Single-Side Domain Generalization for Face Anti-Spoofing

Yunpei Jia[1,2], Jie Zhang[1,2], Shiguang Shan[1,2,3], Xilin Chen[1,2]
[1]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing, 100190, China
[2]University of Chinese Academy of Sciences, Beijing, 100049, China
[3]CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai, 200031, China
yunpei.jia@vipl.ict.ac.cn, {zhangjie, sgshan, xlchen}@ict.ac.cn

Abstract

Existing domain generalization methods for face anti-spoofing endeavor to extract common differentiation features to improve the generalization. However, due to large distribution discrepancies among fake faces of different domains, it is difficult to seek a compact and generalized feature space for the fake faces. In this work, we propose an end-to-end single-side domain generalization framework (SSDG) to improve the generalization ability of face anti-spoofing. The main idea is to learn a generalized feature space, where the feature distribution of the real faces is compact while that of the fake ones is dispersed among domains but compact within each domain. Specifically, a feature generator is trained to make only the real faces from different domains undistinguishable, but not for the fake ones, thus forming a single-side adversarial learning. Moreover, an asymmetric triplet loss is designed to constrain the fake faces of different domains separated while the real ones aggregated. The above two points are integrated into a unified framework in an end-to-end training manner, resulting in a more generalized class boundary, especially good for samples from novel domains. Feature and weight normalization is incorporated to further improve the generalization ability. Extensive experiments show that our proposed approach is effective and outperforms the state-of-the-art methods on four public databases. The code is released online[1].

1. Introduction

In recent years, face recognition techniques have been widely exploited in our daily life, especially in the fields of smartphones login, access control, etc. However, many presentation attacks have emerged (*e.g.*, print attack, video attack, and 3D mask attack), which has led to a huge security risk on face recognition systems and become an increasingly critical concern in the face recognition field. To tackle this issue, various face anti-spoofing methods have been proposed, which can be coarsely categorized into texture-based methods and temporal-based methods. The texture-based methods utilize hand-craft descriptors or data-driven deep learning to extract texture cues discriminative between the real faces and the fake ones, such as the color [5], distortion cues [11, 39], etc. In contrast, the temporal-based methods leverage various temporal cues in consecutive face frames, such as rPPG [20, 21] and optical flow [2, 4].

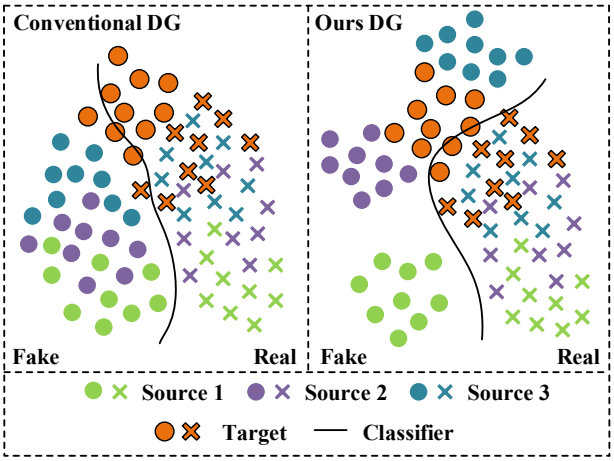


Figure 1. Left: Conventional domain generalization methods align source domains to learn a shared feature space, which fail to get a discriminative class boundary on the unseen domain. Right: Our single-side domain generalization method aggregates all the real examples while separates the fake ones from different domains to learn a class boundary, generalizing better to the novel domain.

Although existing state-of-the-art methods have obtained promising results under intra-database testing scenarios, they cannot generalize well in case of cross-database testing, where training (source domain[2]) and testing (target do-

[1]https://github.com/taylover-pei/SSDG-CVPR2020

[2]The term domain in this paper represents a database.

main) data come from different domains. The reason behind is that traditional methods take no consideration of the intrinsic distribution relationship among different domains and thus extract discriminative features of database biased [35], leading to poor generalization to unseen domains. To address the problem, recent face anti-spoofing methods [19, 38] adopt domain adaptation techniques to minimize the distribution discrepancy between the source and the target domain by utilizing unlabeled target data. However, in many real-world scenarios, it is difficult and expensive to collect a lot of unlabeled target data for training, and even no information about the target domain is available.

Therefore, some researchers start to address the face anti-spoofing problem from the perspective of domain generalization (DG), which aims to train a model by utilizing multiple existing source domains without seeing any target data. Conventional DG approaches [18, 33] aim to learn a generalized feature space by aligning the distributions among multiple source domains. And they assume that the extracted features of unseen faces can be mapped nearby the shared feature space so that the model can generalize well to the novel domains. Since the real faces from both the source and the target domains are collected by imaging real people, their distribution discrepancies are small, which makes it relatively easy to learn a compact feature space. In contrast, due to the diversity of attack types and collecting ways, it is relatively hard to aggregate the features of fake faces from different domains together. Therefore, seeking a generalized feature space for the fake faces is difficult to optimize and may also affect the classification accuracy for the target domain [1, 40]. For this reason, as illustrated in the left of Figure 1, although a compact feature space for both the real and the fake examples is achieved, it still fails to learn a discriminative class boundary for the novel target domain. In consideration of the above arguments, besides constraining the real faces and the fake ones to be as distinguishable as possible, we propose to pull all the real faces aggregated while push the fake ones of different domains separated. As illustrated in the right of Figure 1, our method aims at forcing the features of fake faces more dispersed in the feature space while those of the real ones more compact, thus leading to a class boundary, which generalizes better to the target domain.

With the above thoughts in mind, we propose an end-to-end single-side domain generalization framework (SSDG), as shown in Figure 2. Specifically, a feature generator is trained competing with a domain discriminator to make the features of real faces from different domains undistinguishable, forming a single-side adversarial learning. Since the fake faces are rather diverse than the real ones, we treat the fake faces of different domains as different categories while the real ones of all domains as the other category to perform the asymmetric triplet mining, which ensures three properties: 1) fake faces of different categories are separated; 2) all the real ones regardless of domains are aggregated; 3) all the real faces and the fake ones are distinguishable. As a result, two feature distributions with different characteristics can be achieved, leading to a better generalized class boundary for the target domains. Meanwhile, feature and weight normalization is incorporated to further improve the generalization ability during training.

The main contributions of this work are summarized as follows: 1) Based on the analysis that the fake faces are rather diverse than the real ones, we propose a novel end-to-end single-side domain generalization framework. 2) We design the single-side adversarial learning and the asymmetric triplet loss to achieve different optimization goals for the real and the fake faces and perform the feature and weight normalization to further improve the performance. 3) We make comprehensive comparisons and achieve the state-of-the-art performance on four public databases.

2. Related Work

2.1. Face Anti-spoofing Methods

In this subsection, we review the most representative face anti-spoofing methods, which can be generally divided into two groups: texture-based methods and temporal-based methods, as already mentioned previously.

Texture-based methods distinguish the real faces from the fake ones through various texture cues. Many prior works adopt hand-craft descriptors for face anti-spoofing, such as LBP [9, 26], HOG [14], SURF [5], SIFT [29], etc. In recent years, with the rapid development of deep learning in computer vision, various methods turn to employ CNNs to extract more discriminative features. Yang *et al.* [42] are the first to use CNN with binary supervision for face anti-spoofing. Atoum *et al.* [3] propose a two-stream CNN architecture to extract depth features combining with the texture features to detect attacks. And the face de-spoofing method [16] inversely decomposes a spoof face into a live face and a spoof noise for classification.

Temporal-based methods make use of temporal cues in consecutive face frames for spoofing face detection. Mouth-motion detection [17] and eye-blinking detection [28, 34] are among the earliest solutions for face anti-spoofing based on the temporal cues. Recently, there exist more general methods relying on more effective temporal cues, instead of the particular liveness information. CNN-LSTM architecture is proposed in [41] to take multiple frames as input to extract temporal features for face anti-spoofing. Liu *et al.* [23] utilize the rPPG signal as the auxiliary supervision with a novel CNN-RNN network to detect attacks. More robust rPPG features are extracted by [20, 21] to detect 3D mask attack effectively. And Yang *et al.* [43] take into consideration the global temporal and local spatial cues to distinguish

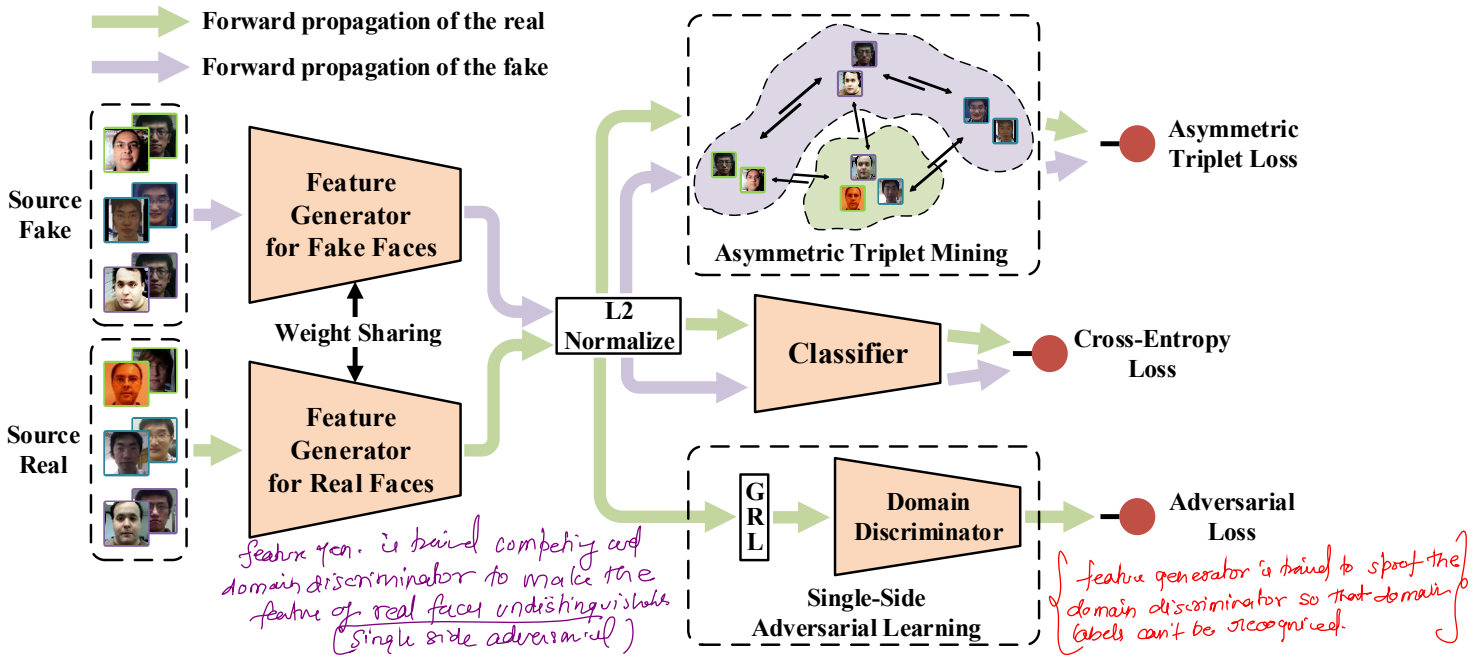


Figure 2. An overview of the proposed method. The input faces with different color borders represent examples of different domains. The parameter sharing feature generator is trained to make the feature distributions of different domains undistinguishable for the real faces but not for the fake ones under the single-side adversarial learning. Moreover, the asymmetric triplet mining is implemented to separate the fake faces while aggregate the real ones of different domains to force the features of fake faces to be more dispersed in the feature space. The feature and weight normalization is incorporated to further improve the generalization ability.

the real faces from the fake ones.

Although the above methods have obtained remarkable results under intra-database testing scenarios, they cannot mine the distribution relationship among different domains and might suffer from extracting database-biased features, leading to poor generalization to unseen domains.

2.2. Domain Generalization

Both the domain adaptation methods [19, 36, 38] and zero-shot face anti-spoofing methods [24, 30] aim to improve the generalization ability. In contrast, the domain generalization (DG) methods explicitly mine the relationship among multiple source domains without accessing any target data, which generalize better to unseen domains. Most of the previous DG methods focus on minimizing the distribution discrepancies among multiple source domains to extract domain-invariant features. Motiian *et al.* [27] propose a new loss to guide the features of the same class to be close in the latent feature space. Autoencoders are exploited in [13, 18] to align the distributions of source domains for generalized features. The most related work to ours is proposed in [33], where multiple feature extractors are trained to learn a generalized feature space via adversarial learning. However, as the first attempt to address the face anti-spoofing problem from the DG point of view, its training process is not end-to-end. Moreover, due to the diversity of attack types and collecting ways, it is difficult to seek a generalized feature space for the fake faces, usually leading to a sub-optimal solution for face anti-spoofing.

3. Proposed Method

3.1. Overview

Since the distribution discrepancies are much larger among the fake faces than the real ones, it is nontrivial to align the features of fake faces from different domains. Therefore, seeking a compact and generalized feature space for both the real and the fake faces is difficult to optimize and may bring negative influences on the classification accuracy for unseen domains. In this work, we focus on asymmetric optimization goals for the real and the fake faces belonging to different domains to learn a feature space with higher generalization ability to unseen domains. As illustrated in Figure 2, we propose a single-side domain generalization framework (SSDG) for face anti-spoofing. Specifically, the feature generator is trained competing with the domain discriminator to make the features of real faces undistinguishable, forming a single-side adversarial learning process. Moreover, we propose the asymmetric triplet loss to explicitly separate the fake faces of different domains while aggregate the real ones. Additionally, feature and weight normalization is further incorporated to improve the gener-

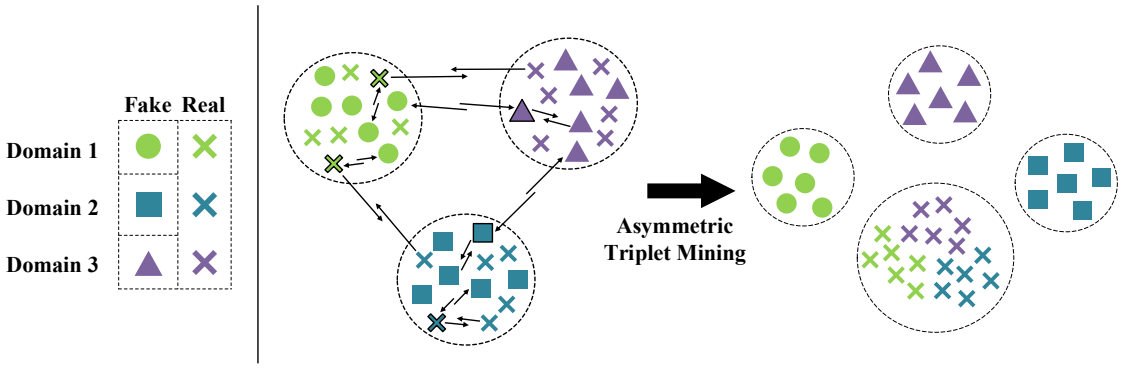


Figure 3. Illustration of the asymmetric triplet loss. A shape with the black border represents the anchor point, while the other two points linked with it are the positive and the negative ones, respectively. Asymmetric triplet mining is conducted to separate the fake faces of different domains while aggregate the real ones together. Meanwhile, the fake faces are also pulled apart away from all the real ones. After adopting the asymmetric triplet loss, the features of fake faces can be more dispersed in the feature space, leading to a class boundary better generalized.

alization ability during the training process. Therefore, the proposed SSDG method forces the fake faces to be more dispersed in the feature space while the real ones to be more compact, leading to a more generalized class boundary to unseen domains.

3.2. Single-Side Adversarial Learning

Assume there are N source domains, denoted as $D = \{D_1, D_2, ..., D_N\}$. Each of them contains two categories of face images, *i.e.*, the real faces X_r and the fake faces X_f. Since all the real faces are collected by imaging real people, we conjecture that the distribution discrepancies among them are much smaller compared to the fake ones. Therefore, seeking a generalized feature space for the real faces is relatively easy, which promotes to capture more common discriminative cues. Specifically, we propose the single-side adversarial learning to learn a generalized feature space, which is conducted only on the extracted features of real faces. In contrast, the adversarial learning is not performed for the fake ones.

We firstly separate the real faces from the fake ones of all source domains, and then feed them into the corresponding feature generators, which transform the input faces into a latent feature space as follows:

$$Z_r = G_r(X_r), Z_f = G_f(X_f), \tag{1}$$

where G_r, G_f represent the feature generators for the real and the fake faces, respectively, and Z_r, Z_f are the corresponding extracted features. Since a parameter sharing strategy is adopted to make all the parameters of G_r and G_f identical, we refer them collectively as G in the following for the sake of convenience. The domain discriminator, denoted as D, is implemented based on Z_r to determine which source domain the input features stem from. On the contrary, the feature generator is trained to spoof the domain discriminator so that the domain labels cannot be recognized. Therefore, a single-side adversarial learning procedure is designed between the feature generator and the domain discriminator to learn a generalized feature space for the real faces. During the learning procedure, the parameters of feature generator are optimized by maximizing the loss of domain discriminator while those of domain discriminator are optimized with the opposite objective. Since there are multiple source domains for classification, we utilize the standard cross-entropy loss to optimize the network under the single-side adversarial learning:

$$\min_D \max_G \mathcal{L}_{Ada}(G, D) = - \mathbb{E}_{x,y\sim X_r,Y_D} \sum_{n=1}^{N} \mathbb{1}_{[n=y]} \log D(G(x)), \tag{2}$$

where Y_D represents the set of domain labels.

In order to optimize the feature generator and the domain discriminator simultaneously, a gradient reverse layer (GRL) [12] is inserted after the feature generator, which multiplies the gradient of the adversarial loss by $-\lambda$ during backward propagation. We set $\lambda = \frac{2}{1+\exp(-10k)} - 1$ and $k = \frac{current\_iters}{total\_iters}$ with the same purpose introduced by [12] to suppress the effect of the noisy signals at the early training stage. With the single-side adversarial learning, a generalized feature space for the real faces is achieved, where common discriminative cues can be further exploited.

3.3. Asymmetric Triplet Mining

Due to the diversity of attack types and database collection ways, the distribution discrepancies are much larger among the fake faces than the real ones. Therefore, seeking a dispersed feature space for the fake is relatively easy compared to seeking a compact one. In consideration of this, we explicitly separate the fake faces of different domains to

force them to be more dispersed in the feature space. In contrast, we aggregate all the real ones to force them to be more compact. To achieve the asymmetric optimization goals for the real and fake faces, we propose the asymmetric triplet loss to perform the asymmetric triplet mining according to the categories, which promotes to learn a better class boundary for unseen domains.

Specifically, assuming there are three source domains available, we recombine the real and the fake faces coming from three different domains into four categories. As shown in the left of Figure 3, the fake faces of three different domains are treated as distinct categories (circle, square, and triangle, respectively), while all the real ones are put together into one category (cross). And then, four-category asymmetric triplet mining is conducted on the real and the fake faces to achieve the following optimization goals: 1) separate the fake faces of different domains; 2) aggregate the real faces of all source domains; 3) pull apart the fake faces away from all the real ones. After that, as shown in the right of Figure 3, the extracted features of fake faces are more dispersed than before in the feature space and those of real ones are more aggregated, leading to a better generalized class boundary for unseen domains. The feature generator is optimized as follows:

$$\min_{G} \mathcal{L}_{AsTrip}(G) = \sum_{x_i^a, x_i^p, x_i^n} (\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha), \tag{3}$$

where the labels of anchor x_i^a and positive example x_i^p are the same, while those of x_i^a and negative example x_i^n are different. The α is a pre-defined margin.

3.4. Feature and Weight Normalization

Normalization approaches have been verified effective in the field of face recognition. In this work, both feature normalization and weight normalization are incorporated to further improve the generalization ability of the proposed method.

Feature Normalization. The feature norms are highly related to the quality of the images, as discussed in [31, 37]. Due to the diversity of database collecting conditions (*e.g.*, illustration, camera quality, etc.), large differences exist among the feature norms of different face images under both the intra-database and cross-database scenarios, which hinder the feature learning process and also affect the generalization ability. Thus, we perform the l_2 normalization on the outputs of the feature generator to constrain all the features share the same Euclidean norm to further improve the performance of face anti-spoofing.

Weight Normalization. In this work, the face anti-spoofing problem is regarded as a binary classification task. Since the softmax function is utilized for training, the decision boundary can be achieved between the real and the fake faces as $\|\boldsymbol{W}_1^T\|\|\tilde{\mathbf{z}}\|\cos(\theta_1)+b_1 = \|\boldsymbol{W}_0^T\|\|\tilde{\mathbf{z}}\|\cos(\theta_0)+b_0$, where $\boldsymbol{W_i}$ is the i-th column of the parameter matrix in last fully connected layer, b_i is the corresponding bias, and θ_i is the angle between the normalized feature $\tilde{\mathbf{z}}$ and $\boldsymbol{W_i}$. Following the works of [10, 22, 37], we perform l_2 normalization on $\boldsymbol{W_i}$ to fix $\|\boldsymbol{W}_i\| = 1$ and set $b_i = 0$, which makes the decision boundary becomes $\cos(\theta_1) - \cos(\theta_0) = 0$. Therefore, we further constrain the feature learning process by the weight normalization, which promotes to learn more discriminative cues between the real and the fake faces.

3.5. Loss Function

Since all the source domain data contain labels, a face anti-spoofing classifier is implemented after the feature generator, as illustrated in Figure 2. Both the face anti-spoofing classifier and the feature generator are optimized by the standard cross-entropy loss, denoted as $\mathcal{L}_{Cls}$. Integrating all things mentioned above together, the objective of the proposed single-side domain generalization framework for face anti-spoofing is:

$$\mathcal{L}_{SSDG} = \mathcal{L}_{Cls} + \lambda_1 \mathcal{L}_{Ada} + \lambda_2 \mathcal{L}_{AsTrip}, \tag{4}$$

where λ_1 and λ_2 are the balanced parameters. Instead of decomposing the training process into two phases in [33], we train all the components in an end-to-end manner.

4. Experiment

4.1. Experimental Settings

Databases. Four public face anti-spoofing databases are utilized to evaluate the effectiveness of our method: OULU-NPU [7] (denoted as O), CASIA-FASD [45] (denoted as C), Idiap Replay-Attack [8] (denoted as I), and MSU-MFSD [39] (denoted as M). We randomly select one database as the target domain for testing and the remaining three as the source domains for training. Thus, we have four testing tasks in total: O&C&I to M, O&M&I to C, O&C&M to I, and I&C&M to O. Many differences (*e.g.*, background, resolution, illustration, ethnicity, etc.) exist under both intra-database and cross-database testing scenarios, especially for the fake ones, which cause great distribution discrepancies among them.

Implementation Details. MTCNN algorithm [44] is adopted for face detection and face alignment to perform the data pre-processing. All the detected faces are normalized to 256×256×3 as the input of the network, where only RGB channels are utilized for training to further reduce the network complexity. We train our model using only one frame information randomly selected from each video. The SGD optimizer with momentum of 0.9 and weight decay of 5e-4 is used for the optimization. The hyperparameter α is set to 0.1.

Popular datasets.
- OULU NPU
- CASIA-FASD
- MSU-MFSD
- Replay Attack.

Table 1. Evaluations of different components of the proposed method with different architectures.

Method	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(%)
SSDG-M w/o triplet	21.19	83.54	26.78	79.10	23.93	74.86	25.43	80.52
SSDG-M w/o ssad	24.05	81.94	28.11	80.15	21.29	84.52	26.62	79.59
SSDG-M w/o norm	17.86	89.76	30.11	78.38	25.57	73.92	29.74	75.48
SSDG-M	**16.67**	**90.47**	**23.11**	**85.45**	**18.21**	**94.61**	**25.17**	**81.83**
SSDG-R w/o triplet	8.81	96.85	14.33	92.28	15.21	83.09	21.98	85.54
SSDG-R w/o ssad	11.19	95.10	12.89	94.08	12.14	**96.63**	18.06	90.43
SSDG-R w/o norm	10.24	96.58	12.78	95.06	12.64	92.92	15.99	91.26
SSDG-R	**7.38**	**97.17**	**10.44**	**95.94**	**11.71**	96.59	**15.61**	**91.54**

Table 2. Comparison results between the proposed method and the corresponding baseline method with different architectures.

Method	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(%)
BDG-M	17.14	87.70	28.00	73.42	20.93	87.06	26.27	79.99
SSDG-M	**16.67**	**90.47**	**23.11**	**85.45**	**18.21**	**94.61**	**25.17**	**81.83**
BDG-R	9.52	93.52	12.78	94.38	12.86	93.06	16.46	91.39
SSDG-R	**7.38**	**97.17**	**10.44**	**95.94**	**11.71**	**96.59**	**15.61**	**91.54**

Our framework is implemented by PyTorch. And two different architectures of the feature generator are adopted for comparisons. The first one combines the feature generator with the feature embedder defined in MADDG [33]. For the second one, we replace the last average pooling layer of ResNet-18 [15] by the global pooling layer (GAP) and utilize all the above layers of GAP. Specifically, we add a fully connected layer (FC) as the bottleneck layer on top of each feature generator, which consists of 512 hidden units. The face anti-spoofing classifier is a simple linear model with a 2 nodes FC layer. And the domain discriminator contains two FC layers with 512 and 3 nodes, respectively. We denote these two different architectures by M and R for short in the following (*i.e.*, SSDG-M and SSDG-R).

Evaluation Metrics. Following the work of [33], we use the Half Total Error Rate (HTER) and the Area Under Curve (AUC) as the evaluation metrics. Moreover, the Receiver Operating Characteristic (ROC) and some visualizations (t-SNE [25] and CAM [32]) are also reported to further evaluate the performance.

4.2. Discussion

4.2.1 Influences of Each Network Component

We perform the ablation study to evaluate the performance gained by each component for different network architectures, *i.e.*, the single-side adversarial learning (denoted as ssad), the asymmetric triplet loss (denoted as triplet), and the feature and weight normalization (denoted as norm). The comparison results are shown in Table 1.

It can be seen that the performances of the proposed method with different architectures both degrade if any component is removed. The comparison results verify that each component of SSDG contributes to performance improvement and the incorporation of all these components can achieve the best results.

4.2.2 Comparisons with the Baseline Method

We further compare the SSDG method with the corresponding baseline method, which aims to seek a generalized feature space for both the real and the fake faces. Specifically, we add another domain discriminator after the feature generator to perform adversarial learning on both the real and the fake features. Moreover, the proposed asymmetric triplet loss is replaced with a two-category triplet loss to aggregate all the fake faces as well as the real ones together. Two different network architectures are adopted in the baseline method for more comparisons (denoted as BDG-M and BDG-R, respectively). The comparison results are shown in Table 2.

Firstly, it can be seen that the performance of BDG-M method is comparable with that of the state-of-the-art MADDG method [33] shown in Table 5. The average HTER results of them for all total testing tasks are 23.09% and 23.05%, respectively. This is because the above two methods both aim to seek a generalized feature space not only for the real faces but also for the fake ones. In contrast, our SSDG method outperforms the BDG method as well as the MADDG method on all testing tasks with different network architectures, which demonstrates that seeking a generalized feature space for fake faces is sub-optimal. As

Popular evaluation Metric
- Half total Error Rate
- Area under Cerve
- Attack presentation classification Error Rate
- Bona fide Presentation CER

Table 3. Comparison results of two different network architectures of the feature generator.

Backbones	Flops(G)	Params(M)	Speed(FPS)	Avg HTER(%)	Avg AUC(%)
MADDG-based	47.59	**3.35**	36.40	20.79	88.09
ResNet18-based	**2.38**	11.18	**149.15**	**11.29**	**93.81**

Table 4. Comparison results of domain generalization with limited source domains for face anti-spoofing.

Method	M&I to C		M&I to O	
	HTER	AUC	HTER	AUC
MS-LBP [26]	51.16	52.09	43.63	58.07
IDA [39]	45.16	58.80	54.52	42.17
CT [6]	55.17	46.89	53.31	45.16
LBP-TOP [9]	45.27	54.88	47.26	50.21
MADDG [33]	41.02	64.33	39.35	65.10
SSDG-M	**31.89**	**71.29**	**36.01**	**66.88**

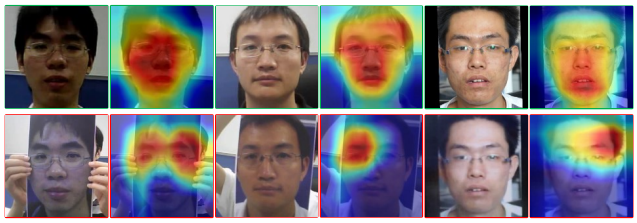


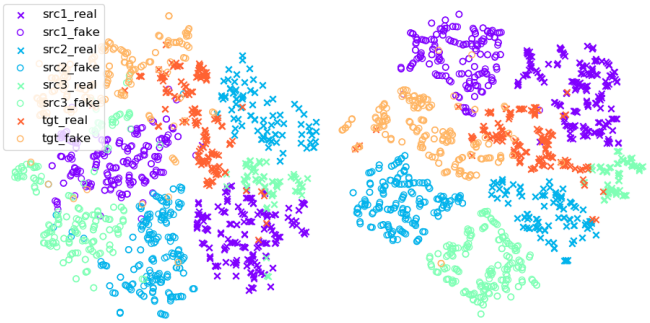
Figure 4. Grad-CAM [32] visualizations of the SSDG method under O&M&I to C. The first row shows the real faces and the second shows the fake ones.

a result, it is more feasible for the face anti-spoofing task to apply asymmetric optimization goals for the real and the fake faces, which can get a class boundary, generalizing better to unseen domains.

4.2.3 Visualizations of the Proposed Method

As shown in Figure 4, we adopt the Grad-CAM [32] to provide the class activation map (CAM) visualizations of our method. It shows that the SSDG method always focuses on the region of the internal face to seek discriminative cues instead of the domain-specific backgrounds, illuminations, etc., which is more likely to generalize well to unseen domains. Specifically, for the fake faces, our method can pay attention to different regions according to different attacks, such as the eyes region of the face for the cut attack.

Moreover, as shown in Figure 5, we randomly select 200 samples of each category from four databases and plot the t-SNE [25] visualizations to analyze the feature space learned by the SSDG method and the corresponding baseline BDG method. It can be seen that the SSDG method can make the features of fake faces more dispersed in the feature space compared to those of the BDG method. In contrast, the feature distribution of the real faces is more compact. Therefore, a better class boundary can be achieved by the SSDG



(a) BDG-R (b) SSDG-R

Figure 5. The t-SNE [25] visualizations of the extracted features by the BDG-R method (a) and the SSDG-R method (b) under the O&M&I to C testing tasks (best viewed in color).

method, which generalizes well to the target domain.

4.2.4 Limited Source Domains

We also evaluate our method when extremely limited source domains are available (*i.e.*, only two source databases). Specifically, MSU and Idiap databases are selected as the source domains for training and the remaining two, *i.e.*, CASIA and OULU, respectively, are used as the target domains for testing. As shown in Table 4, our proposed method achieves the best performance, which has a significant improvement over other methods. Although only two source domains are available, the SSDG method can still force the features of fake faces to be dispersed in the feature space, which promotes to learn a more generalized class boundary for unseen domains.

4.2.5 Comparisons of Different Architectures

As shown in Table 3, we also compare two different architectures of the feature generator, *i.e.*, MADDG-based and ResNet18-based networks, respectively, to evaluate the effects of different backbones. Specifically, the Avg HTER and AUC represent the average results of four testing tasks. And the inference speed of each architecture is tested on the OULU database on a single NVIDIA TITAN 1080 GPU with 256×256 image resolution. It can be seen that the ResNet18-based network is more suitable than the MADDG-based one for face anti-spoofing not only in terms of accuracy but also in terms of speed. And we believe that much better performance can be achieved by the SSDG method with more effective networks.

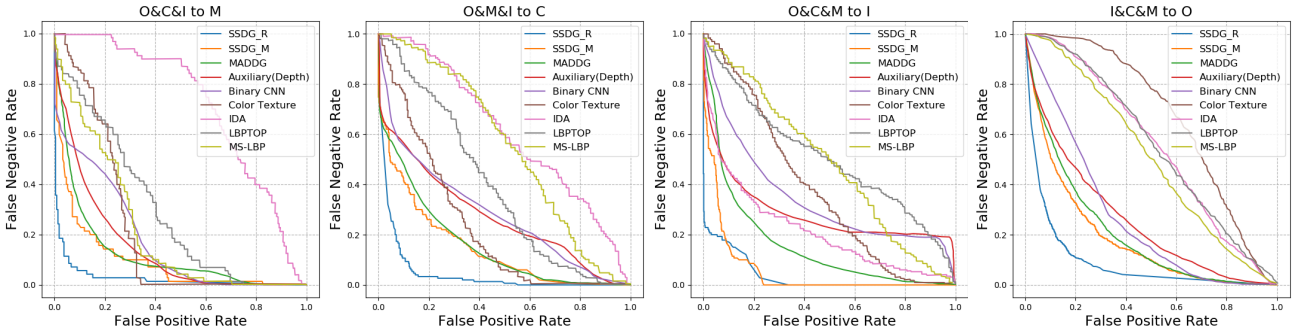


Figure 6. ROC curves of four testing tasks for domain generalization on face anti-spoofing.

Table 5. Comparison results between the proposed method and state-of-the-art methods for domain generalization on face anti-spoofing.

Method	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(%)
MS-LBP [26]	29.76	78.50	54.28	44.98	50.30	51.64	50.29	49.31
Binary CNN [42]	29.25	82.87	34.88	71.94	34.47	65.88	29.61	77.54
IDA [39]	66.67	27.86	55.17	39.05	28.35	78.25	54.20	44.59
Color Texture [6]	28.09	78.47	30.58	76.89	40.40	62.78	63.59	32.71
LBP-TOP [9]	36.90	70.80	42.60	61.05	49.45	49.54	53.15	44.09
Auxiliary (Depth)	22.72	85.88	33.52	73.15	29.14	71.69	30.17	77.61
Auxiliary [23]	-	-	28.40	-	27.60	-	-	-
MADDG [33]	17.69	88.06	24.50	84.51	22.19	84.99	27.89	80.02
SSDG-M	**16.67**	**90.47**	**23.11**	**85.45**	**18.21**	**94.61**	**25.17**	**81.83**
SSDG-R	**7.38**	**97.17**	**10.44**	**95.94**	**11.71**	**96.59**	**15.61**	**91.54**

4.3. Comparison with State-of-the-art Methods

As shown in Table 5 and Figure 6, our method outperforms all the state-of-the-art methods under four testing tasks, which demonstrates the effectiveness of the SSDG method. This is because all other face anti-spoofing methods [6, 9, 23, 26, 39, 42] except for the MADDG [33] method pay no attention to the intrinsic distribution relationship among different domains. Therefore, only database-biased features can be extracted, which causes significant performance degradation in case of cross-database testing scenarios. Although the MADDG method exploits the DG approach to extract common discriminative cues, the results show that seeking a generalized feature space for both the real and the fake faces is difficult to optimize, usually leading to a sub-optimal solution. Due to the diversity of attack types and database collection ways, the extracted features of fake faces are more widely distributed in the feature space than those of real ones, making it nontrivial to aggregate all of them from different domains together. Therefore, our SSDG method applies asymmetric optimization goals for the real and the fake faces to learn a more generalized feature space. Moreover, it shows that when we resort to using the ResNet18-based network, a significant improvement can be made, indicating that better performance can be achieved when the SSDG approach is combined with a more effective network.

4.4. Conclusion

To improve the generalization ability of face anti-spoofing, we propose a novel end-to-end single-side domain generalization framework. Our SSDG learns a generalized feature space, where the feature distribution of real faces is compact while that of fake ones is dispersed across domains. This is quite different from existing methods treating both real and fake faces symmetrically. To achieve this "single-side" goal, the single-side adversarial learning and the asymmetric triplet loss are designed to train the model aggregating the real faces and separating the fake ones from different domains. Extensive experiments show that our SSDG is effective and achieves state-of-the-art results on four public databases. In summary, our work implies that the distribution of real faces and that of fake ones are indeed different, and thus suggests that treating them asymmetrically can lead to better generalization ability to unseen domains. Other possible asymmetric design can be further explored in the future, for instance, dividing the fake faces according to the attack types rather than the databases.

Acknowledgements

This work is partially supported by National Key R&D Program of China (No. 2017YFA0700800) and Natural Science Foundation of China (Nos. 61806188, 61772496).

References

[1] Kei Akuzawa, Yusuke Iwasawa, and Yutaka Matsuo. Adversarial invariant feature learning with accuracy constraint for domain generalization. *arXiv preprint arXiv:1904.12543*, 2019.

[2] André Anjos, Murali Mohan Chakka, and Sébastien Marcel. Motion-based counter-measures to photo attacks in face recognition. *Biometrics IET*, pages 147–158, 2013.

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

[4] Wei Bao, Hong Li, Nan Li, and Wei Jiang. A liveness detection method for face recognition based on optical flow field. In *International Conference on Image Analysis and Signal Processing (IASP)*, pages 233–236, 2009.

[5] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face antispoofing using speeded-up robust features and fisher vector encoding. *Signal Processing Letters (SPL)*, pages 141–145, 2016.

[6] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face spoofing detection using colour texture analysis. *Transactions on Information Forensics and Security (TIFS)*, pages 1818–1830, 2016.

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

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

[9] Tiago de Freitas Pereira, Jukka Komulainen, André Anjos, José Mario De Martino, Abdenour Hadid, Matti Pietikäinen, and Sébastien Marcel. Face liveness detection using dynamic texture. *EURASIP Journal on Image and Video Processing*, page 2, 2014.

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

[11] Javier Galbally, Sébastien Marcel, and Julian Fierrez. Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. *Transactions on Image Processing (TIP)*, pages 710–724, 2013.

[12] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. *arXiv preprint arXiv:1409.7495*, 2014.

[13] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, and David Balduzzi. Domain generalization for object recognition with multi-task autoencoders. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2551–2559, 2015.

[14] Diego Gragnaniello, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. An investigation of local descriptors for biometric spoofing detection. *Transactions on Information Forensics and Security (TIFS)*, pages 849–863, 2015.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

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

[17] Klaus Kollreider, Hartwig Fronthaler, Maycel Isaac Faraj, and Josef Bigun. Real-time face detection and motion analysis with application in liveness assessment. *Transactions on Information Forensics and Security (TIFS)*, pages 548–558, 2007.

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

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

[20] Siqi Liu, Pong C Yuen, Shengping Zhang, and Guoying Zhao. 3d mask face anti-spoofing with remote photoplethysmography. In *European Conference on Computer Vision (ECCV)*, pages 85–100, 2016.

[21] Si-Qi Liu, Xiangyuan Lan, and Pong C Yuen. Remote photoplethysmography correspondence feature for 3d mask face presentation attack detection. In *European Conference on Computer Vision (ECCV)*, pages 558–573, 2018.

[22] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 212–220, 2017.

[23] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 389–398, 2018.

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

[25] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research (JLMR)*, pages 2579–2605, 2008.

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

[27] Saeid Motiian, Marco Piccirilli, Donald A Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5715–5725, 2017.

[28] Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *International Conference on Computer Vision (ICCV)*, pages 1–8, 2007.

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

[30] Yunxiao Qin, Chenxu Zhao, Xiangyu Zhu, Zezheng Wang, Zitong Yu, Tianyu Fu, Feng Zhou, Jingping Shi, and Zhen Lei. Learning meta model for zero-and few-shot face anti-spoofing. *arXiv preprint arXiv:1904.12490*, 2019.

[31] Rajeev Ranjan, Carlos D Castillo, and Rama Chellappa. L2-constrained softmax loss for discriminative face verification. *arXiv preprint arXiv:1703.09507*, 2017.

[32] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.

[33] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C Yuen. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10023–10031, 2019.

[34] Lin Sun, Gang Pan, Zhaohui Wu, and Shihong Lao. Blinking-based live face detection using conditional random fields. In *International Conference on Biometrics (ICB)*, pages 252–260, 2007.

[35] Antonio Torralba, Alexei A Efros, et al. Unbiased look at dataset bias. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, page 7, 2011.

[36] Xiaoguang Tu, Hengsheng Zhang, Mei Xie, Yao Luo, Yuefei Zhang, and Zheng Ma. Deep transfer across domains for face antispoofing. *Journal of Electronic Imaging*, page 043001, 2019.

[37] Feng Wang, Jian Cheng, Weiyang Liu, and Haijun Liu. Additive margin softmax for face verification. *Signal Processing Letters (SPL)*, pages 926–930, 2018.

[38] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Improving cross-database face presentation attack detection via adversarial domain adaptation. *International Conference on Biometrics (ICB)*, pages 1–8, 2019.

[39] Di Wen, Hu Han, and Anil K Jain. Face spoof detection with image distortion analysis. *Transactions on Information Forensics and Security (TIFS)*, pages 746–761, 2015.

[40] Qizhe Xie, Zihang Dai, Yulun Du, Eduard Hovy, and Graham Neubig. Controllable invariance through adversarial feature learning. In *Advances in Neural Information Processing Systems (NIPS)*, pages 585–596, 2017.

[41] Zhenqi Xu, Shan Li, and Weihong Deng. Learning temporal features using lstm-cnn architecture for face anti-spoofing. In *Asian Conference on Pattern Recognition (ACPR)*, pages 141–145, 2015.

[42] Jianwei Yang, Zhen Lei, and Stan Z Li. Learn convolutional neural network for face anti-spoofing. *arXiv preprint arXiv:1408.5601*, 2014.

[43] Xiao Yang, Wenhan Luo, Linchao Bao, Yuan Gao, Dihong Gong, Shibao Zheng, Zhifeng Li, and Wei Liu. Face anti-spoofing: Model matters, so does data. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3507–3516, 2019.

[44] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *Signal Processing Letters (SPL)*, pages 1499–1503, 2016.

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