [19]	90.0	97.3	97.5	**99.9**	**99.9**	**99.6**	81.6	**99.9**	97.5
Ours	**98.3**	**97.7**	**97.6**	**99.9**	99.5	99.3	**86.3**	99.6	**97.8**

4.3 Ablation Studies

Each Component of the Network. The proposed framework and its variants are evaluated on multi-source domains cross-dataset test. Table 5 shows that, based on ST, SR improves the performance by introducing source-oriented regularization to alleviate the self-biasing problem. Besides, the performance improves with CDA added, demonstrating the effectiveness of conditional

domain alignment to mitigate distribution discrepancies and enhance the discrimination ability. Moreover, TSE can further improve the performance, especially on the large test dataset (e.g., I & C&M→O), reflecting its power in self-exploring valuable information in large target data.

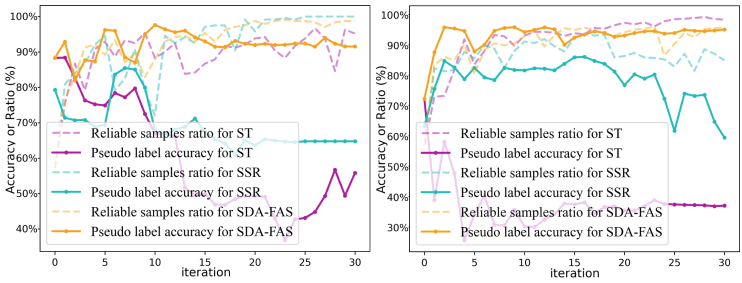
Portion of Target Data Used. Firstly, we randomly sample 10% and 50% of live and spoof faces in the training set for adaptation. Table 8 shows, even with 10% training samples, SDA-FAS improves the performance a lot, manifesting the validity for real scenarios with few data. For example, SDA-FAS reduces HTER by 9.44% after adaptation using only 24 unlabeled samples in C. Secondly, for extreme cases in FAS where live faces are much larger than spoof faces, we randomly sample 5%, 10% and 50% of spoof faces in the training set. With only 5% spoof faces (i.e., 9 samples), SDA-FAS reduces HTER by 8.71% after adaptation to C, demonstrating the effectiveness for more challenging scenarios.

Unseen Attack Types. To further evaluate TSE in self-exploring the target data, we reconstitute CA test set with all real faces and only 3D mask attack faces (unseen in the source data where only 2D attack types exist), and conduct experiments under M & C & O→CA (3D mask). As shown in Table 6, TSE significantly improves the performance, i.e., 9.25% HTER reduction and 7.15% AUC gain, demonstrating its effectiveness in self-exploring novel attack types in the case where the source knowledge fail to apply. Corresponding qualitative analysis is conducted in the supplementary material by visualizing a few hard 3D mask faces. Moreover, following protocols in [19], only partial attack types with unlabeled target data are tested. Table 7 shows our method outperforms DTN [19] that is fully supervised with labeled data. By adapting source knowledge, our method achieves better performance in an unsupervised manner.

Statistics of Pseudo Labels. As shown in Fig. 5, self-training (ST) results in a self-biasing problem and the accuracy of pseudo labels gradually drops to less than 50%. Self-training with source-oriented regularization (SSR) can alleviate the self-biasing problem, and the accuracy achieves a steady improvement to 70%. Moreover, with CDA mitigating domain discrepancies and TSE self-exploring target data, SDA-FAS achieves the highest accuracy exceeding 90%.

Table 8. Experiments on different portion of target train data and spoof faces. L denotes live faces and S denotes spoof faces, respectively.

Protocols	O & C & I→M		O & M & I→C		O & C & M→I		I & C & M→O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
0% (L+S)	12.50	93.71	20.00	90.53	16.25	90.99	17.26	91.80
10% (L+S)	10.00	96.10	10.56	95.86	8.50	98.21	10.07	96.60
50% (L+S)	7.14	96.45	5.37	99.04	4.87	98.93	7.08	98.30
100%L+5%S	10.00	95.74	11.29	95.24	9.28	95.70	10.76	96.24
100%L+10%S	8.57	96.04	8.33	97.53	7.50	97.57	9.65	96.38
100%L+50%S	5.71	98.52	3.52	99.37	3.75	99.28	5.83	98.70



(a) $O\&M\&I \rightarrow C$ (b) $I\&C\&M \rightarrow O$

Fig. 5. Reliable samples ratio (dashed line) and pseudo labels accuracy (solid line) with respect to the updating iteration.

4.4 Visualizations

Attention Map. Figure 6 shows that, for real faces in rows 1 and 3, our method exhibits dense attention maps to effectively capture the physical structure of human faces. For the cut attack in row 2, the cut area of eyes is precisely specified, whereas the finger hint holding the paper is detected for the print attack in row 4. The attention maps suggest that SDA-FAS can model the features of live faces well and also precisely capture the intrinsic and detailed spoofing cues. Therefore, it can generalize well to the target domain.

Feature Space. We select all samples of target data for t-SNE visualizations. As shown in Fig. 7, after adaptation, the features of fake faces and real faces are better separated on the target domain compared to those before adaptation.

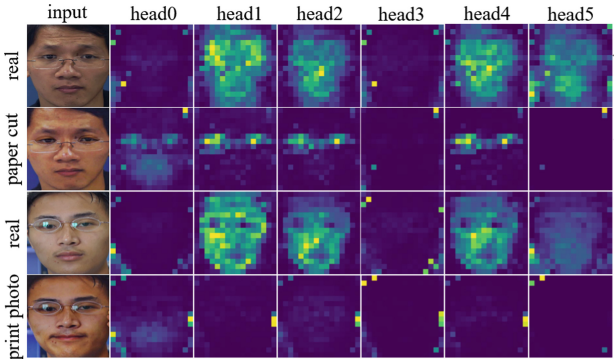
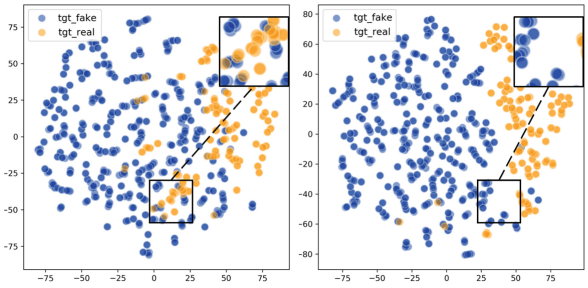


Fig. 6. Attention maps [3] from the last layer of the transformer encoder under O & M & I→C. Column 1: cropped input image. Columns 2–7: six heads of the transformer encoder. Rows 1–2: attention maps for subject 1's real face and paper-cut attack. Rows 3–4: attention maps for subject 2's real face and print photo attack.



(a) Before adaptation (b) After adaptation

Fig. 7. The t-SNE [21] visualization of the extracted features by our model with adaptation (right) and without adaptation (left) under O & M & I→C.

5 Conclusion

In this paper, we propose a novel adaptation framework for face anti-spoofing under a practical yet challenging source-free setting, which protects the security and privacy of human faces. Specifically, source-oriented regularization is introduced to alleviate the self-biasing problem of self-training. Besides, we propose a novel contrastive domain alignment module to align the conditional distribution across domains for mitigating the discrepancies. Moreover, self-supervised learning is adopted to self-explore the target data for robust features under enormous domain discrepancies where source knowledge is inapplicable. Extensive experiments validate the effectiveness of our method statistically and visually.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China under Grants 61932022, 61931023, 61971285, 62120106007, and in part by the Program of Shanghai Science and Technology Innovation Project under Grant 20511100100.

References

1. Ahmed, S.M., Raychaudhuri, D.S., Paul, S., Oymak, S., Roy-Chowdhury, A.K.: Unsupervised multi-source domain adaptation without access to source data. In: CVPR, pp. 10103–10112. IEEE (2021)
2. Boulkenafet, Z., Komulainen, J., Li, L., Feng, X., Hadid, A.: OULU-NPU: a mobile face presentation attack database with real-world variations. In: FG, pp. 612–618. IEEE (2017)
3. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: ICCV, pp. 9630–9640. IEEE (2021)
4. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: ICML, pp. 1597–1607. PMLR (2020)
5. Chen, X., He, K.: Exploring simple Siamese representation learning. In: CVPR, pp. 15750–15758. IEEE (2021)

6. Chen, Z., et al.: Generalizable representation learning for mixture domain face anti-spoofing. In: AAAI, pp. 1132–1139. AAAI Press (2021)
7. Chingovska, I., Anjos, A., Marcel, S.: On the effectiveness of local binary patterns in face anti-spoofing. In: BIOSIG, pp. 1–7 (2012)
8. de Freitas Pereira, T., Anjos, A., De Martino, J.M., Marcel, S.: *LBP*TOP based countermeasure against face spoofing attacks. In: Park, J.-I., Kim, J. (eds.) ACCV 2012. LNCS, vol. 7728, pp. 121–132. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37410-4_11
9. Freitas Pereira, T., et al.: Face liveness detection using dynamic texture. EURASIP J. Image Video Process. **2014**(1), 1–15 (2014). https://doi.org/10.1186/1687-5281-2014-2
10. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: ICML, pp. 1180–1189. PMLR (2015)
11. Jia, Y., Zhang, J., Shan, S., Chen, X.: Single-side domain generalization for face anti-spoofing. In: CVPR, pp. 8484–8493. IEEE (2020)
12. Jia, Y., Zhang, J., Shan, S., Chen, X.: Unified unsupervised and semi-supervised domain adaptation network for cross-scenario face anti-spoofing. Pattern Recogn. **115**, 107888 (2021)
13. Khosla, P., et al.: Supervised contrastive learning. In: NeurIPS, pp. 18661–18673. Curran Associates, Inc. (2020)
14. Kim, Y., Hong, S., Cho, D., Park, H., Panda, P.: Domain adaptation without source data. IEEE Trans. Artif. Intell. **2**(6), 508–518 (2020)
15. Komulainen, J., Hadid, A., Pietikäinen, M.: Context based face anti-spoofing. In: BTAS, IEEE (2013)
16. Li, H., Li, W., Cao, H., Wang, S., Huang, F., Kot, A.C.: Unsupervised domain adaptation for face anti-spoofing. IEEE Trans. Inf. Forensics Secur. **13**(7), 1794–1809 (2018)
17. Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: ICML, pp. 6028–6039. PMLR (2020)
18. Liu, Y., Jourabloo, A., Liu, X.: Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In: CVPR, pp. 389–398. IEEE (2018)
19. Liu, Y., Stehouwer, J., Jourabloo, A., Liu, X.: Deep tree learning for zero-shot face anti-spoofing. In: CVPR, pp. 4680–4689. IEEE (2019)
20. Lv, L., et al.: Combining dynamic image and prediction ensemble for cross-domain face anti-spoofing. In: ICASSP, pp. 2550–2554 (2021)
21. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(86), 2579–2605 (2008)
22. van den Oord, A., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
23. Panwar, A., Singh, P., Saha, S., Paudel, D.P., Van Gool, L.: Unsupervised compound domain adaptation for face anti-spoofing. In: FG, IEEE (2021)
24. Patel, K., Han, H., Jain, A.K.: Secure face unlock: spoof detection on smartphones. IEEE Trans. Inf. Forensics Secur. **11**(10), 2268–2283 (2016)
25. Quan, R., Wu, Y., Yu, X., Yang, Y.: Progressive transfer learning for face anti-spoofing. IEEE Trans. Image Process. **30**(3), 3946–3955 (2021)
26. Saha, S., et al.: Domain agnostic feature learning for image and video based face anti-spoofing. In: CVPR Workshops, pp. 802–803. IEEE (2020)
27. Shao, R., Lan, X., Li, J., Yuen, P.C.: Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In: CVPR, pp. 10023–10031. IEEE (2019)

28. Shao, R., Lan, X., Yuen, P.C.: Regularized fine-grained meta face anti-spoofing. In: AAAI, pp. 11974–11981. AAAI Press (2020)
29. Sohn, K., et al.: FixMatch: simplifying semi-supervised learning with consistency and confidence. In: NeurIPS, pp. 596–608. Curran Associates, Inc. (2020)
30. The European Parliament and The Council of the European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of European Union (OJ) 59(L119), pp. 1–88 (2016)
31. Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H.: Training data-efficient image transformers & distillation through attention. In: ICML, pp. 10347–10357. PMLR (2021)
32. Tu, X., Zhang, H., Xie, M., Luo, Y., Zhang, Y., Ma, Z.: Deep transfer across domains for face antispoofing. J. Electron. Imaging **28**(4), 043001 (2019)
33. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: CVPR, pp. 7167–7176. IEEE (2017)
34. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. In: ICLR (2021)
35. Wang, G., Han, H., Shan, S., Chen, X.: Improving cross-database face presentation attack detection via adversarial domain adaptation. In: ICB, IEEE (2019)
36. Wang, G., Han, H., Shan, S., Chen, X.: Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection. IEEE Trans. Inf. Forensics Secur. **16**, 56–69 (2021)
37. Wang, J., Zhang, J., Bian, Y., Cai, Y., Wang, C., Pu, S.: Self-domain adaptation for face anti-spoofing. In: AAAI, pp. 2746–2754. AAAI Press (2021)
38. Wen, D., Han, H., Jain, A.K.: Face spoof detection with image distortion analysis. IEEE Trans. Inf. Forensics Secur. **10**(4), 746–761 (2015)
39. Xu, Z., Li, S., Deng, W.: Learning temporal features using LSTM-CNN architecture for face anti-spoofing. In: ACPR, pp. 141–145. IEEE (2015)
40. Yang, J., Lei, Z., Li, S.Z.: Learn convolutional neural network for face anti-spoofing. arXiv preprint arXiv:1408.5601 (2014)
41. Yang, S., Wang, Y., van de Weijer, J., Herranz, L., Jui, S.: Exploiting the intrinsic neighborhood structure for source-free domain adaptation. In: NeurIPS, pp. 29393–29405. Curran Associates, Inc. (2021)
42. Yang, S., Wang, Y., van de Weijer, J., Herranz, L., Jui, S.: Generalized source-free domain adaptation. In: ICCV, pp. 8978–8987. IEEE (2021)
43. Yu, Z., Li, X., Niu, X., Shi, J., Zhao, G.: Face anti-spoofing with human material perception. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12352, pp. 557–575. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58571-6_33
44. Yu, Z., Li, X., Shi, J., Xia, Z., Zhao, G.: Revisiting pixel-wise supervision for face anti-spoofing. IEEE Trans. Biomet. Behav. Ident. Sci. **3**(3), 285–295 (2021)
45. Yu, Z., Wan, J., Qin, Y., Li, X., Li, S.Z., Zhao, G.: NAS-FAS: static-dynamic central difference network search for face anti-spoofing. IEEE Trans. Pattern Anal. Mach. Intell. **43**(9), 3005–3023 (2021)
46. Yu, Z.,et al.: Searching central difference convolutional networks for face anti-spoofing. In: CVPR, pp. 5295–5305. IEEE (2020)
47. Zhang, K.Y., et al.: Structure destruction and content combination for face anti-spoofing. In: IJCB, IEEE (2021)

48. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process. Lett. **23**(10), 1499–1503 (2016)
49. Zhang, Y., et al.: CelebA-spoof: large-scale face anti-spoofing dataset with rich annotations. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12357, pp. 70–85. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58610-2_5
50. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: A face antispoofing database with diverse attacks. In: ICB, pp. 26–31. IEEE (2012)