FEW-SHOT ONE-CLASS DOMAIN ADAPTATION BASED ON FREQUENCY FOR IRIS PRESENTATION ATTACK DETECTION

Yachun Li, Ying Lian, Jingjing Wang, Yuhui Chen, Chunmao Wang, Shiliang Pu[*]

Hikvision Research Institute

ABSTRACT

Iris presentation attack detection (PAD) has achieved remarkable success to ensure the reliability and security of iris recognition systems. Most existing methods exploit discriminative features in the spatial domain and report outstanding performance under intra-dataset settings. However, the degradation of performance is inevitable under cross-dataset settings, suffering from domain shift. In consideration of real-world applications, a small number of bonafide samples are easily accessible. We thus define a new domain adaptation setting called Few-shot One-class Domain Adaptation (FODA), where adaptation only relies on a limited number of target bonafide samples. To address this problem, we propose a novel FODA framework based on the expressive power of frequency information. Specifically, our method integrates frequency-related information through two proposed modules. Frequency-based Attention Module (FAM) aggregates frequency information into spatial attention and explicitly emphasizes high-frequency fine-grained features. Frequency Mixing Module (FMM) mixes certain frequency components to generate large-scale target-style samples for adaptation with limited target bonafide samples. Extensive experiments on LivDet-Iris 2017 dataset demonstrate the proposed method achieves state-of-the-art or competitive performance under both cross-dataset and intra-dataset settings.

***Index Terms*—** Iris Presentation Attack Detection, Few-shot One-class Domain Adaptation, Frequency-based Attention

1. INTRODUCTION

Iris has unique and abundant texture information and has been widely used in biometric recognition applications of high-security levels. However, iris recognition system is vulnerable to various presentation attacks, causing security concerns in our society.

To alleviate these issues, presentation attack detection (PAD) has drawn growing attention. As deep learning has proved its ability in many applications, many methods employ CNNs to develop PAD systems to detect spoof samples [1–5]. Nevertheless, these methods exploit discriminative features for iris PAD only in the spatial domain. Bonafide iris has rich textures, while attacks such as the colored contact lenses have specific patterns and printouts would show some artifacts due to print quality. These differences can be easily captured in the frequency domain and are rarely exploited in iris PAD yet. Besides, frequency information has demonstrated its power in many related fields. It has been seamlessly inserted in networks to incorporate frequency components to color space [6, 7]. Additional frequency domain analysis explores spoofing clues in both face liveness detection [8, 9] and forgery detection [10, 11] tasks.

Therefore, we propose a novel Frequency-based Attention Module (FAM). FAM aggregates frequency information into spatial attention and is embedded in multiple layers to explicitly highlight high frequency components. Specifically, we employ Discrete Cosine Transform (DCT) to obtain frequency-based features. Then a learnable mask is applied to filter out low frequency components and keep high frequency ones. The masked frequency-based features are finally converted back to spatial domain through inverse DCT and act as the frequency-based attention to original features. FAM is different from SE [12] and CBAM [13] since they only focus on spatial domain. The most related work is FcaNet [7]. The similarity is that we both integrate frequency information into the attention map. However, FcaNet presents multi-spectral channel attention aiming at more efficient channel representation, while our FAM generates spatial attention in multiple layers to highlight high frequency components for iris PAD.

Although current iris PAD works achieve promising results in intra-dataset settings, the performance inevitably degrades under cross-dataset testing, due to domain shift. To alleviate this, domain adaptation (DA) techniques have been introduced to face anti-spoofing [14, 15]. It has not been employed in iris PAD field yet. Moreover, face anti-spoofing DA methods utilize a large number of unlabeled images, including both live and spoof ones. It is of high cost to collect attack samples, but collecting several bonafide samples from target domain should be convenient. Therefore, we define a new few-shot domain adaptation setting called Few-shot One-class Domain Adaptation (FODA), where only a small number of bonafide samples from target domain are available.

Inspired by [16], we propose a new Frequency Mixing Module (FMM), which generates large-scale target-style sam-

[*] Corresponding author (pushiliang.hri@hikvision.com).

978-1-6654-0540-9/22/$31.00 ©2022 IEEE

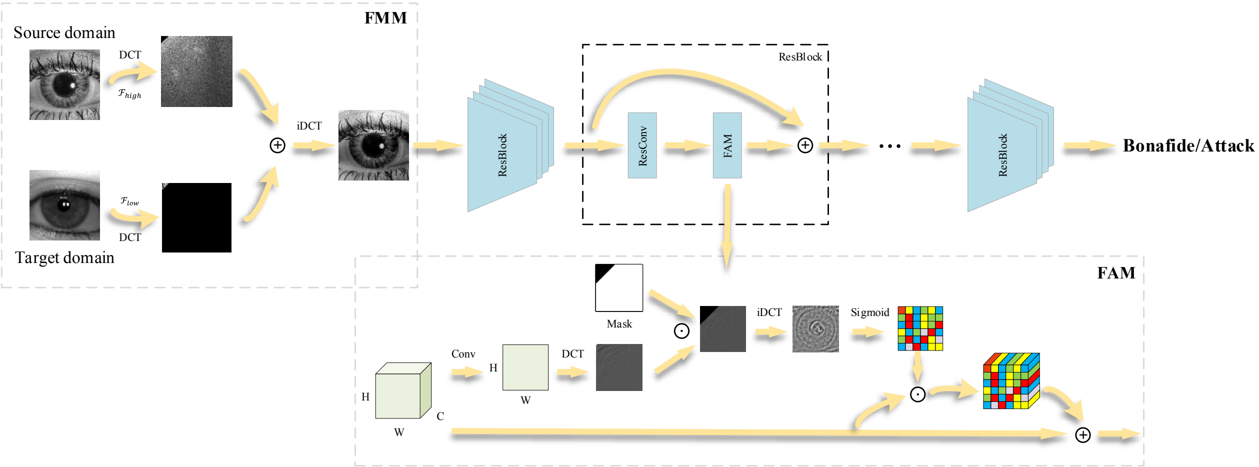

Fig. 1. Overview of few-shot one-class domain adaptation framework. Frequency information is leveraged in both frequency mixing module and frequency-based attention module.

ples with limited target bonafide samples by mixing certain frequency components. To be specific, low frequency components are more style-related and contribute more to domain shift, while high frequency components contain more bonafide/attack-related contents. We mix low frequency components of images from target domain with high frequency components of images from source domain, in order to generate new samples with source labels and target styles.

Finally, our proposed FODA framework combines FAM and FMM. FMM first generates target-style samples with few target bonafide samples. Original and adapted source samples are fed to the FAM-embedded network. Multi-scale FAM helps to explicitly emphasize high frequency components and benefits PAD-related feature learning. Two frequency-based modules integrate efficient frequency information and facilitate iris presentation attack detection.

2. PROPOSED METHOD

Fig. 1 shows an overview of our few-shot one-class domain adaptation framework, including two modules. Frequency-based Attention Module (FAM) adaptively decouples PAD-related high frequency features. High frequency components are extracted from features through discrete cosine transform (DCT), and then converted back to spatial domain via inverse DCT to get frequency-based attention. Frequency Mixing Module (FMM) is designed to address the domain shift problem. It generates a large number of target-style samples with very few target bonafide samples. By mixing certain frequency components of source and target images, the mixing images have source labels and target style at the same time. This helps to reduce the domain shift caused by style discrepancy. Both source and mixed samples are fed to the network, and the multi-scale FAM network explicitly emphasizes high frequency components to facilitate PAD feature learning. We describe the two frequency-based modules in detail next.

2.1. Frequency-based Attention Module

Images can be decomposed into low and high frequency components. Low frequency components are usually related to style content, while high frequency components contain more fine-grained information. As iris PAD task is more concerned with subtle textures, emphasis on high frequency ones could bring more generalizability to the model. To this end, we propose to explicitly enhance high frequency part via FAM.

Specifically, given a feature map $\mathbf{f} \in \mathbb{R}^{C \times H \times W}$, FAM first aggregates multiple channels to single one, where the output feature $\mathbf{f}_a \in \mathbb{R}^{H \times W}$ is calculated by:

$$\mathbf{f}_a = F_{aggre}(\mathbf{f}) = \sum_{i=1}^{C} W_i \cdot \mathbf{f}_i \tag{1}$$

Discrete Cosine Transform (DCT) is applied to one-channel feature $\mathbf{f}_a$. And then we leverage a learnable mask $\mathbf{M}$, initialized to remove the low-frequency band, in order to adaptively filter out low frequency components while keep high frequency ones in frequency domain:

$$\mathbf{f}_{highfreq} = \mathcal{D}(\mathbf{f}_a) \odot \mathbf{M} \tag{2}$$

where $\mathcal{D}$ is DCT and $\odot$ is the element-wise product.

Finally, we employ inverse DCT to transform high frequency feature to spatial domain and scale the original feature $\mathbf{f}$ with this frequency-aware attention map. The frequency-based attention is obtained by:

$$\mathbf{f}_{freqatt} = \sigma(\mathcal{D}^{-1}(\mathbf{f}_{highfreq})) \tag{3}$$

where sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ aims at squeezing x within the range $(0, 1)$. We adopt a residual connection to get the final frequency-enhanced feature $\mathbf{f}_{FAM} = \mathbf{f} + \mathbf{f}_{freqatt} \odot \mathbf{f}$. Thus, high frequency components are explicitly emphasized through the proposed frequency-aware attention mechanism.

Table 1. Ablation studies under cross-dataset settings.

Trained Dataset		IIITD-WVU		NotreDame		Clarkson		Average
Tested Dataset		NotreDame	Clarkson	IIITD-WVU	Clarkson	IIITD-WVU	NotreDame	
wo/DA	Baseline	7.33	45.69	20.97	11.23	29.46	28.83	23.92
	SE	5.75	42.26	19.17	13.33	31.47	34.14	24.35
	CBAM	9.11	30.72	13.47	14.08	36.78	31.17	22.56
	FAM	12.86	28.90	14.38	9.39	25.88	19.81	18.54
w/DA	DANN*	10.53	25.57	14.71	20.45	27.83	19.94	19.84
	MMD*	20.31	40.58	**11.36**	23.31	26.39	**14.14**	22.68
	FMM	4.50	36.31	18.49	**8.91**	22.83	24.61	19.27
	FMM+SE	**3.83**	27.72	21.47	14.72	29.03	29.72	21.08
	FMM+CBAM	6.36	**24.99**	14.76	11.82	36.85	35.53	21.72
	FMM+FAM	5.81	26.03	15.07	10.51	**22.06**	20.92	**16.73**

* Entire unlabeled target domain images were used for adaptation.

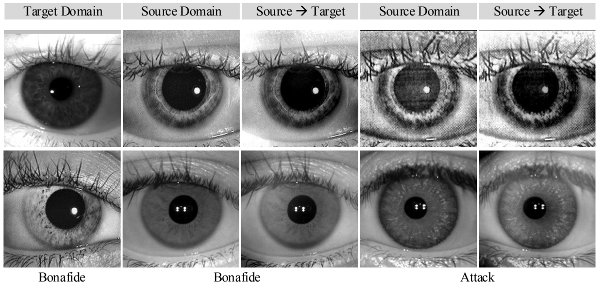

Fig. 2. Mixed samples from FMM.

2.2. Frequency Mixing Module

Considering domain adaptation scenarios, there are distribution biases between training and testing domains. The discrepancy of domains might be from capturing equipment, illuminations, backgrounds, etc. Therefore, we would like to generate samples that are able to imitate the possible changes. The high frequency components contain textures features that are crucial to detect presentation attacks. Domain style is more related to low-frequency components, which may help to distinguish different domains. Based on this, we present a novel module called FMM to adapt model towards target domain with few costs.

For presentation attack detection task, bonafide samples are more accessible than attack ones. We thus define the typical domain adaptation problem, where only a small number of bonafide samples from target domain are available (e.g. 10 bonafide samples). Given source domain $\mathcal{S} = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ and target domain $\mathcal{T} = \{(x_i^t, y_i^t)\}_{i=1}^{N_t}$ (where N_t is a small number and label $\{y_i^t\}_{i=1}^{N_t}$ is bonafide), we adopt domain adaptation by mixing high frequency components from source domain with low frequency ones from target domain. To achieve this, we design a binary low frequency filter $\mathcal{F}_{low}$ to extract style contents from target domain, while a reversed high frequency filter $\mathcal{F}_{high}$ decompose bonafide/attack-related contents from source domain:

$$\mathbf{x}_{low}^t = \mathcal{F}_{low}(\mathcal{D}(\mathbf{x}^t)), \tag{4}$$

$$\mathbf{x}_{high}^s = \mathcal{F}_{high}(\mathcal{D}(\mathbf{x}^s)), \tag{5}$$

Then, we apply frequency-base mixing to the obtained $\mathbf{x}_{low}^t$ and $\mathbf{x}_{high}^s$. The mixing is conducted in frequency domain and finally converted back to spatial domain via inverse DCT:

$$\mathbf{x}^{s \to t} = \mathcal{D}^{-1}(\mathbf{x}_{low}^t + \mathbf{x}_{high}^s) \tag{6}$$

As bonafide/attack-related contents are contained in high frequency components from source domain, the mixed images should have identical labels as source images. We set low-frequency band to 2.5% in our experiments. The mixed samples are shown in Fig. 2.

3. EXPERIMENTS

3.1. Experimental Setup

Datasets. The proposed method is evaluated on the LivDet-Iris 2017 dataset [18], which consists of 4 different datasets. However, the Warsaw dataset is no longer publicly available, so we use the remaining Clarkson, Notre Dame, and IIITD-WVU datasets for evaluation. Metrics Attack Presentation Classification Error Rate (APCER), Bonafide Presentation Classification Error Rate (BPCER), and Half Total Error Rate (HTER) are deployed to measure the performance.

Implementation Details. The input iris image is grayscale and of 200 × 200 size. Random cropping is performed in the training phase. We employ ResNet18 as backbone and initial it with ImageNet pre-trained model. In domain adaptation task, the number of bonafide samples N_t from target domain is 10, and source samples are randomly replaced with frequency-mixing samples by $p = 0.5$. All the experiments are based on PyTorch.

3.2. Ablation Studty

In this section, we conduct ablation experiments to verify the effectiveness of frequency-based components in our proposed domain adaptation framework.

Effect of frequency-based attention. To testify the effectiveness of FAM where high frequency components are explicitly emphasized, we compare FAM with SE [12] and

Table 2. Comparison to existing SoTA methods on LivDet-Iris 2017 dataset under cross-dataset settings.

Trained Dataset	IIITD-WVU		NotreDame		Clarkson	
Tested Dataset	NotreDame	Clarkson	IIITD-WVU	Clarkson	IIITD-WVU	NotreDame
PBS [17]	<u>16.86</u>	47.17	17.49	45.31	42.48	32.42
A-PBS [17]	27.61	**21.99**	**9.49**	<u>22.46</u>	<u>34.17</u>	<u>23.08</u>
FAM+FMM(Ours)	**5.81**	<u>26.03</u>	<u>15.07</u>	**10.51**	**22.06**	**20.92**

Table 3. Comparison to existing SoTA methods on LivDet-Iris 2017 dataset under intra-dataset settings.

Database	Metric	PAD Algorithm(%)						
		CASIA [18]	SpoofNet [2]	D-NetPAD [19]	MSA [5]	PBS [17]	A-PBS [17]	FAM
Clarkson	APCER	13.39	33.00	5.78	-	8.97	6.16	6.10
	BPCER	0.81	0.00	0.94	-	0.00	0.81	0.81
	HTER	7.10	16.50	**3.36**	-	4.48	3.48	<u>3.45</u>
NotreDame	APCER	7.78	18.05	10.38	12.28	8.89	7.88	8.06
	BPCER	0.28	0.94	3.32	0.17	1.06	0.00	0.00
	HTER	<u>4.03</u>	9.50	6.81	6.23	4.97	**3.94**	<u>4.03</u>
IIIT-WVU	APCER	29.40	0.34	36.41	2.31	5.76	8.86	1.00
	BPCER	3.99	36.89	10.12	19.94	8.26	4.13	12.68
	HTER	16.70	18.62	23.27	11.13	7.01	**6.50**	<u>6.84</u>

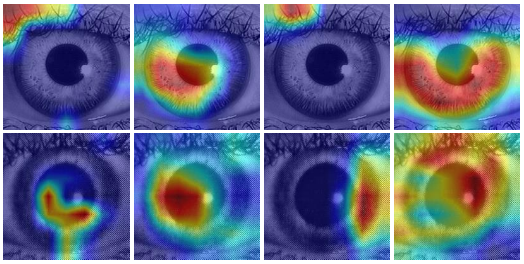
(a) Baseline (b) SE (c) CBAM (c) FAM

Fig. 3. Grad-CAM [20] visualization results. The first row is colored contact lens sample. The second row is printout attack sample.

CBAM [13]. Experiments are conducted both without domain adaptation (donated as wo/DA) and with domain adaptation (w/DA). As shown in Table 1, average performance of FAM is constantly superior to SE and CBAM, especially under w/DA scenario. We apply the Grad-CAM [20] to visualize class activation maps of attack class (see Fig. 3). Compared with other methods, it is clear that our FAM shows more interpretable attention and focuses on discriminative high-frequency regions according to different attacks. The improvement of FAM demonstrates the generalization ability of high frequency components.

Effect of frequency mixing domain adaptation. FMM is designed for few-shot domain adaptation. To verify the effectiveness of its adaptation ability, we compare it with widely-used domain adaptation methods DANN [21] and MMD [22]. Following the papers, we used the entire unlabeled target images for adaptation, while our FMM had limited access to only 10 bonafide images from target domain. Although DANN and MMD gain great improvement in certain settings, FMM still performs best considering the overall HTER. Adaptation results of FMM illustrate its efficacy even with few shots. Further integration of FAM, also referred to as our full FODA framework, increases average HTER to 16.73%, improving 2.54%. Thus, frequency mixing in our adaptation framework really helps to extract more adaptive features across datasets.

3.3. Comparison with State-of-the-Art Methods

We compare our method with current state-of-the-art methods in Table 2. The results show that our method performs better than PBS [17] consistently and surpasses A-PBS [17] in the majority of settings. It is worth noting that Clarkson dataset has a quite distinct style compared to the others, and we achieve 11.95%, 12.11% improvement when adapting from NotreDame to Clarkson, Clarkson to IIITD-WVU respectively. Our method integrates source bonafide/attack-related components to style contents from target domain and boosts the performance with very few costs.

As few methods in the literature focus on cross-dataset iris PAD, Table 3 also reports results of FAM under intra-dataset settings. We still observe superior or comparable results compared to SoTA methods. It confirms high frequency components are effective in iris presentation attack detection.

4. CONCLUSION

In this paper, we propose a novel few-shot one-class domain adaptation framework, including frequency-based attention module (FAM) for high frequency component highlighting and frequency mixing module (FMM) for style-adapted domain adaptation. FAM aggregates frequency information into spatial attention and emphasizes fine-grained features. FMM mixes low frequency components of source images with high frequency components of very few target bonafide images. Extensive experiments show the effectiveness of the proposed method under both cross-dataset and intra-dataset settings.

5. REFERENCES

[1] Cunjian Chen and Arun Ross, "A multi-task convolutional neural network for joint iris detection and presentation attack detection," in *WACVW*, 2018.

[2] Gabriela Y Kimura, Diego R Lucio, Alceu S Britto Jr, and David Menotti, "Cnn hyperparameter tuning applied to iris liveness detection," *arXiv preprint arXiv:2003.00833*, 2020.

[3] Lingxiao He, Haiqing Li, Fei Liu, Nianfeng Liu, Zhenan Sun, and Zhaofeng He, "Multi-patch convolution neural network for iris liveness detection," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2016.

[4] Ramachandra Raghavendra, Kiran B Raja, and Christoph Busch, "Contlensnet: Robust iris contact lens detection using deep convolutional neural networks," in *WACV*, 2017.

[5] Meiling Fang, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper, "Micro stripes analyses for iris presentation attack detection," in *IJCB*, 2020.

[6] Kai Xu, Minghai Qin, Fei Sun, Yuhao Wang, Yen-Kuang Chen, and Fengbo Ren, "Learning in the frequency domain," in *CVPR*, 2020.

[7] Zequn Qin, Pengyi Zhang, Fei Wu, and Xi Li, "Fcanet: Frequency channel attention networks," in *ICCV*, 2021.

[8] Gahyun Kim, Sungmin Eum, Jae Kyu Suhr, Dong Ik Kim, Kang Ryoung Park, and Jaihie Kim, "Face liveness detection based on texture and frequency analyses," in *2012 5th IAPR international conference on biometrics (ICB)*, 2012.

[9] Meiling Fang, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper, "Learnable multi-level frequency decomposition and hierarchical attention mechanism for generalized face presentation attack detection," *arXiv preprint arXiv:2109.07950*, 2021.

[10] Shen Chen, Taiping Yao, Yang Chen, Shouhong Ding, Jilin Li, and Rongrong Ji, "Local relation learning for face forgery detection," *arXiv preprint arXiv:2105.02577*, 2021.

[11] Yuyang Qian, Guojun Yin, Lu Sheng, Zixuan Chen, and Jing Shao, "Thinking in frequency: Face forgery detection by mining frequency-aware clues," in *ECCV*, 2020.

[12] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," in *CVPR*, 2018.

[13] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "Cbam: Convolutional block attention module," in *ECCV*, 2018.

[14] Haoliang Li, Wen Li, Hong Cao, Shiqi Wang, Feiyue Huang, and Alex C Kot, "Unsupervised domain adaptation for face anti-spoofing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1794–1809, 2018.

[15] Jingjing Wang, Jingyi Zhang, Ying Bian, Youyi Cai, Chunmao Wang, and Shiliang Pu, "Self-domain adaptation for face anti-spoofing," in *AAAI*, 2021.

[16] Yanchao Yang and Stefano Soatto, "Fda: Fourier domain adaptation for semantic segmentation," in *CVPR*, 2020.

[17] Meiling Fang, Naser Damer, Fadi Boutros, Florian Kirchbuchner, and Arjan Kuijper, "Iris presentation attack detection by attention-based and deep pixel-wise binary supervision network," in *IJCB*, 2021.

[18] David Yambay, Benedict Becker, Naman Kohli, Daksha Yadav, Adam Czajka, Kevin W Bowyer, Stephanie Schuckers, Richa Singh, Mayank Vatsa, Afzel Noore, et al., "Livdet iris 2017—iris liveness detection competition 2017," in *IJCB*, 2017.

[19] Renu Sharma and Arun Ross, "D-netpad: An explainable and interpretable iris presentation attack detector," in *IJCB*, 2020.

[20] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *ICCV*, 2017.

[21] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky, "Domain-adversarial training of neural networks," *The journal of machine learning research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[22] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell, "Deep domain confusion: Maximizing for domain invariance," *arXiv preprint arXiv:1412.3474*, 2014.