Fourier-Based Frequency Space Disentanglement and Augmentation for Generalizable Face Anti-Spoofing

Yang Yu, Zhekai Du, Heng Luo, Chengwei Xiao, and Jiang Hu

***Abstract*—Generalizing face anti-spoofing (FAS) models to unseen distributions is challenging due to domain shifts. Previous domain generalization (DG) based FAS methods focus on learning invariant features across domains in the spatial space, which may be ineffective in detecting subtle spoof patterns. In this paper, we propose a novel approach called Frequency Space Disentanglement and Augmentation (FSDA) for generalizable FAS. Specifically, we leverage Fourier transformation to analyze face images in the frequency space, where the amplitude spectrum captures low-level texture information that forms distinct visual appearances, and the phase spectrum corresponds to the content information. We hypothesize that the liveness of a face is more related to these low-level patterns rather than high-level content information. To locate spoof traces, we disentangle the amplitude spectrum into domain-related and spoof-related components using either empirical or learnable strategies. We then propose a frequency space augmentation technique that mixes the disentangled components of two images to synthesize new variations. By imposing a distillation loss and a consistency loss on the augmented samples, our model learns to capture spoof patterns that are robust to both domain and spoof type variations. Extensive experiments on four FAS datasets demonstrate the superiority of our method in improving the generalization ability of FAS models in various unseen scenarios.**

***Index Terms*—Face anti-spoofing, domain generalization, Fourier transformation, frequency space augmentation.**

I. Introduction

FACE recognition (FR) [1] plays an important role in many real-world authentication applications such as access control, online payment and other security scenarios. However, the great convenience also raises security concerns in the community since FR systems can be easily deceived by some well-designed presentation attacks (PA). Aiming at securing FR systems from these malicious spoof attacks, face anti-spoofing (FAS) [2], [3], [4] has continuously drawn attention from researchers. FAS has become an increasingly challenging task due to the ever-evolving spoofing techniques and the need for generalization across different domains. In real-world applications, such as biometric systems and authentication processes, a slight increase in detection accuracy or a decrease in false positive rates can lead to significant enhancements in security and user experience.

Methods for PA detection have evolved from traditional liveness detection [5], [6] and texture analysis [7], [8] algorithms based on handcrafted features, to more advanced deep learning-based ones [3], [9] that perform binary or pixel-wise supervision on high-level deep features. Despite the profound improvements achieved in intra-dataset scenarios, generalizing FAS models to out-of-distribution (OOD) data is still a huge challenge for algorithms that strongly rely on the i.i.d assumption. On the other side, different from common classification problems that make predictions on high-level dominating semantic information, the judgment for FAS is the opposite: the key spoof traces often lie in subtle low-level image patterns [10], e.g., face texture, blurring, distortion, moire pattern, and reflection artifacts. As a result, existing FAS research still has two open challenges, i.e., the OOD generalization problem raised by distribution shift among different domains, and the inferior low-level spoof feature learning capacity of deep networks.

For the first concern, considerable efforts have been devoted to improving the generalization ability of FAS models in light of domain generalization (DG) [11], [12]. Concretely, DG-based FAS methods can be generally grouped into three regimes: adversarial learning-based methods [13], [14], disentanglement-based methods [15], [16] and meta-learning-based [17], [18], [19], [20] methods. Although these methods have made impressive progresses in detection accuracy under inter-dataset scenarios, they consistently utilize convolutional neural networks that excavate spoof traces of a face image in the entire spatial (RGB) space, which often fails to show a sufficient discriminative power when the key spoof clue lies in low-level local features.

Inspired by prior studies [10], [21] that analyze texture differences between real and spoof faces in the frequency space, we

Received 25 March 2024; revised 19 May 2024 and 16 June 2024; accepted 17 June 2024. Date of publication 24 June 2024; date of current version 7 August 2025. *(Corresponding author: Jiang Hu.)*

Yang Yu, Chengwei Xiao, and Jiang Hu are with the Department of Orthopaedics, Sichuan Provincial People's Hospital, and School of Medicine, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: hujiang8711@163.com).

Zhekai Du is with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zhekaid@std.uestc.edu.cn).

Heng Luo is with the Second Affiliated Hospital of Chengdu Medical College, China National Nuclear Corporation 416 Hospital, Chengdu 610051, China.

Digital Object Identifier 10.1109/JBHI.2024.3417404

2168-2194 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: UNIVERSITY OF NEVADA RENO. Downloaded on October 22,2026 at 20:18:24 UTC from IEEE Xplore. Restrictions apply.

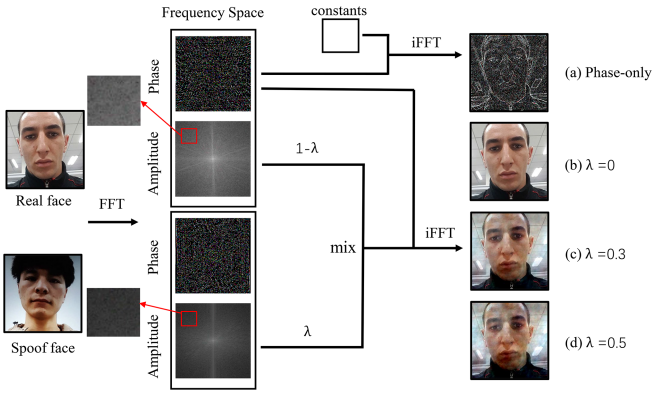


Fig. 1. Reconstruction by Fast Fourier Transformation (FFT) and its inverse (i.e., iFFT). (a) reconstructed image with pure phase information. (b)-(d) reconstructed images when $\lambda = 0, 0.3$ and 0.5, respectively. It shows that spoof face carries less high-frequency information than real face. Gradually mixing their amplitude spectrums makes the appearance of the real face gradually transfer to that of the fake face.

propose Frequency Space Disentanglement and Augmentation (FSDA) that effectively detects and manipulates spoof-related traces based on the Fourier transformation. Our motivation is that the phase component of the Fourier spectrum retains high-level semantic information which defines the visual structure of content objects, while the amplitude component depicts the low-level texture that forms a distinct visual appearance (as shown in Fig. 1). Typically, the liveness of a face image is often less sensitive to content objects (e.g., subjects) but more related to fine-grained texture details. These liveness-related texture information includes subtle patterns such as skin pores, wrinkles, and natural shading, which are characteristic of genuine faces. In contrast, spoof face images often exhibit spoof-related style factors, such as moire patterns, color distortions, or reflections, due to the limitations of spoofing techniques. These style factors contribute to the distinct appearance of spoof faces and are less prominent in real faces. Therefore, we propose to distill spoof-related style factors from the amplitude spectrum that controls the appearance of a face. To distinguish domain-related frequency components from spoof-related ones, we divide the whole amplitude spectrum into multiple blocks and design two alternatives to disentangle them based on spectrum blocks. The first is an empirical method that identifies domain- and spoof-related blocks based on the outputs of the task-specific classifier and the domain discriminator with two binary masks. The second is a learnable method that iteratively learns soft masks by backpropagation in an end-to-end manner. We then conduct a frequency space augmentation by linearly interpolating between the domain-related and spoof-related parts of the amplitude spectrums of two images. A distillation loss and a consistency regularization are then imposed on the augmented samples to make our model discover the liveness-related texture information in the amplitude spectrum rather than the domain-related ones or the task-irrelevant content information in the phase spectrum. Moreover, our model simultaneously learns the continuous variation patterns of the domain and spoof traces without interference, thereby achieving generalization.

In summary, the contributions of this work can be phrased as follows: 1) We propose to explore and manipulate subtle low-level spoof traces for FAS in the frequency space, which is efficiently implemented based on the amplitude component of the Fourier spectrum. 2) We design two schemas to disentangle the whole amplitude spectrum into domain-related and spoof-related frequency components in the unit of spectrum blocks. 3) We conduct frequency space-based augmentation with the disentangled amplitude spectrum. Two regularization terms are imposed to make the model generalizable to unseen domains and spoofs. 4) Our method is validated on four widely used FAS datasets and shows superior results.

II. Related Work

A. Face Anti-Spoofing

The research on FAS is usually carried out along two technical routes, i.e., texture-based methods [22], [23], [24] that distinguish real and spoof faces based on color and texture clues, and temporal-based methods [25], [26] that leverage temporal change in face area for judgment. Traditional methods use various handcrafted features such as LBP [7], SIFT [8] and SURF [23], etc., to detect spoof patterns, which have limited representation capacity and are not robust to variations like illuminations. With the advent of deep learning, various deep networks have been equipped to extract more discriminative features for FAS [9], [27], [28], [29]. However, the data-driven nature makes deep learning-based methods still unable to escape the curse of the i.i.d assumption. Therefore, the accuracy and generalization ability are still unsatisfactory.

Several works have been conducted to generalize FAS models to unseen scenarios by leveraging various DG techniques. As an early work, MADDG [13] learns a common feature space with multiple source domains via a multi-adversarial framework. SSDG [14] points out the uniqueness of real faces and the diversity of spoof faces, conducting adversarial learning only on real faces with a triplet loss. Inspired by meta-learning [30], RFM [18] learns a generic learning direction with a depth regularized meta-learning schema. In addition to meta-learning, D ^{2} AM [17] addresses a domain-agnostic scenario, while SDA [19] utilizes target data for improved adaptation during testing. Apart from them, ANRL [31] and SSAN [32] aim to disentangle domain-related and spoof-related style information using normalization layers and a style transfer framework, respectively. These approaches share some conceptual similarities with our work. However, they both analyze image statistics in the spatial space, which can be less effective when spoof clues are subtle in visual space.

B. Fourier-Based FAS and Transfer Learning

Some texture-based FAS methods have tried to find discriminative features in the frequency space. Li et al. [21] empirically find that real faces have more high-frequency components than spoof ones and propose a high-frequency descriptor based on the Fourier spectrum of a face. Huang et al. [10] later leverage Fourier spectrums in multiple video frames as extra inputs to

Authorized licensed use limited to: UNIVERSITY OF NEVADA RENO. Downloaded on October 22,2026 at 20:18:24 UTC from IEEE Xplore. Restrictions apply.

compensate for the lack of discrimination in spatial domain. However, they do not consider the domain shift problem in FAS, nor make an in-depth investigation of different components in the Fourier spectrum.

Research on psychophysics [33] shows that the phase component of the Fourier spectrum carries high-level semantic information like edges and contours of an object, while the amplitude encodes the low-level style information. Based on that, Yang et al. [34] propose to preserve the phase spectrum when conducting image translation for domain adaptation [35]. A concurrent work [36] aims to achieve domain alignment by replacing low-frequency amplitude components of source images with that of target images. Later, Xu et al. [37] mix the amplitude spectrums of two images for augmentation, and impose a consistency loss on the semantic label to achieve DG. Differently, the liveness label in FAS is determined by the image style rather than the content object. Besides, how to disentangle domain-related and spoof-related styles also remains to be solved in FAS.

It worth noting that a recent FAS work [38] leverages Fourier-based augmentation for source-free domain adaptation. However, it simply adopts vanilla amplitude mix [37] for target data augmentation, and the *de facto* adaptation is achieved by statistics matching in the source-trained model. Besides, fine-grained spoof patterns in the amplitude spectrum still remain underexplored in FAS research.

III. Method

A. Problem Definition

Following DG-based FAS methods, we have access to a collection of labeled face images $\mathcal{D}_s = \{x_i, y_i, d_i\}_{i=1}^{N_s}$ from K distinct domains that conform to different distributions, where $x_i \in [0, 255]^{3\times H\times W}$ is an RGB space image with $H \times W$ the spatial size, $y_i \in \{0, 1\}^2$ is the one-hot binary label, $d_i \in \{0, 1\}^K$ denotes the one-hot domain label, and N_s amounts to the total number of training samples. Our goal is to train an FAS model f that performs well on unseen target samples $\{x_i^t\}_{i=1}^{N_t}$. Typically, the model f can be further decomposed into a feature extractor G and a task-specific classifier C. Besides, we employ a domain discriminator D which is necessary to explore domain-related knowledge.

B. Frequency Space Augmentation for FAS

Many existing DG-based FAS works [13], [14] learn a domain-shared feature space for the intermediate features $f_i = G(x_i)$ with limited source data in hand, hoping that the learned feature space is also applicable to the target domain. However, we argue it is too excessive for the extractor G to find such a mapping function with only several discrete domains, which make the latent feature space not smooth and continuous enough. Therefore, it is a natural idea to increase the diversity of the training set by augmentation, which has been verified effective for generalization [39].

Different from common DG methods [37], [40], [41] that synthesize novel stylized samples for domain augmentation, for generalizable FAS problem we tackle not only the variation of domains but also the variation of spoof traces. Therefore it is necessary to conduct augmentation towards above two directions. A recent work SSAN [32] proposes to transfer liveness-related style information across samples to construct a stylized feature space. However, it highlights the liveness-related style and suppresses domain-related style in the spatial space with a contrastive loss. As we discussed, the low-level style information is hard to be captured in the RGB space. In this work, we aim to find these two kinds of style information in the frequency space. Specifically, we first transform each channel of an RGB image x_i to frequency space by the Fourier transformation $\mathcal{F}(\cdot)$:

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

We use the notation $\mathcal{F}^{-1}(\cdot)$ for the inverse Fourier transformation, and both of them can be effectively calculated by the FFT algorithm [42]. The original Fourier spectrum $\mathcal{F}(x_i) \in \mathbb{R}^{3\times H\times W}$ is in the form of complex numbers and can be further decomposed into an amplitude component $\mathcal{A}(x_i)$ and a phase component $\mathcal{P}(x_i)$ by:

$$\begin{aligned}\mathcal{A}(x_i)(u, v) &= |\mathcal{F}(x_i)(u, v)|,\\ \mathcal{P}(x_i)(u, v) &= \arctan\left[\frac{I(x_i)(u, v)}{R(x_i)(u, v)}\right], \end{aligned} \quad (2)$$

where $R(\cdot)$ and $I(\cdot)$ denotes the real and the imaginary part of $\mathcal{F}(\cdot)$. Research [33] has shown that the phase component mainly carries the high-level semantic information like object edges and contours (we also observed that in Fig. 1(a)), and the amplitude component encodes the low-level style information. Inspired by MixUp-based DG methods [43], [44], [45], we aim to further increase the diversity of the training set by linearly interpolating between the amplitude spectrums of two face images. However, different from [37] that directly mixes the whole amplitude spectrums, we need to deploy different mix schemas for the domain- and spoof-oriented augmentation, respectively. To this end, we propose a masked amplitude mix strategy.

Denoted by $M_D \in \mathbb{R}^{H\times W}$ and $M_S \in \mathbb{R}^{H\times W}$ the masks that indicate whether a frequency component contributes to domain-related and spoof-related appearance style, respectively, we use the following strategy to mix domain- and spoof-related amplitude components between two images:

$$\begin{aligned}\hat{\mathcal{A}}^D(x_i) &= (\mathbf{1} - M_D) \odot \mathcal{A}(x_i) + M_D \odot [(1 - \lambda)\mathcal{A}(x_i)\\ &\quad + \lambda\mathcal{A}(x_i')],\\ \hat{\mathcal{A}}^S(x_i) &= (\mathbf{1} - M_S) \odot \mathcal{A}(x_i) + M_S \odot [(1 - \lambda)\mathcal{A}(x_i)\\ &\quad + \lambda\mathcal{A}(x_i')],\end{aligned}$$

where $\lambda \sim U(0, \alpha)$, (x_i', y_i', d_i') is a randomly sampled reference image paired with x_i. A figurative illustration of the masked amplitude mix is provided in Fig. 2.

We combine the mixed amplitude spectrums with the original phase spectrum to calculate the corresponding Fourier spectrums, respectively:

$$\mathcal{F}^D(x_i)(u, v) = \hat{\mathcal{A}}^D(x_i)(u, v) * e^{-j*\mathcal{P}(x_i)(u,v)},$$

Authorized licensed use limited to: UNIVERSITY OF NEVADA RENO. Downloaded on October 22,2026 at 20:18:24 UTC from IEEE Xplore. Restrictions apply.

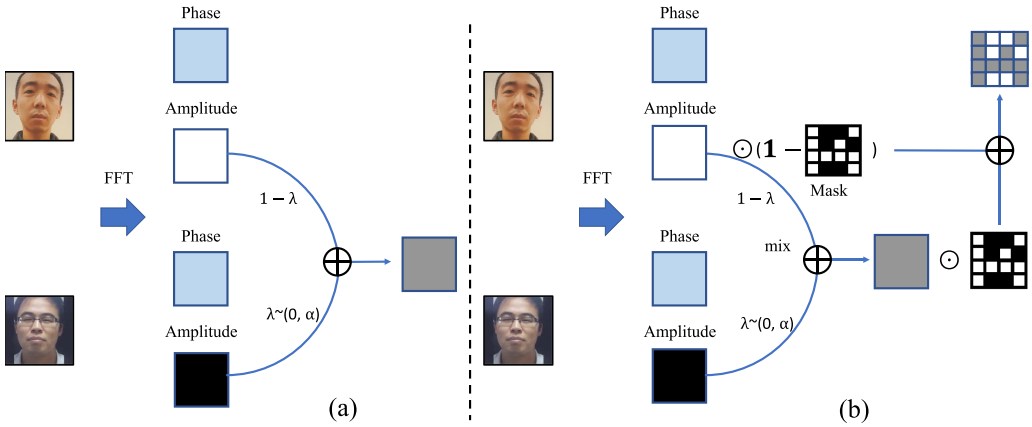


Fig. 2. Comparison between (a) amplitude mix [37] and our (b) masked amplitude mix.

$$\mathcal{F}^S(x_i)(u,v) = \hat{\mathcal{A}}^S(x_i)(u,v) * e^{-j*\mathcal{P}(x_i)(u,v)}. \quad (3)$$

The obtained Fourier spectrums are used to construct the domain-augmented image x_i^d and spoof-augmented image x_i^s by the inverse FFT, i.e., $x_i^d = \mathcal{F}^{-1}(\mathcal{F}^D(x_i))$ and $x_i^s = \mathcal{F}^{-1}(\mathcal{F}^S(x_i))$. We associate x_i^d and x_i^s with a mixed domain label d_i^d and a mixed task label y_i^s respectively to preserve the semantic relationship after augmentation. Specifically, $d_i^d = (1-\lambda)d_i + \lambda d_i'$ and $y_i^s = (1-\lambda)y_i + \lambda y_i'$.

C. Frequency Space Disentanglement

We now describe how we disentangle domain-related and spoof-related style information to obtain M_D and M_S in Section III-B. Intuitively, some styles like illumination and background are more related to domain variation, while others like moire pattern and reflection artifacts are associated with spoof attacks. [10] shows that spoof patterns can be found in some blocks in high-frequency domain. Considering that adjacent pixels in the Fourier spectrum represent signals with similar directions and frequencies, we propose to achieve disentanglement in the unit of spectrum blocks.

To this end, we divide the whole amplitude spectrum into $N_b = L \times L$ blocks, and define two low-scale masks $m_D, m_S \in \mathbb{R}^{L\times L}$ to identify domain- and spoof-related blocks, respectively. We first introduce an empirical method to achieve this goal. Concretely, at the begining of each epoch, we randomly retrieve a batch of images $\{\hat{x_i}\}_{i=1}^B$ from the training set. For each $\hat{x_i}$, we conduct the Fourier transformation on it and employ iFFT on its phase spectrum and each amplitude block respectively, constructing N_b images $\{\hat{x_i}^n\}_{n=1}^{N_b}$. For a domain-related block, we expect that the image reconstructed by it will yield a high output response on the domain discriminator. Meanwhile, we hope the discriminator is also confident about the prediction.

In this work, we consider the difference between the first and second highest output values as a measure of confidence, which can be formulated as

$$V_i^n = D(G(\hat{x_i}^n))[y_i^+] - D(G(\hat{x_i}^n))[y_i^-], \quad (4)$$

where y_i^+, y_i^- stand for the positions of the highest and second-highest outputs, respectively. We then construct a sample-level identification vector $V_i = [V_i^1, V_i^2, \ldots, V_i^{N_b}]$ for each $\hat{x_i}$. By averaging V_i in the batch, the batch-level identification vector can be obtained: $V = \frac{1}{B}\sum_{i=1}^B V_i$. The values in V indicate the significance of each amplitude block for domain classification. We therefore select blocks with values in the top $P\%$ as the domain-related amplitude blocks, thus yielding a binary mask m_D:

$$m_D[i,j] = \begin{cases} 1, & \text{if } V[L*i+j] \in top\ P\% \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Analogously, the spoof-related mask m_S can be obtained by substituting D with C in (4). Both m_S and m_D can be simply scaled to M_S and M_D for masked amplitude mix. We detail this process in the bottem part of Fig. 3.

In practical scenarios, disentanglement with binary masks may be too arbitrary. It is possible that some frequencies within a block are simultaneously related to the domain and spoof with different degrees. Therefore, apart from the binary masks based on empirical identification, we provide a feasible solution to automatically learn m_D and m_S by making them learnable parameters with a Sigmoid activation, leading to soft masks with continuous values in [0,1]. We thereafter refer to our method with binary masks and soft masks as FSDA-B and FSDA-S for differentiation.

D. Training Strategy

In this section, we introduce all training losses we designed to constitute the complete method. As we described in Section III-A, we employ a task-specific C and a domain discriminator D to learn spoof-related and domain-related knowledge, respectively. We first utilize the original training samples to set a basic model for our method. Specifically, for an original training sample x_i, we use a stand cross-entropy loss to train the classifier, i.e.,

$$\mathcal{L}_{cls} = -\mathbb{E}_{(x_i,y_i,d_i)} y_i \log\left(\sigma\left(C\left(G\left(x_i\right)\right)\right)\right), \quad (6)$$

where σ denotes the softmax operation. For training the domain discriminator, we adopt an adversarial learning-based strategy

Authorized licensed use limited to: UNIVERSITY OF NEVADA RENO. Downloaded on October 22,2026 at 20:18:24 UTC from IEEE Xplore. Restrictions apply.

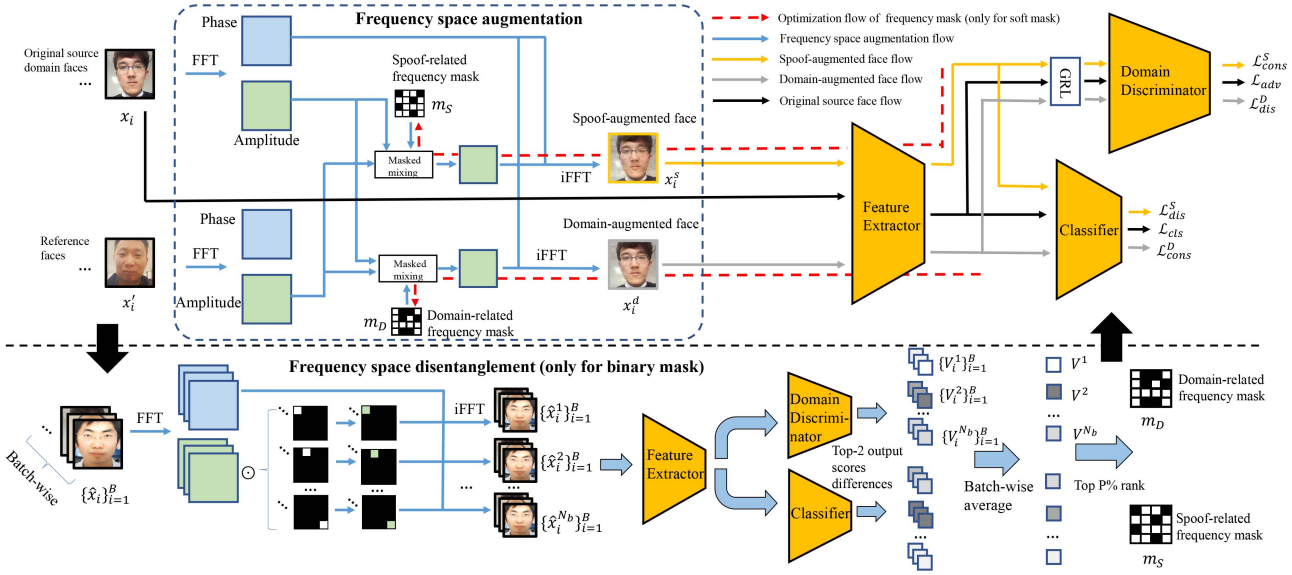


Fig. 3. Overview of the proposed method. RGB images from different domains are firstly transfered to the frequency space using FFT. Each image is then augmented to a domain-oriented and a spoof-oriented conterparts by blending the domain-related and spoof-related amplitude components with another reference image from the training set. The disentanglement masks can be learned by backpropagation from supervised losses (red dashed arrows) with continuous values, or via an empirical identification based on the classifier and the domain discriminator outputs for each frequency block with binary values. The augmented samples are utilized to train a domain- and spoof-robust model exploited by a distillation loss $\mathcal{L}_{dis} := \{\mathcal{L}^S_{dis}, \mathcal{L}^D_{dis}\}$ and a consistency loss $\mathcal{L}_{cons} := \{\mathcal{L}^S_{cons}, \mathcal{L}^D_{cons}\}$.

to pursue domain-invariant features:

$$\min_D \max_G \mathcal{L}_{adv} = -\mathbb{E}_{(x_i,y_i,d_i)} d_i \log\left(\sigma(D(G(x_i)))\right), \quad (7)$$

the gradient reversal layer (GRL) [46] is used to make the training proceed in an end-to-end fashion. Together training with above two losses improves the accuracy and transferability of the model within the source domains. However, the OOD generalization ability has not been explicitly considered yet. Fortunately, the augmented samples provide us with an explicit way to explore unseen domains and spoof types. Specifically, the domain-augmented samples x^d_i and spoof-augmented samples x^s_i contain rich knowledge about the domain relations and spoof relations between the original and augmented images, respectively. We aim to internalize this knowledge by enforcing the model to be consistent between the outputs of the augmented images and their soft labels, which is implemented by a distillation loss:

$$\mathcal{L}^D_{dis} = \mathbb{E}_{(x^d_i,y_i,d^d_i)} \mathcal{D}_{KL}\left(\sigma\left(D\left(G\left(x^d_i\right)\right)/T\right) \| d^d_i/T\right)$$
$$\mathcal{L}^S_{dis} = \mathbb{E}_{(x^s_i,y^s_i,d_i)} \mathcal{D}_{KL}\left(\sigma\left(C\left(G\left(x^s_i\right)\right)/T\right) \| y^s_i/T\right), \quad (8)$$

where $\mathcal{D}_{KL}(\cdot)$ denotes the Kullback-Leibler divergence, T is the temperature that controls the smoothness of the distribution. Such a distillation regularization enables an implicit learning on both latent domains and spoofs, and also forces the model to make judgments based on the information in the amplitude spectrum instead of the spoof- and domain-irrelevant content information in the phase spectrum.

Furthermore, to constrain our model not to detect presentation attacks based on domain-specific information and vice versa, we impose a consistency regularization on the model output, which is defined as

$$\mathcal{L}^D_{cons} = -\mathbb{E}_{(x^d_i,y_i,d^d_i)} y_i \log\left(\sigma\left(C\left(G\left(x^d_i\right)\right)\right)\right),$$
$$\mathcal{L}^S_{cons} = -\mathbb{E}_{(x^s_i,y^s_i,d_i)} d_i \log\left(\sigma\left(D\left(G\left(x^s_i\right)\right)\right)\right). \quad (9)$$

With all losses together, we summary the full training objective as follows:

$$\mathcal{L}_{FSDA} = \mathcal{L}_{cls} + \mathcal{L}_{adv} + \lambda_{aug}(\mathcal{L}^D_{dis} + \mathcal{L}^S_{dis} + \mathcal{L}^D_{cons} + \mathcal{L}^S_{cons}), \quad (10)$$

where λ_{aug} controls the training trade-off between the original samples and the augmented samples.

IV. Experiments

A. Datasets and Evaluation Protocols

We evaluate our method on four widely adopted FAS datasets, i.e., CASIA-MFSD [47] (C), OULU-NPU [48] (O), Idiap Replay-Attack [49] (I) and MSU-MFSD [24] (M). The datasets involve different subjects, capture devices, environments, and even attack types, resulting in substantial domain shifts. Following previous DG-based methods [14], [31], [50], we conduct experiments on two domain generalization protocols: a standard leave-one-out DG protocol and a limited-source domain DG protocol. The Half Total Error Rate (HTER) and the Area Under Curve (AUC) are used as evaluation metrics.

B. Implementaiton Details

The proposed method is implemented on Pytorch. For the network architecture, we adopt the embedding part of DepthNet [3]

Authorized licensed use limited to: UNIVERSITY OF NEVADA RENO. Downloaded on October 22,2026 at 20:18:24 UTC from IEEE Xplore. Restrictions apply.

TABLE I
RESULTS ON LEAVE-ONE-OUT DG PROTOCOL RESPECT TO HTER (%) AND AUC (%)

Method	O&C&M to I		O&C&I to M		O&M&I to C		I&C&M to O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
Binary CNN [9]	34.47	65.88	29.25	82.87	34.88	71.94	29.61	77.54
LBPTOP [25]	49.45	49.54	36.90	70.80	42.60	61.05	53.15	44.09
MMD-AAE [11]	31.58	75.18	27.08	83.19	44.59	58.29	40.98	63.08
MADDG [13]	22.19	84.99	17.69	88.06	24.50	84.51	27.98	80.02
SSDG-M [14]	18.21	**94.61**	16.67	90.47	23.11	85.45	25.17	81.83
RFM [18]	17.30	90.48	13.89	93.98	20.27	88.16	16.45	91.16
D^2AM [17]	15.43	91.22	12.70	95.66	20.98	85.58	15.27	90.87
SDA [19]	15.60	90.10	15.40	91.80	24.50	84.40	23.10	84.30
ANRL [31]	16.03	91.04	10.83	96.75	17.83	89.26	15.67	91.90
SSAN-M [32]	**14.00**	94.58	10.42	94.76	**16.47**	90.81	19.51	88.17
FSDA-B (Ours)	14.68	90.86	**8.63**	**97.25**	21.47	86.87	19.10	88.01
FSDA-S (Ours)	14.67	92.79	9.22	96.72	16.49	**90.91**	**15.07**	**92.92**

Smaller HTER and larger AUC stands for better performance. FSDA-B and FSDA-S denote our method with binary and soft masks, respectively. The bold formatting to highlight the best performance in each column.

as the feature backbone to be consistent with previous DG-based FAS methods. The structure of the domain discriminator is adapted from that in SSDG [14], and the classifier is a simple linear layer. For data preparation, we sample frames in each video at certain intervals and then crop the face regions by MTCNN [51] and resize them to $3 \times 256 \times 256$ RGB images. To ensure a balanced number of extracted images across different datasets, we sample a frame from videos every 5 to 10 frames, depending on the dataset, which results in total 12239, 10297, 10448, 12623 frames for C, I, M, O, respectively after pre-processing. For training details, we use the SGD optimizer with a momentum of 0.9 and weight decay of 5e-4 to update the model and the soft masks for FSDA-S. The initial learning rate for the network is set to 1 e^{-3} and decreased by 0.5 times after 200 and 400 epochs. The batch size is 10 for each source domain, therefore 30 for the whole batch. The mix hyper-parameter α is 0.5 and the identification percentage P is 25. The number of spectrum blocks N_b and the temperature T are empirically set to 64 (8×8) and 1.0 respectively for all experiments. For the hyper-parameter λ_{aug}, we linearly increase the value from 0 to 1.0 in the first 10 epochs and keeps it thereafter. For frequency space augmentation, we randomly select reference images without replacement for each training sample during the augmentation step. Specifically, for each training image, we randomly select an image from the same training batch as the reference image.

C. Comparisons With the State-of-The-Art

Leave-one-out protocol: We compare our method with other FAS methods on four typical DG tasks in Table I. It shows that our method obtains the best or comparable performances on all tasks with respect to both AUC and HTER. For instance, our FSDA-S achieves 90.91% on AUC and 16.49% on HTER on the most challenging task O&M&I to C, which outperforms the adversarial learning-based method SSDG [14] by 5.46% on AUC and 6.62% on HTER. It indicates that frequency space augmentation plays a vital role in generalization. For task I&C&M to O, our FSDA-S yields the highest AUC (92.92%) and HTER (15.07%), whereas the strong baseline SSAN obtains 88.17% (AUC) and 19.51% (HTER) on that. We conjecture the reason is that SSAN searches the spoof-related style information in the spatial space, which shows inferior discriminative power for subtle spoof traces. In contrast, our method not only focuses on the more discriminative frequency information but also explicitly performs augmentation on both domains and spoof clues. While the performance gains may be relatively small in some cases, it is important to consider the challenging nature of the FAS task and the saturation of performance in this field. In the context of FAS, even small improvements in performance can have significant practical implications. A slight increase in the detection accuracy or a decrease in the false positive rate can lead to substantial benefits in real-world applications, such as enhanced security in biometric systems or reduced user friction in authentication processes. The above results demonstrate that our method is more effective in cross-domain scenarios.

TABLE II
RESULTS ON LIMITED-SOURCE DOMAIN DG PROTOCOL

Method	M&I to C		M&I to O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)
MS_LBP [22]	51.16	52.09	43.63	58.07
Color Texture [23]	55.17	46.89	53.31	45.16
LBPTOP [25]	45.27	54.88	47.26	50.21
MADDG [13]	41.02	64.33	39.35	65.10
SSDG-M [14]	31.89	71.29	36.01	66.88
RFM [18]	36.34	67.52	29.12	72.61
D^2AM [17]	32.65	72.04	27.70	75.36
ANRL [31]	31.06	72.12	30.73	74.10
SSAN [32]	30.00	76.20	29.44	76.62
FSDA-B (Ours)	32.97	73.24	**26.89**	**78.82**
FSDA-S (Ours)	**29.58**	**76.63**	27.76	78.30

The bold formatting to highlight the best performance in each column.

Limited source domain protocol: Table II reports the results on two DG scenarios where only two source domains are available. Despite the limited source data, our method achieves the highest AUC and the lowest HTER on both M&I to C and M&I to O. It is worth noting that we outperform the SSDG [14] by a large margin on both tasks. We believe it is due to the difficulty of learning generalized feature representations with limited data. Our augmentation alleviates this concern and is therefore more

Authorized licensed use limited to: UNIVERSITY OF NEVADA RENO. Downloaded on October 22,2026 at 20:18:24 UTC from IEEE Xplore. Restrictions apply.

TABLE III
COMPARISON BETWEEN OUR METHOD AND BASELINES WITH AMPLITUDE MIX STRATEGY [37] AND ADVERSARIAL LEARNING [14]

Method	O&C&M to I		O&C&I to M		O&M&I to C		I&C&M to O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
BDG-M [12]	20.93	87.06	17.14	87.70	28.00	73.42	26.27	79.99
FSDA-FACT [35]	16.94	88.40	12.39	94.88	31.90	74.26	19.43	87.68
FSDA-S (Ours)	**14.67**	**92.79**	**9.22**	**96.72**	**16.49**	**90.91**	**15.07**	**92.92**

The bold formatting to highlight the best performance in each column.

TABLE IV
ABLATION STUDY ON DIFFERENT COMPONENTS OF THE PROPOSED METHOD ON TASK O&C&M TO I

Method	O&C&M to I	
	HTER(%)	AUC(%)
Binary CNN	34.47	65.88
w/ $\mathcal{L}_{adv}$	24.40	83.58
w/ $\mathcal{L}_{dis}$	25.69	81.94
w/ $\mathcal{L}_{adv}$ + $\mathcal{L}_{dis}$	15.32	89.22
w/ $\mathcal{L}_{adv}$ + $\mathcal{L}_{cons}$	16.51	87.64
FSDA-B (full model)	14.68	90.86
FSDA-S (full model)	**14.67**	**92.79**

The bold formatting to highlight the best performance in each column.

suitable for learning generalized patterns in such challenging scenarios.

D. Ablation Study

Effectiveness of Masked Amplitude Mix: To illustrate the effectiveness of the masked amplitude mix strategy, we implement a baseline method that adopts the amplitude mix strategy used in FACT [37] for augmentation (denoted as FSDA-FACT). Specifically, we mix the whole amplitude spectrums of two images without disentanglement, thus the $\mathcal{L}_{cons}$ is also disabled. Besides, we also adopt a pure adversarial learning baseline (i.e., BDG-M) from [14] for comparison. The results on four typical DG tasks are reported in Table III.

It turns out that compared with BDG-M, the FSDA-FACT brings significant improvement on O&C&I to M and I&C&M to O, and maintains comparable results on the other two tasks. We attribute the phenomenon to the frequency space augmentation adopted in FACT. However, it directly mixes the whole amplitude spectrums, overlooking the fact that domain- and spoof clues may entangle in the frequency space. As a result, the model is encouraged to explore spoof traces in the entire spectrum. In contrast, our method explicitly pursues a disentangled frequency space and guides the model to make classifications based on only the spoof-related frequency components, which is more effective and therefore achieves the best results on four DG tasks.

Effectiveness of Different Components: To study how different components of our method contribute to the final results, we implement four variants and evaluate them on task O&C&M to I. The comparison results are shown in Table IV. Specifically, w/ $\mathcal{L}_{adv}$ means a binary CNN model with pure adversarial learning. w/ $\mathcal{L}_{dis}$ denotes the variant that discards adversarial learning and only conducts spoof-oriented augmentation with the distillation loss. w/ $\mathcal{L}_{adv}$ + $\mathcal{L}_{dis}$ and w/ $\mathcal{L}_{adv}$ + $\mathcal{L}_{cons}$ stand for adversarial learning with the distillation loss and the consistency loss on augmented samples, respectively. It shows that adversarial learning does play an crucial role in learning domain-invariant features since it significantly boosts the performance of the source-only model. However, it is still difficult to capture discriminative and generalizable factors with limited data in the spatial space. On the other hand, our augmentation strategy (i.e., w/ $\mathcal{L}_{dis}$) alleviates the data-scarce problem and operates in more discriminative frequency space, while it neglects the domain shift problem, thus leading to sub-optimal performance. Combining $\mathcal{L}_{adv}$ with $\mathcal{L}_{dis}$ (or $\mathcal{L}_{cons}$) makes the model learn on the augmented data for latent domains and spoofs, and encourages it to predict domain or liveness labels based on the amplitude spectrum in the frequency space, which further benefits generalization and discriminability. Finally, combining all above losses yields the best results among all variants, indicating all components in our method are beneficial.

Comparison Between Binary and Soft Masks: From Tables I and II, we empirically find that a soft disentanglement yields superior results on most tasks, since it explores more fine-grained information in the frequency space without being affected by hyperparameters (e.g., P). However, we found that learning soft masks requires more GPU resources due to additional parameters. Specifically, it requires 44.3 GB memory for FSDA-S and 41.8 GB for FSDA-B. For training time, it takes about 150 seconds to complete a training epoch for FSDA-S and 140 seconds for FSDA-B. Furthermore, we empirically found FSDA-B often converges faster than FSDA-S. For instance, we observed the optimal result on task O&M&I to C at epoch 284 for FSDA-B, While 412 for FSDA-S. We conjecture the reason is that the small gradient makes the convergence of soft masks often take more time. In practice, we suggest choosing the strategy on demand adaptively.

E. Visualization Analysis

Disentanglement Visualization: We depict the disentanglement results in Fig. 4(a) and (b), which show that many frequencies in the amplitude spectrum are both domain-related and spoof-related. For instance, changes in brightness, color and background can cause a distribution shift. However, variations in brightness and color can also be caused by replay devices, and changes in the background may be caused by some spoof-related means, such as the hand holding the cut photo. This is the direct reason why we learn two masks independently instead of

Authorized licensed use limited to: UNIVERSITY OF NEVADA RENO. Downloaded on October 22,2026 at 20:18:24 UTC from IEEE Xplore. Restrictions apply.

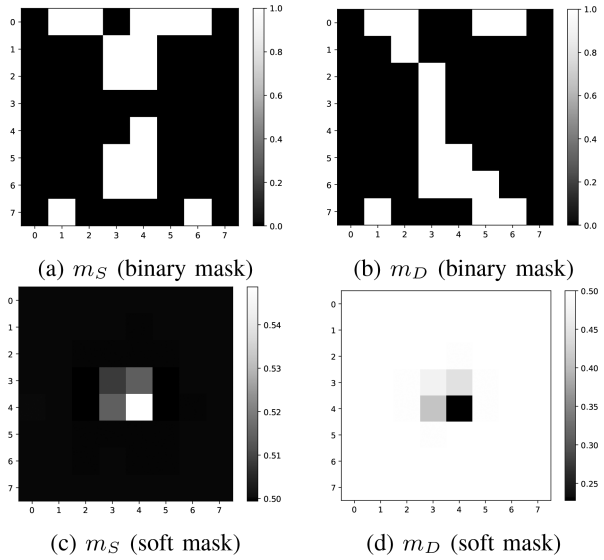


Fig. 4. The binary and soft masks (O&M&I to C).

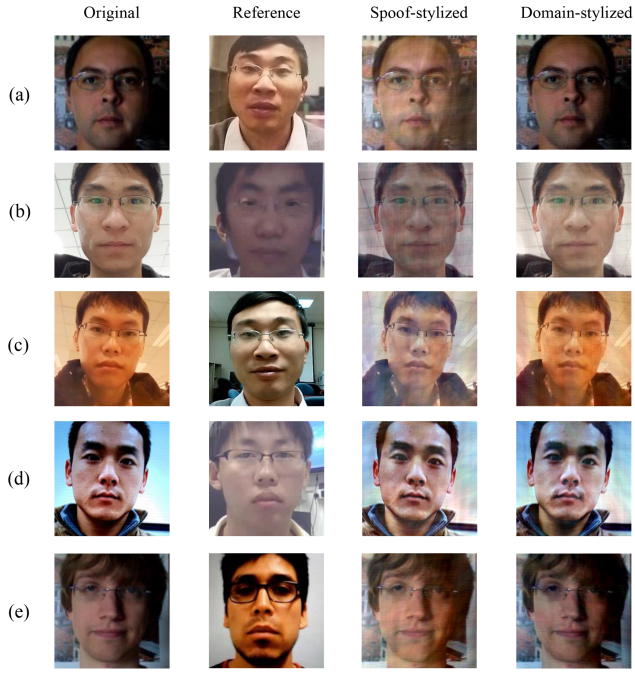


Fig. 5. Visualization of spoof-oriented and domain-oriented augmentation by masked amplitude mix with binary masks when $\lambda = 0.5$. (a)-(b) Real original images with spoof reference images. (c)-(d) Spoof original images with real reference images. (e) Original (real) and reference (spoof) images are from the same domain.

making them mutually exclusive. To study how much a frequency component is related to each style, we further show soft masks that present more fine-grained disentanglement. From Fig. 4(c) and (d), we can observe that most of the frequencies are equally important for domain and spoof, since they all have a value around 0.5. However, there are certain frequency blocks in the middle part that can have very different effects on domain and spoof variations, which verifies that domain- and spoof-related information is indeed hidden in different frequency components.

Augmentation Visualization: We visualize the images augmented by our masked amplitude mix in Fig. 5. From the first two rows, we can feel that when a real face mixes the spoof-related amplitude spectrum with that of a spoof face, the texture and brightness of the stylized image will be more like a fake face, e.g., the depth information in the image is weakened. The domain-oriented augmentation only changes the information that is not intuitively related to spoofs, such as the resolution of the image and the brightness. In contrast, we can observe from the last two rows that a spoof face will be more like a real face when it mixes the spoof-related amplitude spectrum with a real face. Specifically, the color distortion is weakened after augmentation for the third sample, which makes it more natural in presentation. Besides, the display artifacts is significantly alleviated for the last sample. These results indicate that our frequency space augmentation is able to generate meaningful and rational images.

Attention Visualization: To further investigate which parts indeed affect the classification results, we visualize the attention maps on some samples in the CASIA dataset using Grad-CAM [52]. It can be observed in Fig. 6 that a CNN model trained only on source domains shows irregular attention, since there is a significant domain shift between CASIA and other datasets. After adversarial learning, the model presents some interesting phenomena. Specifically, for the real faces, the model mainly focuses on the ficial regions for classification. In contrast, it relies on various non-facial regions for PA detection. However, some background parts are also thought to be useful for classification, such as the second real face. When training with our method, the attention of our model can cover the crucial regions accurately and comprehensively.

Feature Visualization: To vividly illustrate why our method enjoys a superior performance, we visualize the feature space learned by our FSDA-S and the source-only model in Fig. 7. We can see from Fig. 7(a) that a large number of spoof faces in the target domain are confused with real faces in source domains, which indicates that the model supervised only in source domains cannot learn discriminative and generalized spoof clues, since spoof clues are usually subtle and various. Our method guides the model to search spoof clues in the frequency space, which is more suitable for PA detection than spatial space. In addition, we explicitly augment the training set to make the model more generalizable to manually mimicked unseen domains and spoof traces, resulting in a discriminative and generalizable feature distribution in Fig. 7(b).

F. Parameter Sensitivity Analysis

We conduct additional experiments on O&C&M to I and O&M&I to C to investigate the impact of the hyperparameter α on the performance of our method (FSDA-S). Specifically, α controls the mixing ratio between the amplitude spectrums of the original and reference images during the frequency space augmentation. As shown in Fig. 8, the optimal value of α lies around 0.5, where the model achieves the best balance between

Authorized licensed use limited to: UNIVERSITY OF NEVADA RENO. Downloaded on October 22,2026 at 20:18:24 UTC from IEEE Xplore. Restrictions apply.

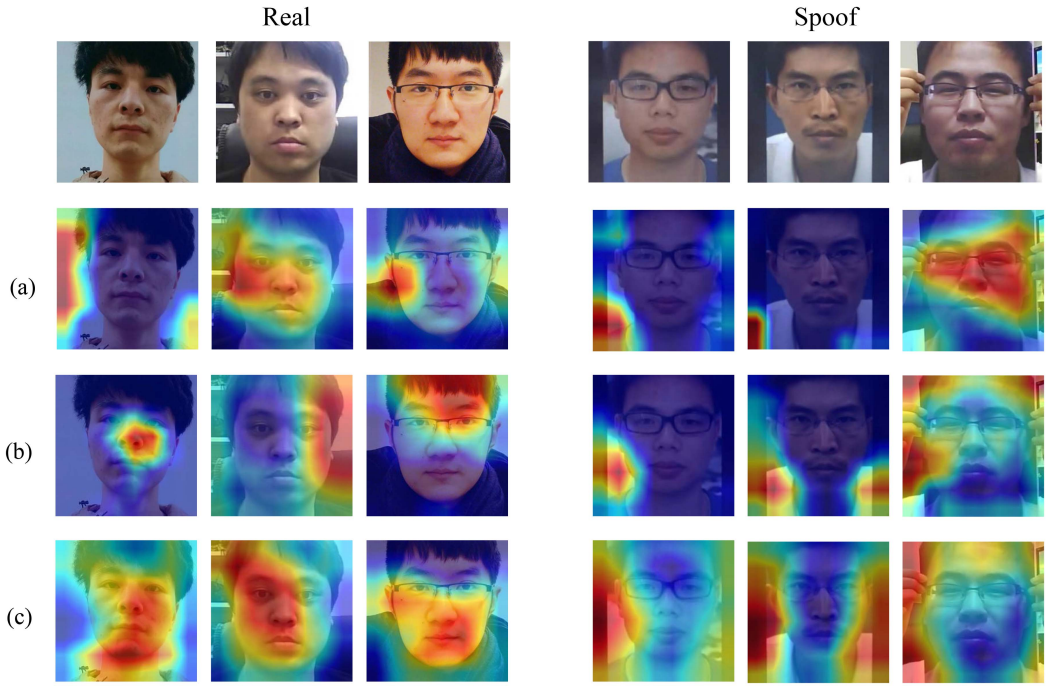
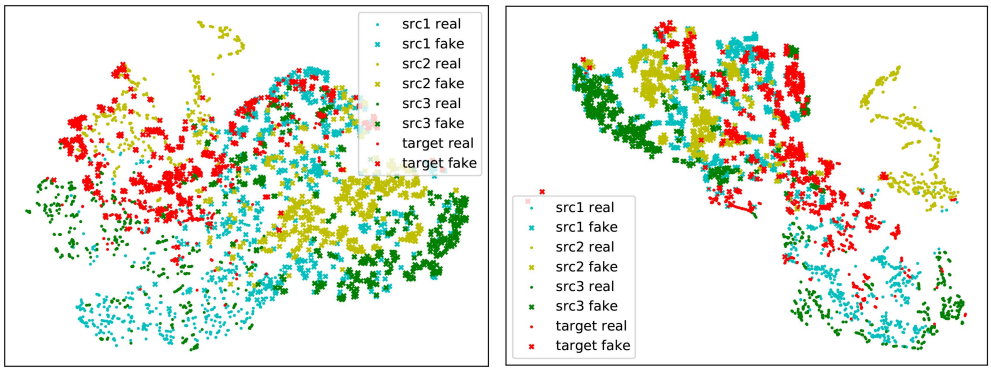


Fig. 6. Attention visualization using Grad-CAM [52] on task O&I&M to C. (a) Source-only model. (b) Training with only adversarial learning. (c) Training with adversarial learning and augmentation (FSDA-S).



(a) Binary CNN (Baseline) (b) FSDA-S

Fig. 7. The t-SNE feature visualization for (a) the baseline model and (b) our method on task O&M&I to C.

leveraging the augmented data and preserving the essential information from the original images. This observation is consistent across different DG tasks, demonstrating the robustness of our method. Based on these findings, we recommend setting α to 0.5 as a default value. However, the optimal value may vary slightly depending on the specific dataset and task. One can fine-tune the α parameter around 0.5 to further optimize the performance for specific use case.

G. Comparison Between Different Sampling Strategies for Augmentation

As we claimed in Section IV-B, for each original image, we randomly select an image from the same training batch as the reference image. This sampling is performed without replacement. To see the influence of this configuration, we evaluate another sampling strategy where the same reference image can be used for different original image, i.e., sampling with replacement. From Table V, we can observe that sampling without replacement achieves comparable or slightly better performance compared to the one with replacement. We conjecture the reason is that sampling without replacement ensures a diverse set of reference images for each training sample, preventing the model from overfitting to specific image pairs. By selecting reference images without replacement, we encourage the model to learn from a wide range of variations and combinations of training samples. This diversity in the augmented samples helps the model capture more robust and generalizable features.

Authorized licensed use limited to: UNIVERSITY OF NEVADA RENO. Downloaded on October 22,2026 at 20:18:24 UTC from IEEE Xplore. Restrictions apply.

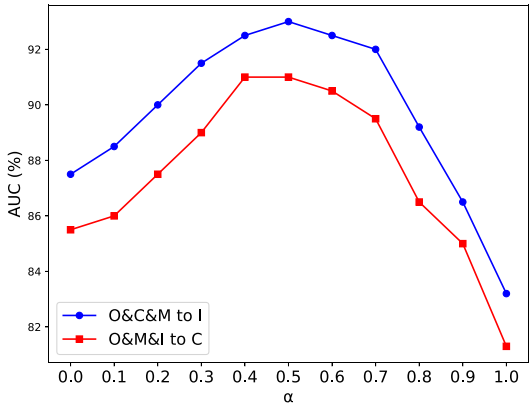


Fig. 8. Parameter sensitivity of α on O&C&M to I and O&M&I to C. We use FSDA-S in this experiment.

TABLE V
COMPARISON BETWEEN DIFFERENT SAMPLING STRATEGIES FOR FREQUENCY SPACE AUGMENTATION

Strategy	O&C&M to I		O&M&I to C	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)
FSDA-S (w/ rep.)	14.89	92.06	18.22	89.97
FSDA-S (w/o rep.)	**14.67**	**92.79**	**16.49**	**90.91**

'w/ rep.' and 'w/o rep.' represent sampling with or without replacement, respectively. The bold formatting to highlight the best performance in each column.

V. CONCLUSION

In this work, we aim to tackle generalizable FAS in the frequency space based on the amplitude spectrum of the Fourier transformation. Specifically, we disentangle domain-related and spoof-related frequency components and propose two strategies for this goal. A frequency space augmentation is then proposed based on the disentangled amplitude spectrum. We conduct training on both original and augmented samples, and experimental results on four FAS datasets verify the effectiveness of our method.

REFERENCES

[1] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4685–4694.
[2] Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao, "Deep learning for face anti-spoofing: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5609–5631, May 2023.
[3] Y. Liu, A. Jourabloo, and X. Liu, "Learning deep models for face anti-spoofing: Binary or auxiliary supervision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 389–398.
[4] J.-D. Lin et al., "Lightweight face anti-spoofing network for telehealth applications," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 5, pp. 1987–1996, May 2022.
[5] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
[6] A. Ali, F. Deravi, and S. Hoque, "Liveness detection using gaze collinearity," in *Proc. IEEE 3rd Int. Conf. Emerg. Secur. Technol.*, 2012, pp. 62–65.
[7] T. d. Freitas Pereira, A. Anjos, J. M. D. Martino, and S. Marcel, "LBP-top based countermeasure against face spoofing attacks," in *Proc. Comput. Vis.-ACCV*, 2012, pp. 121–132.
[8] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 10, pp. 2268–2283, Oct. 2016.
[9] J. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," 2014, *arXiv:1408.5601*.
[10] Y. Huang, W. Zhang, and J. Wang, "Deep frequent spatial temporal learning for face anti-spoofing," 2020, *arXiv:2002.03723*.
[11] H. Li, S. J. Pan, S. Wang, and A. C. Kot, "Domain generalization with adversarial feature learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5400–5409.
[12] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, "Deeper, broader and artier domain generalization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5542–5550.
[13] R. Shao, X. Lan, J. Li, and P. C. Yuen, "Multi-adversarial discriminative deep domain generalization for face presentation attack detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10015–10023.
[14] Y. Jia, J. Zhang, S. Shan, and X. Chen, "Single-side domain generalization for face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8481–8490.
[15] Y. Liu and X. Liu, "Spoof trace disentanglement for generic face anti-spoofing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3813–3830, Mar. 2023.
[16] Y. Liu, J. Stehouwer, and X. Liu, "On disentangling spoof trace for generic face anti-spoofing," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 406–422.
[17] Z. Chen et al., "Generalizable representation learning for mixture domain face anti-spoofing," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 1132–1139.
[18] R. Shao, X. Lan, and P. C. Yuen, "Regularized fine-grained meta face anti-spoofing," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 11974–11981.
[19] J. Wang, J. Zhang, Y. Bian, Y. Cai, C. Wang, and S. Pu, "Self-domain adaptation for face anti-spoofing," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 2746–2754.
[20] Y. Qin, Z. Yu, L. Yan, Z. Wang, C. Zhao, and Z. Lei, "Meta-teacher for face anti-spoofing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6311–6326, Oct. 2022.
[21] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of Fourier spectra," *Proc. SPIE*, vol. 5404, pp. 296–303, 2004.
[22] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. IEEE Int. Joint Conf. Biometrics*, 2011, pp. 1–7.
[23] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 8, pp. 1818–1830, Aug. 2016.
[24] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 4, pp. 746–761, Apr. 2015.
[25] T. d. F. Pereira et al., "Face liveness detection using dynamic texture," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, pp. 1–15, 2014.
[26] S.-Q. Liu, X. Lan, and P. C. Yuen, "Remote photoplethysmography correspondence feature for 3D mask face presentation attack detection," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 558–573.
[27] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li, and A. Hadid, "An original face anti-spoofing approach using partial convolutional neural network," in *Proc. 6th Int. Conf. Image Process. Theory, Tools Appl.*, 2016, pp. 1–6.
[28] Z. Wang et al., "Deep spatial gradient and temporal depth learning for face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5042–5051.
[29] H. Kuang et al., "Multi-modal multi-layer fusion network with average binary center loss for face anti-spoofing," in *Proc. 27th ACM Int. Conf. Multimedia*, 2019, pp. 48–56.
[30] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1126–1135.
[31] S. Liu et al., "Adaptive normalized representation learning for generalizable face anti-spoofing," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 1469–1477.
[32] Z. Wang et al., "Domain generalization via shuffled style assembly for face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 4113–4123.
[33] L. N. Piotrowski and F. W. Campbell, "A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase," *Perception*, vol. 11, no. 3, pp. 337–346, 1982.
[34] Y. Yang, D. Lao, G. Sundaramoorthi, and S. Soatto, "Phase consistent ecological domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9008–9017.

Authorized licensed use limited to: UNIVERSITY OF NEVADA RENO. Downloaded on October 22,2026 at 20:18:24 UTC from IEEE Xplore. Restrictions apply.

[35] Y. Gu, Z. Ge, C. P. Bonnington, and J. Zhou, "Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 5, pp. 1379–1393, May 2020.

[36] Y. Yang and S. Soatto, "FDA: Fourier domain adaptation for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 4084–4094.

[37] Q. Xu, R. Zhang, Y. Zhang, Y. Wang, and Q. Tian, "A Fourier-based framework for domain generalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 14378–14387.

[38] Q. Zhou et al., "Generative domain adaptation for face anti-spoofing," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 335–356.

[39] Z. Wang, Y. Luo, R. Qiu, Z. Huang, and M. Baktashmotlagh, "Learning to diversify for single domain generalization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 814–823.

[40] K. Zhou, Y. Yang, T. Hospedales, and T. Xiang, "Learning to generate novel domains for domain generalization," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 561–578.

[41] J. Huang, D. Guan, A. Xiao, and S. Lu, "FSDR: Frequency space domain randomization for domain generalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 6887–6898.

[42] H. J. Nussbaumer, "The fast Fourier transform," in *Fast Fourier Transform and Convolution Algorithms*. Berlin, Germany: Springer, 1981, pp. 80–111.

[43] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Representations*, 2017.

[44] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang, "Domain generalization with mixstyle," in *Proc. Int. Conf. Learn. Representations*, 2021.

[45] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6022–6031.

[46] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.

[47] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proc. IEEE 5th IAPR Int. Conf. Biometrics*, 2012, pp. 26–31.

[48] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "OULU-NPU: A mobile face presentation attack database with real-world variations," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2017, pp. 612–618.

[49] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. IEEE Int. Conf. Biometrics Special Int. Group*, 2012, pp. 1–7.

[50] Z. Du, J. Li, L. Zuo, L. Zhu, and K. Lu, "Energy-based domain generalization for face anti-spoofing," in *Proc. 30th ACM Int. Conf. Multimedia*, 2022, pp. 1749–1757.

[51] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[52] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2921–2929.

Authorized licensed use limited to: UNIVERSITY OF NEVADA RENO. Downloaded on October 22,2026 at 20:18:24 UTC from IEEE Xplore. Restrictions apply.