Adversarial learning and decomposition-based domain generalization for face anti-spoofing

Mingxin Liu[a], Jiong Mu[a,**], Zitong Yu[b,**], Kun Ruan[a], Baiyi Shu[a], Jie Yang[a]

[a] *Sichuan Agricultural University, Ya'an 625000, Sichuan Province, China*
[b] *Center for Machine Vision and Signal Analysis, University of Oulu, Oulu 90014, Finland*

ABSTRACT

Face anti-spoofing (FAS) plays a critical role in the face recognition community for securing the face presentation attacks. Many works have been proposed to regard FAS as a domain generalization problem for robust deployment in real-world scenarios. However, existing methods focus on extracting intrinsic spoofing cues to improve the generalization ability, yet neglect to train a robust classifier. In this paper, we propose a framework to improve the generalization ability of face anti-spoofing in two folds:) a generalized feature space is obtained via aggregation of all live faces while dispersing each domain's spoof faces; and) a domain agnostic classifier is trained through low-rank decomposition. Specifically, a Common Specific Decomposition for Specific (CSD-S) layer is deployed in the last layer of the network to select common features while discarding domain-specific ones among multiple source domains. The above-mentioned two components are integrated into an end-to-end framework, ensuring the generalization ability to unseen scenarios. The extensive experiments demonstrate that the proposed method achieves state-of-the-art results on four public datasets, including CASIA-MFSD, MSU-MFSD, Replay-Attack, and OULU-NPU. *Keywords*: Face anti-spoofing; Domain Generalization; Low-rank Decomposition

1. Introduction

Owing to their convenience and security, face recognition systems have been widely deployed as an efficient and advanced bio-identification technology. Face recognition systems have been applied in payments, entrance guard systems, and municipal public security systems. However, private information (e.g., face images) is easily stolen or leaked in this age of high-developed networks, thus leading to all kinds of presentation attacks (PA) against facial recognition systems. These attacks (e.g., print, video, and 3D mask attack) can easily deceive face recognition systems, thus creating significant and unknown security and property damages. Therefore, face anti-spoofing (FAS) (Yu et al., 2021) (Liu et al., 2021) plays a critical role in the current face recognition fields in both academia and industry.

In recent years, various FAS methods have been proposed and divided into two stages according to the development process: traditional handcrafted feature based and deep learning based. Previous researchers carried out FAS approaches mainly by adopting handcrafted features, such as LBP (Määttä et al., 2011) (de Freitas Pereira et al., 2012) (de Freitas Pereira et al., 2013), HoG (Komulainen et al., 2013) (Yang et al., 2013), and SURF (Boulkenafet et al., 2016a). However, these traditional methods suffer from poor generalization due to the obtained texture information varying with capture devices. Later in the deep learning era, deep learning techniques have been used to extract features and leverage various temporal cues, such as convolutional neural network (CNN) and recurrent neural network (RNN), which have richer semantic information and more robust feature representation than traditional methods (Atoum et al., 2017) (Liu et al., 2018) (Wang et al., 2020b) (Liu et al., 2016).

Although current methods have shown validity and promis-

[**] Corresponding authors: Jiong Mu and Zitong Yu
e-mail: `lmx@stu.sicau.edu.cn` (Mingxin Liu), `jmu@sicau.edu.cn` (Jiong Mu), `zitong.yu@oulu.fi` (Zitong Yu), `qkmc@outlook.com` (Kun Ruan), `shubaiyi100@stu.sicau.edu.cn` (Baiyi Shu), `yangjie1@stu.sicau.edu.cn` (Jie Yang)

ing performance under intra-dataset experiments, they cannot achieve reliable performance on unseen datasets, where models are trained in source domain and tested in target domain on different datasets. The reason behind is that previous methods did not take into account the diversity of feature distribution in different datasets. As a result, captured spoofing cues are dataset-biased (Torralba and Efros, 2011) and cannot generalize well to unseen domains (caused by different materials of attacks or recording environments).

Considerable effort has been made by adapting domain adaption (DA) techniques to align the feature distribution between source and target domain. In (Li et al., 2018), Li et al. proposed an unsupervised DA framework, where the classifier for target domain is trained on the basis of a different source domain. However, such kinds of methods require a large amount of unlabeled target data for training to satisfy the final test, which is expensive and challenging. To make the research on FAS more valuable for practical applications, this paper considers FAS as a domain generalization (DG) problem.

Classical DG methods (Shao et al., 2019) (Jia et al., 2020) (Piratla et al., 2020) (Muandet et al., 2013) (Li et al., 2017) aim to learn a feature representation on multiple source domains. By doing so, the method can be generalized to a target domain in which the data are not used in training. In (Shao et al., 2019), Shao et al. sought to learn a generalized feature space that is discriminative and shared by multiple source domains via a multi-adversarial framework. Jia at el. proposed a novel single-side adversarial learning method in (Jia et al., 2020), where the feature distribution of fake samples is induced to be dispersed, but a reverse operation is performed on real samples. Although these approaches have demonstrated remarkable improvement on the generalization ability for FAS, they only focus on obtaining a universal feature space across domains with a simple binary classifier as output layer, thus neglecting to obtain a robust classifier across domains.

In order to improve the generalization ability, we propose a novel framework for FAS. To obtain a generalized feature space on existing source domains for unseen domains, we follow and extend the preliminary method (Jia et al., 2020), which compresses all real samples and disperses fake ones (for source data) and has shown outstanding testing performance. Compared to (Jia et al., 2020), our proposed method not only focus on domain-invariant features representation but also seeking a domain-agnostic classifier.

Our work is inspired by (Piratla et al., 2020) (Li et al., 2017) and based on the assumption that there are common features and domain-specific features in source data, whose correlation with its corresponding label is consistent/inconsistent across domains. Thus, we propose a Common Specific Decomposition for Specific (CSD-S) layer. Our proposed method is different from the early decomposition method (Li et al., 2017), which deals with domain-specific features by adding a matrix of constraint variables but leads to excessive use of domain-specific features in the condition of binary classification tasks. Our method is also different from the method (Piratla et al., 2020), in which they compute losses on the common and domain-specific features and force them to be orthogonal. The orthogonal operation further improves the model's generalization ability to unseen domains when applied in multi-classification tasks, e.g., Hand-written and speech recognition. However, it entirely ignores the critical cues hidden in the domain-specific features needed to be utilized further in intricate face-related tasks. Hence, to better apply low-rank decomposition for the FAS task and effectively explore hidden important clues in the domain-specific components, we add an extra constraint to the domain-specific features into the orthogonal process.

The main contributions of our work are summarized below:

- We propose a novel framework to extract domain-invariant features via adversarial learning, and combine low-rank decomposition methods for face anti-spoofing.
- A CSD-S layer is designed to highlight the discriminative and common features across domain while ignoring domain-specific features. Hence, a domain-agnostic classifier is obtained.
- Comprehensive experimental results show that, the proposed method further improves the model's generalization ability, and achieves state-of-the-art performance on four public FAS databases with DG based evaluation protocol.

2. Related Work

In this section, we review papers in two categories: handcrafted features based and deep learning based methods.

2.1. Face anti-spoofing

Handcrafted Features based FAS. Traditional works focus on revealing textural differences between live samples and PA, specifically utilizing the micro-texture information of images to counter face spoofing. In most prior works, handcrafted features were extracted first by texture descriptors, such as LBP (Määttä et al., 2011) (de Freitas Pereira et al., 2012) (de Freitas Pereira et al., 2013), HOG (Komulainen et al., 2013) (Yang et al., 2013), DoG (Tan et al., 2010) (Peixoto et al., 2011), SIFT (Patel et al., 2016), and SURF (Boulkenafet et al., 2016a). Classifiers (e.g., SVM, LDA) are then adopted to obtain one binary classification result eventually as output to determine whether the input is alive or not. To reduce the influence of various illumination, Boulkenafet et al. proposed a solution that explores different color spaces such as HSV and YCbCr, which discard chrominance information (Boulkenafet et al., 2015) (Boulkenafet et al., 2016b). However, these texture-based methods are not robust enough. They are helpless in handling PA such as 3D masks and replay attacks. As the resolution and quality of existing datasets varies, such methods become less reliable.

Deep Learning based FAS. Compared with traditional handcrafted features, deep learning-based methods (Atoum et al., 2017) (Liu et al., 2018) (Wang et al., 2020b) (Yu et al., 2020) (Yu et al., 2020a) (Guo et al., 2019) (Yang et al., 2019) (Qin et al., 2020) (Yu et al., 2020c) use Deep Neural Networks (DNN) to extract multistage information and discriminative cues between live and spoof samples. Several works have been developed in recent years. For example, Atoum et al. proposed depth information as the discrepancy between live and spoof samples, and designed a multi-task network with fusion local and holistic features (Atoum et al., 2017). Moreover, Liu et al., deployed a CNN–RNN architecture supervised by auxiliary information (i.e., depth map and rPPG signal) separately, which achieved outstanding performance in PAD (Liu et al., 2018). Recently, Wang et al. used multiple frames to detect PA from two aspects, where discriminative cues came from spatial gradient magnitude and dynamic faces (Wang et al., 2020b). In (Yu et al., 2020) (Yu et al., 2020b), Yu et al. proposed central difference convolutional networks that captures detailed instinct patterns via aggregating both intensity and gradient information. Furthermore, they dug deeper by seeing FAS as a material recognition problem for getting intrinsic material-based patterns (Yu et al., 2020a). In addition to leverage auxiliary information, a specific trend improves the accuracy and generalization ability of a model by synthesizing train data (Guo et al., 2019; Yang et al., 2019). In (Guo et al., 2019), Guo et al. trained CNN from numerous synthetic spoof samples, whereas Yang et al. collected data to simulate real-life scenarios (Yang et al., 2019). Another train of thought relies on meta-learning to solve the overfitting and poor generalization problem of the FAS methods. Qin et al. proposed training a meta-learner to detect unseen spoofing types by learning from predefined real and spoofing faces and a few examples of new attacks (Qin et al., 2020). Yu et al. proposed NAS-FAS, which utilized meta neural architecture search to discover the well-suitable networks with strong domain generalization capacity (Yu et al., 2020). Orthogonal to NAS-FAS focusing on architecture design, our work pays more attentions on efficient learning strategies to enhance domain generalization capacity.

Although traditional and deep learning methods have gained remarkable results, their performance may severely drop under unseen scenarios.

2.2. Domain Generalization

Domain Generalization is an area that mining the consistent relationship between data and their corresponding label on multiple source domains without accessing any target data. Through its unique training strategy, it has promising performance in improving the generalization ability of FAS models.

In (Muandet et al., 2013), Muandet et al. proposed an algorithm named Domain-invariant Component Analysis (DICA) to learn an invariant transformation by minimizing dissimilarity across domains. However, conventional DG such as DICA may suffer from overfitting to seen source domains. To obtain better generalization performance, adversarial learning-based methods (Li et al., 2018) have been designed and have achieved remarkable performance. In (Li et al., 2018), Li et al. aligned distributions among source domains and then matched the aligned distributions via adversarial feature learning. Unlike the adversarial learning methods that focus on learning generalized feature space, decomposition-based domain generalization methods (Piratla et al., 2020) (Li et al., 2017) seek a domain-agnostic classifier by dividing network parameters into common parameters and domain-specific parameters in training. By retaining the common parts of a network and removing domain-specific parts, not only a lightweight face anti-spoofing model is obtained, but also a robust classifier. Adversarial learning and decomposition-based methods deal with DG from the feature space and the classifier, respectively, and both have shown effectiveness. Following these ideas, our work attempts to combine these two methods to obtain a generalized feature space as well as a robust classifier by using low-rank decomposition.

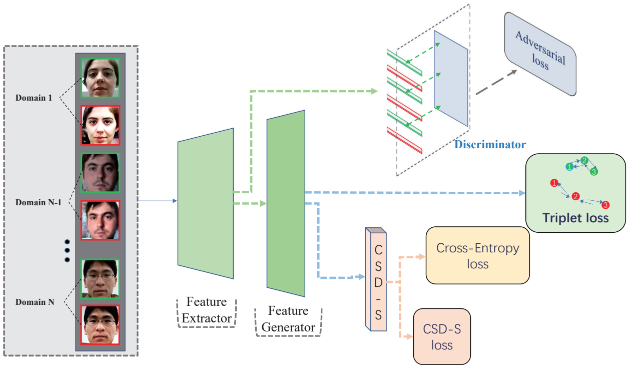


Fig. 1. Overview of our proposed method. The facial images are processed first using a feature extractor. Subsequently, adversarial learning and asymmetric triplet loss are conducted to the extracted features. The CSD-S layer is deployed in the last layer of the proposed network to send the common component of features into the classifier. A generalized feature space and a domain-agnostic classifier are obtained eventually.

3. Proposed Method

3.1. Overview

Figure 1 shows that our overall framework for learning generalized feature space and a robust classifier. After extracting deep features via the feature generator, we compress all live samples and disperse spoof ones of each domain through adversarial learning and Triplet loss. Given that the real images of all domains are collected by simulating real people, obtaining a compact feature space for them is easy. In comparison, the spoof images of each domain vary in many aspects. Thus, to obtain a discriminative class boundary, the opposite operation is conducted on each domain's attack images. By using low-rank decomposition theory, we add a CSD-S block in the last layer of the network. The proposed CSD-S block enhances the generalization ability of the model while pruning it by decomposing features into common parts and domain-specific parts. Domain-invariant feature space and a robust classifier are obtained eventually.

3.2. Singel-side Adversarial Learning and Asymmetric Triplet Mining

In the DG problem of face anti-spoofing, adversarial learning and feature clustering are used to align the feature distribution of given inputs among source domains. However, live and spoof faces are treated equally, which has been proven not the best solution (Jia et al., 2020). Real images are collected by simulating real people. Hence, the distribution discrepancies of real images among multiple source domains may be smaller than the distribution discrepancy of spoof ones. Such differences may be because the spoof faces of each source domain are different in many ways, for example, background, light condition, angle of capture, and attack type. Their feature distribution discrepancies are more significant than that of real ones. Therefore, obtaining a generalized feature space for a real image and an attack image is difficult. For this reason, gathering all live faces and separating attack images across domains is feasible. Consequently, in this paper, single-side adversarial learning and asymmetric triplet mining are conducted to create a discriminative classification boundary that aims to obtain a generalized feature space.

Single-side Adversarial Learning. Assume that K source domains are denoted as $D = \{D_1, D_2, D_3, \ldots, D_{K-1}, D_K\}$. Each domain contains N-labeled instances $\{x_i, y_i\}_{i=1}^{N}$, in which x_i is the input image, and $y_i = 0/1$ is the corresponding label of an input image, that is, the attack image and the real image are labeled with 0 and 1, respectively. To apply single-side adversarial learning, we divide every source domain into two categories: real images X_r and attack images X_a. Then we send them into feature extractor to obtain the corresponding feature as follows:

$$f_r = E\left(X_r\right), f_a = E\left(X_a\right), \tag{1}$$

where E is the shared feature extractor for X_r and X_a. f_r and f_a are the corresponding features extracted by E. D stands for the domain discriminator, against the feature extractor (i.e., extracted feature X_r). D's objective is to identify the source domain of the given features. In contrast, feature extractor E is made to spoof the domain discriminator. Thus, a single-side adversarial learning procedure is designed only for real images to obtain a compact feature space. The objective function can be formulated as:

$$\begin{aligned} \min_{E} \max_{D} \mathcal{L}_{Al}\left(E, D\right) = \\ - \mathbb{E}_{x,y \sim X_r, Y_D} \textstyle\sum_{n=1}^{K} \mathbb{1}_{[n=y]} \log D\left[E\left(x\right)\right]. \end{aligned} \tag{2}$$

Y_D represents the set of domain labels. The function minimizes the loss of feature extractor to optimize its parameters and applies the reverse operation to the loss of feature discriminator, providing a generalized feature space for real images. A gradient reverse layer (GRL) (Ganin and Lempitsky, 2015) is deployed behind the feature extractor in order to optimize the feature extractor and the domain discriminator simultaneously.

Asymmetric Triplet Mining. To obtain a generalized feature space, we implement the concept of gathering all real images and separating attack ones from different domains by applying asymmetric triplet loss. After the asymmetric mining procedure, the features of all real images become compact, and the features of attack ones become increasingly dispersed. The optimized function is as follows:

$$\begin{aligned} \min_{E} \mathcal{L}_{At}\left(E\right) = \sum_{x_i^a, x_i^p, x_i^n} \Big[\left\| E\left(x_i^a\right) - E\left(x_i^p\right) \right\|_2^2 \\ - \left\| E\left(x_i^a\right) - E\left(x_i^n\right) \right\|_2^2 + \alpha \Big]. \end{aligned} \tag{3}$$

For the i-th domain, we randomly select an image that is extracted by the feature extractor and marked as anchor x_i^a. x_i^p represents a positive example with the same label as x_i^a, whereas a negative example x_i^n has an opposite label. $\alpha$is a pre-defined hyperparameter.

3.3. Common Specific Decomposition for Specific Layer

To improve the generalization ability for FAS tasks, extracting common features with a consistent relationship across multiple source domains is important and necessary. The features extracted from the feature extractor can be seen as a combination of parameters that are divided into common component f_c and domain-specific component f_s, which are represented in Eq. (4)

$$x = y\left(f_c + \beta_i f_s\right) + \mathcal{N}\left(0, \Sigma_K\right) \in \mathbb{R}^m, \forall K \in [D], \tag{4}$$

where β_i is a coefficient of f_s, which varies from domain to domain, and $f_c \perp \rho f_s.\rho$ is a constraint variable of f_s. $\mathcal{N}(0, \Sigma_K)$ represents a standard normal random variable with mean zero and covariance matrix Σ_K. m is the dimension of feature space.

Therefore, for each domain K, a good domain specific classifier w_K exists, as follows:

$$\tilde{w}_K = f_c + \gamma_K f_s, \tag{5}$$

$\gamma_K \in \mathbb{R}^m$is a combination of the domain specific components coefficient β_i for $i = 1, 2, \ldots, N$. Each domain's specific classifier can be composed by f_c and f_s. Accordingly, we extract common component f_c and discard the domain-specific ones for all of these source domains to obtain a common classifier w_c. After conducting low-rank decomposition for the sets of all common classifiers and domain-specific ones, we obtain the following formula:

$$W = w_c \mathbb{1}^\top + W_s \Gamma^\top, \tag{6}$$

where $W := [\tilde{w}_1 \tilde{w}_2 \tilde{w}_3 \ldots \tilde{w}_K]$, $\mathbb{1} \in \mathbb{R}^m$ is the all ones vector, $W_s := [f_{s1} f_{s2} \ldots f_{sK}]$, and $\Gamma^\top = [\gamma_1 \gamma_2 \gamma_3 \ldots \gamma_K]$ is the corresponding coefficient matrix of domain specific components W_s.

Therefore, the objective function can be summarized as follows:

$$arg \min_{f} \frac{1}{K} \sum_{j=1}^{K} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\hat{y}_i^{(j)}, y_i^{(j)}\right), \tag{7}$$

where N is the number of instances in Kth domain, $\hat{y}_i = E\left(x_i \mid f_i\right)$. $\hat{y}_i$ is the predicted label of x_i. f_i is the parameter of x_i in this function. f is the combination of common parameters f_c and the domain-specific one f_s. Regarding f as the feature is convenient, and K is the number of source domains.

3.4. Loss Functions

Adversarial learning between domain discriminator and real samples of source domains gathers real samples, similar to asymmetric triplet loss on real and fake ones. Subsequently, the low-rank decomposition-based CSD-S layer extracts the common component among source domains and then sends it into the classifier optimized by the standard cross-entropy loss who is denoted as $\mathcal{L}_{Cls}$. Meanwhile, the common component f_c and domain-specific one f_s are forced to be weak orthogonal to further extract the common features among source domains. The low-rank decomposition process is optimized by $\mathcal{L}_{CSD-S}$. The

Table 1. **Evaluations of different components of the proposed model in domain generalization FAS tasks on four testing sets.**

Baseline	SSA	CSD-S	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(%)
✓			12.74	93.88	13.79	93.11	15.86	92.45	19.98	87.97
✓	✓		7.86	96.29	13.11	93.52	12.42	94.58	16.22	91.19
✓		✓	8.33	95.52	14.00	92.45	14.79	92.68	18.79	88.68
✓	✓	✓	5.00	97.58	10	96.85	12.07	94.68	13.45	94.43

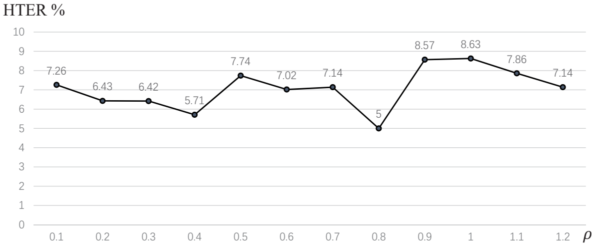


Fig. 2. The performance of the model, with different proportions of the common and domain-specific features, under O&C&I to M testing tasks. Where ρ is a constraint parameter of the domain-specific component.

integration of all these points leads to the overall loss (objective) of our proposed work:

$$\mathcal{L}_{All} = \mathcal{L}_{Cls} + \lambda_1\mathcal{L}_{Al} + \lambda_2\mathcal{L}_{At} + \lambda_3\mathcal{L}_{CSD-S}, \tag{8}$$

where λ_1, λ_2 and λ_3 are the constant constraint parameters of $\mathcal{L}_{Al}$, $\mathcal{L}_{At}$ and $\mathcal{L}_{CSD-S}$, respectively.

4. Experiments

4.1. Dataset and Metrics

Four databases—CASIA-MFSD (Zhang et al., 2012), MSU-MFSD (Wen et al., 2015), Idiap Replay-Attack (Chingovska et al., 2012), and OULU-NPU (Boulkenafet et al., 2017)—were used to evaluate the proposed method. For convenience, they are denoted as C, M, I, and O, respectively. CASIA contains 50 subjects, with three face PA types and a total video number of 700 in three different resolutions and luminous environment. MSU contains 55 subjects with 2 PA types, which are printed, replayed, and captured under two camera devices. Replay-Attack includes 50 subjects in natural light and near-infrared illumination conditions. OULU has 55 subjects with a total video number of 4,950 and 2 PA types. Moreover, OULU has relatively comprehensive scenarios. These four databases contain print attacks and video attacks, but each attack is unique in terms of materials, illumination, background, image capturing and display devices, and resolution. Hence, significant domain shifts are found among these databases.

Following the previous work (Shao et al., 2019), we choose one as the target domain in the four databases, and the remaining three are used for training. Besides, we adopt the half total error rate (HTER) (Bengio and Mariéthoz, 2004) (half of the summation of false acceptance rate and false rejection rate) and the area under curve (AUC) as evaluation metrics.

4.2. Implementation Details

First, we extract facial images with a size of 256 × 256 from existing databases (for each video, we select one frame randomly) by using MTCNN (Zhang et al., 2016) as the input of our network. Similar to (Jia et al., 2020), we only consider RGB channels. Furthermore, we conduct our framework on Pytorch with Restnet18 (He et al., 2016) as a feature generator. We impose the CSD-S layer behind the last convolution layer of Restnet18. The proposed CSD-S layer acts on the features extracted by the feature generator, which decomposes them into common features and domain-specific features. Orthogonality is imposed on the common part and the domain-specific part, in which the domain-specific one is restricted by hyperparameter ρ. Eventually, the optimized common features are sent into the classifier. Thus, an efficient and relatively lightweight network is obtained with $11.44M$ **parameters** and $3.64G$ **FLOPs**. Besides, L2 normalization is adopted on feature and weight to improve the generalization ability during training.

4.3. Ablation Study

In this subsection, we conduct a detailed ablation study to evaluate the performance using different combinations of each component, i.e., the single-side adversarial learning(denoted as SSA) and CSD-S layer. The experimental results on four public FAS datasets are shown in Table 1.

Impact of SSA and CSD-S. For demonstrating the effectiveness of the proposed components, our baseline is comprised of a Resnet-18 backbone and a simple binary classifier with asymmetric triplet mining applied. As shown in Table 1, the results are consistently improved when adding SSA into the model on four testing tasks. Thus a generalized feature space is obtained. Additionally deploying the CSD-S with the baseline, the results are again improved. In contrast, the model that combines baseline and SSA provides slighter better results than the one using baseline and the CSD-S. Finally, the results are improved further when applying all the components. This observation verifies that significant effects are made by deploying the CSD-S to obtain a domain-agnostic classifier under the premise of applying SSA to get a generalized feature space.

Impact of Hyperparameter ρ. In order to evaluate how the hyperparameter ρ influences CSD-S layer, we evaluate the results of the common component and domain-specific one with different proportions obtained in the decomposition process. Specifically, we adjust the constraint of the domain-specific one (i.e., ρ) to get optimal decomposition results in the orthogonal process. As shown in Figure 2, the performance obtained by different proportions of the common features and domain-specific one in the proposed model is varied, and the model gains the maximum impact when $\rho = 0.8$.

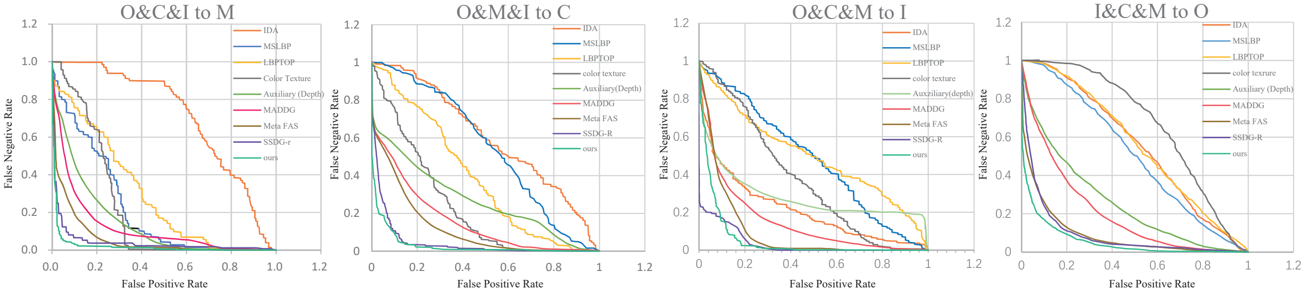


Fig. 3. ROC curves of our method in four testing tasks.

Table 2. Comparison results of our method and the state-of-the-art DG methods 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	29.76	78.5	54.28	44.98	55.3	51.64	50.29	49.31
IDA	66.67	27.86	55.17	39.05	28.35	78.25	54.2	44.59
Color Texture	28.09	78.47	30.58	76.89	40.4	62.78	63.59	32.71
LBP-TOP	36.9	70.8	42.6	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	-	-	28.4	-	27.6	-	-	-
MADDG	17.69	88.06	24.5	84.51	22.19	84.99	27.89	80.02
Cross-domain PAD	17.02	90.1	19.68	87.43	20.87	86.72	25.02	81.47
Meta FAS	13.89	93.98	20.27	88.16	17.3	90.48	16.45	91.16
SSDG-R	7.38	97.17	10.44	95.94	11.71	96.59	15.61	91.54
NAS-FAS	16.85	90.42	15.21	92.64	**11.63**	**96.98**	13.16	94.18
Ours	**5.00**	**97.58**	**10.00**	**96.85**	12.07	94.68	**13.45**	**94.43**

4.4. Comparison with the State-of-the-arts Methods

We evaluate our results by conducting a comparison with a number of state-of-the-art methods: multi-scale LBP (MS-LBP) (Määttä et al., 2011); image distortion analysis (Wen et al., 2015); color texture (Boulkenafet et al., 2016b); LBP-TOP (de Freitas Pereira et al., 2014); auxiliary (Liu et al., 2018); MADDG (Shao et al., 2019); Cross-domain PAD (Wang et al., 2020a); Meta FAS (Shao et al., 2020) and SSDG-R (Jia et al., 2020). As shown in Table 2 and Figure 3, our method outperforms all the state-of-art methods proposed for FAS under three DG testing sub-protocols, because previous methods donot take into account the distribution relationship among different domains (Määttä et al., 2011) (Liu et al., 2018) (Boulkenafet et al., 2016b) (Wen et al., 2015) (de Freitas Pereira et al., 2014). Therefore, their performance drops severely in cross-database experiments. Adversarial learning was used to learn shared and discriminative cues among multiple source domains in (Shao et al., 2019) (Jia et al., 2020). Despite remarkable achievements in improving the generalization ability for face anti-spoofing like SSDG-R method which compacted all real samples and dispersed attack ones of each domain, the state-of-art methods only focus on obtaining a generalized feature space among source domains. By contrast, our method tries to extract domain-invariant features and trains a domain-agnostic classifier. Compared with the state-of-the-art searched generalized architecture in NAS-FAS (Yu et al., 2020), the proposed method based on simple ResNet18 backbone achieves better performance on ‘O&C&I to M’, ‘O&M&I to C’, and ‘I&C&M to O’, indicating the excellent generalization ability with domain adversarial learning and CSD-S. The theory of learning generalized feature space and a domain agnostic classifier is more effective than solely focusing on one.

4.5. Visualization

As shown in Fig.4, the Grad-CAM (Selvaraju et al., 2017) visualization is conducted on different network architectures to provide the class activation map (CAM) under I&M&O to C. As can be seen that the network with SSA and our proposed network focus on the region of internal face, instead of concentrating on the domain-specific information (i.e. the edge area outside the face, such as backgrounds, illuminations, etc.) like the baseline according to Fig.4. It shows that the network with SSA and the proposed network have a better chance of learning robust discriminant clues. Compared with SSA, our proposed method with the CSD-S focuses more on facial areas with rich discriminant clues across domains such as eyes, nose, cheeks, etc., which is more likely to generalize well to unseen domains.

Moreover, as shown in Fig.5, the t-SNE (Van der Maaten and Hinton, 2008) visualizations are plotted to analyze the feature space learned by our proposed method and the Resnet-18. Both experiments were completed under O&C&I to M testing tasks with 200 samples of each category from four databases. It is worth noting that the real features and attack features are separated in Fig.5(a) by a rough and straightforward classification boundary with an unsatisfied performance. In contrast, as can be seen in Fig.5(b), the feature distribution of attack images is dispersed domain to domain, and the distribution of real one is more compact. Therefore, not only a better class boundary can be achieved but also far better generalization ability is gained by our proposed method when testing in Fig.5(c) compared to the performance in Fig.5(a).

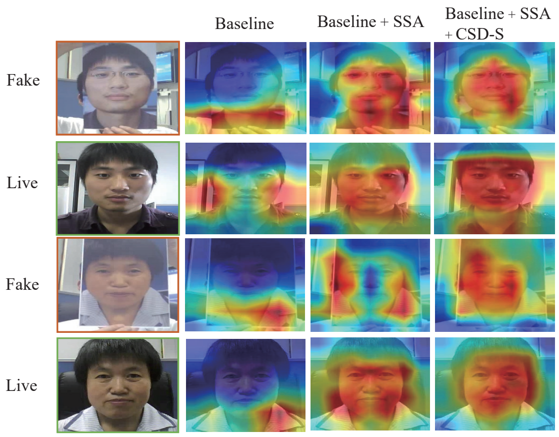
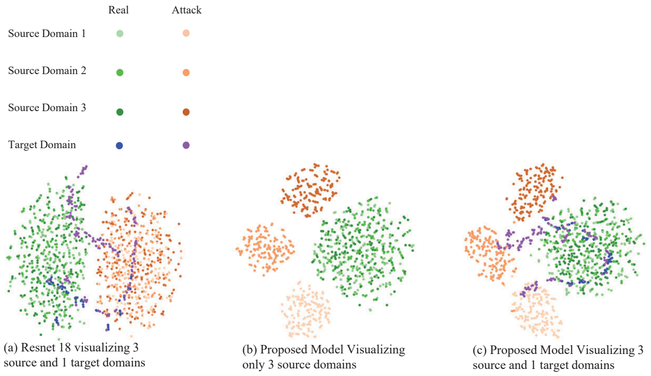


Fig. 4. The Grad-CAM visualization of different methods under I&C&O to M, where the first and second rows show the fake and live samples, respectively. The methods from left to right are the pre-trained Resnet18 (i.e., baseline), the network with SSA, and our proposed method with CSD-S.



(a) Resnet 18 visualizing 3 source and 1 target domains

(b) Proposed Model Visualizing only 3 source domains

(c) Proposed Model Visualizing 3 source and 1 target domains

Fig. 5. A t-SNE plot of the extracted features coming from Resnet (a) and our full model (b, c) under the O&C&I to M testing tasks (best viewed in colors).

To further evaluate the performance of the model on each type of true and false samples, we listed some of the incorrect FAS results under four experiments. As shown in Fig.6, the first two columns are the real faces which are incorrectly classified as fake ones, while the latter two columns are the opposite. It can be observed from the samples that the majority of misclassification is caused by obvious appearance variances, such as partial occlusion, underexposure or oversaturation, and image distortion, etc., which makes the extracted features differences between these real and fake faces smaller. Therefore, it is very challenging in practical applications of the FAS model to solely rely on learning a generalized feature space for all data, but other efforts can be made for instance the proposed CSD-S which tries to gain a domain-agnostic classifier.

5. Conclusion

In this paper, we proposed a novel framework to extract domain-invariant features via adversarial learning and combined low-rank decomposition in face anti-spoofing. To improve the generalization ability of our model, we tried to seek a generalized feature space by aggregating all live faces and separating spoof ones from different domains. Besides, we proposed a CSD-S layer that decomposes the network's parameters into two parts, thus obtaining a domain-agnostic classifier. Comprehensive experimental results show the effectiveness of combining generalized feature space and robust classifier.

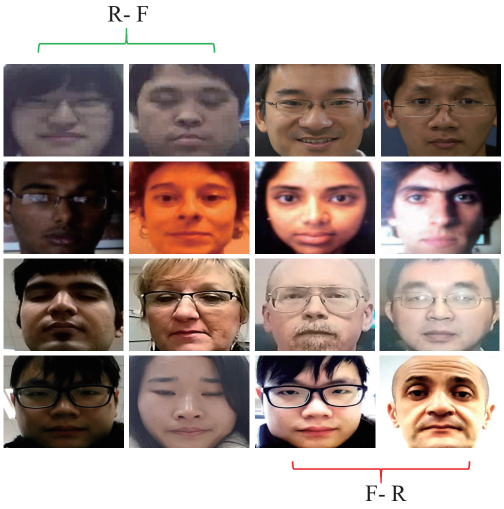


Fig. 6. Examples of the incorrect FAS results by the proposed method under four experiments. The sign "R-F" denotes that a real face is incorrectly classified as a fake one, while the "F-R" is the opposite.

Acknowledgments

This paper was supported by the Key Laboratory of Agricultural Information Engineering of Sichuan Province, Sichuan Agricultural University. This study was financially supported by Sichuan Science and Technology Innovation Seedling Project Funding Project under Grant No.2019025, Sichuan Agricultural University Innovation and Entrepreneurship Training Program (Grant No.201910626026).

References

Atoum, Y., Liu, Y., Jourabloo, A., Liu, X., 2017. Face anti-spoofing using patch and depth-based cnns, in: IJCB, IEEE. pp. 319–328.

Bengio, S., Mariéthoz, J., 2004. A statistical significance test for person authentication, in: Proceedings of Odyssey 2004: The Speaker and Language Recognition Workshop.

Boulkenafet, Z., Komulainen, J., Hadid, A., 2015. Face anti-spoofing based on color texture analysis, in: ICIP, IEEE. pp. 2636–2640.

Boulkenafet, Z., Komulainen, J., Hadid, A., 2016a. Face antispoofing using speeded-up robust features and fisher vector encoding. IEEE SPL 24, 141–145.

Boulkenafet, Z., Komulainen, J., Hadid, A., 2016b. Face spoofing detection using colour texture analysis. TIFS 11, 1818–1830.

Boulkenafet, Z., Komulainen, J., Li, L., Feng, X., Hadid, A., 2017. Oulu-npu: A mobile face presentation attack database with real-world variations, in: FG, IEEE. pp. 612–618.

Chingovska, I., Anjos, A., Marcel, S., 2012. On the effectiveness of local binary patterns in face anti-spoofing, in: BIOSIG, IEEE. pp. 1–7.

de Freitas Pereira, T., Anjos, A., De Martino, J.M., Marcel, S., 2012. Lbp-top based countermeasure against face spoofing attacks, in: ACCV, Springer. pp. 121–132.

de Freitas Pereira, T., Anjos, A., De Martino, J.M., Marcel, S., 2013. Can face anti-spoofing countermeasures work in a real world scenario?, in: ICB, IEEE. pp. 1–8.

de Freitas Pereira, T., Komulainen, J., Anjos, A., De Martino, J.M., Hadid, A., Pietikäinen, M., Marcel, S., 2014. Face liveness detection using dynamic texture. EURASIP Journal on Image and Video Processing 2014, 2.

Ganin, Y., Lempitsky, V., 2015. Unsupervised domain adaptation by backpropagation, in: ICML, pp. 1180–1189.

Guo, J., Zhu, X., Xiao, J., Lei, Z., Wan, G., Li, S.Z., 2019. Improving face anti-spoofing by 3d virtual synthesis, in: ICB, IEEE. pp. 1–8.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: CVPR, pp. 770–778.

Jia, Y., Zhang, J., Shan, S., Chen, X., 2020. Single-side domain generalization for face anti-spoofing, in: CVPR, pp. 8484–8493.

Komulainen, J., Hadid, A., Pietikäinen, M., 2013. Context based face anti-spoofing, in: BTAS, IEEE. pp. 1–8.

Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M., 2017. Deeper, broader and artier domain generalization, in: ICCV, pp. 5542–5550.

Li, H., Li, W., Cao, H., Wang, S., Huang, F., Kot, A.C., 2018. Unsupervised domain adaptation for face anti-spoofing. TIFS 13, 1794–1809.

Liu, A., Li, X., Wan, J., Liang, Y., Escalera, S., Escalante, H.J., Madadi, M., Jin, Y., Wu, Z., Yu, X., et al., 2021. Cross-ethnicity face anti-spoofing recognition challenge: A review. IET Biometrics 10, 24–43.

Liu, S., Yuen, P.C., Zhang, S., Zhao, G., 2016. 3d mask face anti-spoofing with remote photoplethysmography, in: ECCV, Springer. pp. 85–100.

Liu, Y., Jourabloo, A., Liu, X., 2018. Learning deep models for face anti-spoofing: Binary or auxiliary supervision, in: CVPR, pp. 389–398.

Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-sne. Journal of machine learning research 9.

Määttä, J., Hadid, A., Pietikäinen, M., 2011. Face spoofing detection from single images using micro-texture analysis, in: IJCB, IEEE. pp. 1–7.

Muandet, K., Balduzzi, D., Schölkopf, B., 2013. Domain generalization via invariant feature representation, in: ICML, pp. 10–18.

Patel, K., Han, H., Jain, A.K., 2016. Secure face unlock: Spoof detection on smartphones. TIFS 11, 2268–2283.

Peixoto, B., Michelassi, C., Rocha, A., 2011. Face liveness detection under bad illumination conditions, in: ICIP, IEEE. pp. 3557–3560.

Piratla, V., Netrapalli, P., Sarawagi, S., 2020. Efficient domain generalization via common-specific low-rank decomposition, in: International Conference on Machine Learning, PMLR. pp. 7728–7738.

Qin, Y., Zhao, C., Zhu, X., Wang, Z., Yu, Z., Fu, T., Zhou, F., Shi, J., Lei, Z., 2020. Learning meta model for zero-and few-shot face anti-spoofing, in: AAAI.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE international conference on computer vision, pp. 618–626.

Shao, R., Lan, X., Li, J., Yuen, P.C., 2019. Multi-adversarial discriminative deep domain generalization for face presentation attack detection, in: CVPR, pp. 10023–10031.

Shao, R., Lan, X., Yuen, P.C., 2020. Regularized fine-grained meta face anti-spoofing., in: AAAI, pp. 11974–11981.

Tan, X., Li, Y., Liu, J., Jiang, L., 2010. Face liveness detection from a single image with sparse low rank bilinear discriminative model, in: ECCV, Springer. pp. 504–517.

Torralba, A., Efros, A.A., 2011. Unbiased look at dataset bias, in: CVPR, IEEE. pp. 1521–1528.

Wang, G., Han, H., Shan, S., Chen, X., 2020a. Cross-domain face presentation attack detection via multi-domain disentangled representation learning, in: CVPR, pp. 6678–6687.

Wang, Z., Yu, Z., Zhao, C., Zhu, X., Qin, Y., Zhou, Q., Zhou, F., Lei, Z., 2020b. Deep spatial gradient and temporal depth learning for face anti-spoofing, in: CVPR, pp. 5042–5051.

Wen, D., Han, H., Jain, A.K., 2015. Face spoof detection with image distortion analysis. TIFS 10, 746–761.

Yang, J., Lei, Z., Liao, S., Li, S.Z., 2013. Face liveness detection with component dependent descriptor, in: ICB, IEEE. pp. 1–6.

Yang, X., Luo, W., Bao, L., Gao, Y., Gong, D., Zheng, S., Li, Z., Liu, W., 2019. Face anti-spoofing: Model matters, so does data, in: CVPR, pp. 3507–3516.

Yu, Z., Li, X., Niu, X., Shi, J., Zhao, G., 2020a. Face anti-spoofing with human material perception, in: ECCV.

Yu, Z., Li, X., Shi, J., Xia, Z., Zhao, G., 2021. Revisiting pixel-wise supervision for face anti-spoofing. IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM) .

Yu, Z., Qin, Y., Li, X., Wang, Z., Zhao, C., Lei, Z., Zhao, G., 2020b. Multi-modal face anti-spoofing based on central difference networks, in: CVPRW, pp. 650–651.

Yu, Z., Qin, Y., Xu, X., Zhao, C., Wang, Z., Lei, Z., Zhao, G., 2020c. Auto-fas: Searching lightweight networks for face anti-spoofing, in: ICASSP, IEEE. pp. 996–1000.

Yu, Z., Wan, J., Qin, Y., Li, X., Li, S.Z., Zhao, G., 2020. Nas-fas: Static-dynamic central difference network search for face anti-spoofing. IEEE TPAMI , 1–1doi:10.1109/TPAMI.2020.3036338.

Yu, Z., Zhao, C., Wang, Z., Qin, Y., Su, Z., Li, X., Zhou, F., Zhao, G., 2020. Searching central difference convolutional networks for face anti-spoofing, in: CVPR, pp. 5295–5305.

Zhang, K., Zhang, Z., Li, Z., Qiao, Y., 2016. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE SPL 23, 1499–1503.

Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z., 2012. A face antispoofing database with diverse attacks, in: ICB, IEEE. pp. 26–31.