EasyChair Preprint

№ 13542

Domain-Generalized Face Anti-Spoofing with Domain Adaptive Style Extraction

Sunghun Yang, Jungho Lee, Sungjun Jang, Minseok Kang, Yongju Lee and Sangyoun Lee

June 4, 2024

Domain-Generalized Face Anti-Spoofing with Domain Adaptive Style Extraction

Sunghun Yang
Yonsei University
Seoul, Korea
sunghun98@yonsei.ac.kr

Jungho Lee
Yonsei University
Seoul, Korea
2015142131@yonsei.ac.kr

Sungjun Jang
Yonsei University
Seoul, Korea
jeu2250@yonsei.ac.kr

Minseok Kang
Yonsei University
Seoul, Korea
louis0503@yonsei.ac.kr

Yongju Lee
Yonsei University
Seoul, Korea
paulyongju@yonsei.ac.kr

Sangyoun Lee
Yonsei University
Seoul, Koera
syleee@yonsei.ac.kr

***Abstract*—Face anti-spoofing is a important task in securing face recognition systems. In particular, domain generalization for face anti-spoofing has been extensively studied, with the goal of increasing robustness across different datasets and real-world scenarios. Existing methods of domain generalization for face anti-spoofing require re-providing prior information for each new dataset, limiting their applicability. To address this limitation, we introduce a simplified domain-generalized face anti-spoofing (FAS) model that excels in diverse environments without requiring domain-specific modifications. By prioritizing the distinction between textural and non-facial features over conventional facial attributes, our model adapts to various unseen domains, leveraging dynamic kernels and style transfer AdalN for domain-invariant feature extraction. This approach mitigates the model's vulnerability to environmental and attack vector variations, enhancing its generalizability. Our comprehensive evaluation demonstrates the model's superior performance and adaptability, comparing favorably with state-of-the-art methods without the need for predefined domain knowledge or specific attack categorization. The model simplifies the binary classification process between spoof and live samples, showcasing its practical applicability in enhancing biometric security systems. Through this work, we provide valuable insights into domain generalization, offering considerations that are instrumental for future research in face anti-spoofing.**

***Index Terms*—Face anti-spoofing, Domain generalize, Style**

I. Introduction

In the digital security, face anti-spoofing (FAS) stands as a critical barrier against fraudulent access attempts, protecting biometric authentication systems from various present attacks. Traditional approaches, however, often grapple with domain-specific limitations, where the efficacy of an FAS model is tightly bound to the characteristics of the dataset on which it was trained. This dependency poses a significant challenge in real-world applications, where the diversity of attack vectors and environmental conditions cannot be exhaustively covered by a single dataset. To address this challenge, our work introduces a simple domain-generalized face anti-spoofing model designed to operate effectively across diverse conditions without the need for domain-specific tuning. This emphasis on distinguishing between texture and facial information attributes over direct facial characteristics in our model is motivated by the inherent complexities within Face Anti-Spoofing FAS tasks. Both authentic and spoofed samples display facial features, posing a significant challenge in distinguishing them based solely on facial data. The critical reliance on facial information primarily arises in the initial stages, particularly during face detection activities that lay the groundwork for FAS processes. By focusing on the underlying and external features, our model reduces its susceptibility to variations in facial expressions enhancing its ability to adapt and perform across various unseen domains

The main contribution of our approach is the innovative use of dynamic kernels inspired by IADG [1], [2] and style transfer AdalN inspired by SSAN [3] which together facilitate the extraction of domain-invariant features. Considering the advantages of these two methodologies, we have conceived an efficient network structure. By applying distinct normalization techniques to facial features and texture features separately, our model efficiently segregates these two fundamental aspects of an image. This separation is crucial, as it allows for the targeted minimization of domain-specific characteristics that could otherwise lead to model underperformance on novel datasets. Moreover, our model a slightly modified version of employs supervised contrastive loss [4] and smooth L1 loss to further refine the learning process. The former enhances the model's focus on texture features that are indicative of spoofing attempts, while the latter aims to minimize the influence of content features, thereby improving the overall robustness and adaptability of the model.

By combining the proposed methods, our domain-agnostic face anti-spoofing approach achieves better performance with existing domain generalization models. Our contributions not only pave the way for more secure and reliable biometric authentication systems but also offer valuable insights into the field of domain generalization, setting a new benchmark for future research in face anti-spoofing

II. Related Work

A. Face Anti-Spoofing

Face anti-spoofing (FAS) has become a focal point in diverse research domains, drawing considerable attention. Initially, researchers explored human behaviors and predefined movements to discern between genuine and spoofed facial presentations. This paved the way for the utilization of handcrafted features such as LBP, HoG [5], [6], SIFT [7], crucial for characterizing spoof patterns. In recent years, deep neural networks have revolutionized FAS, offering a spectrum of methodologies from classification-based approaches to regression-based and generative models. This paradigm shift has empowered FAS systems to discern intricate patterns and nuances in facial data, enhancing their discriminative capabilities against spoof attacks.

B. Domain-generalize FAS

Face anti-spoofing (FAS) has emerged as a prominent research domain, driven by the pursuit of models capable of generalizing effectively to previously unseen domains. Various strategies have been employed to address this challenge. Some approaches hinge on domain adaptation techniques, which necessitate access to target domain data for model adaptation [8]–[10]. In contrast, others center around learning shared features across domains through adversarial training and triplet loss mechanisms. Notably, while conventional wisdom in the field often regards domain-specific signals as detrimental to model performance, our paper takes a pioneering approach by advocating for the explicit utilization of such signals. Through the innovative application of invariant risk minimization in cross-domain FAS, we harness domain-specific information to enhance model robustness and adaptability.

III. Method

The FAS(Face Anti-Spoofing) task primarily involves distinguishing and classifying information such as texture, lighting, and patterns present in incoming images, rather than focusing on facial features. This is because in FAS, both spoof and live samples contain facial characteristics, making it challenging to differentiate them based on facial information. The necessity for facial information arises in the pre-processing stage, particularly in face detection tasks preceding FAS. Therefore, it can be inferred that in the FAS task, minimizing the influence of the face and focusing more on other textures can yield better results. Similarly, to minimize the influence of the domain, it is necessary to have robust features in other textures that are unaffected by the domain. So in this section, we provide a detailed proposal for the domain-generalized FAS task. We have developed a mechanism by integrating Dynamic Kernel [2] and Style Transfer [11], aimed at reducing instance-specific features while extracting more generalized characteristics. This approach establishes a foundation for the FAS system to generalize across various domains, rather than being confined to specific ones. Subsections will be introduced the specific operational mechanisms of the domain-generalized FAS system utilizing Dynamic Kernel and Style Transfer

A. Seperate Content & Style

we denote the facial component as "content," while referring to other elements essential for anti-spoofing as "style." [3] Given the significant differences between content and style features, attempting to discriminate them simultaneously using the same kernel proves challenging. To address this, we employ distinct kernels and normalization techniques within the backbone network to seperate content and style features.

For content features, which remain consistent regardless of spoof or live labels, we utilize batch normalization(BN) to ensure uniform feature extraction irrespective of the label. [1] Conversely, style features, which vary across images, labels, and domains, are extracted using instance normalization(IN) [1] to preserve the distinct characteristics present in each image.

By initially separating features in this manner, we aim to mitigate the influence of content information in the FAS task, allowing for a more focused anti-spoofing analysis

B. Instance & General Feature

Style refers to the distinctive characteristics present in each domain, as well as the common attributes typically found in all spoof or live images. $e.g$, displays may exhibit moiré patterns, whereas A4 documents may not. However, both spoof domains may exhibit color differences distinct from live images. Similarly, different cameras may produce images with varying atmospheres or textures depending on the camera's filter. These domain-specific differences in style are denoted as instance styles, while the average style exhibited by spoof or live images is categorized as general style.

By utilizing a Dynamic Kernel Generator(DKG), a new kernel is created based on the feature input to the Dynamic Kernel Module. In essence, this means that instead of using the same kernel for all images, as in

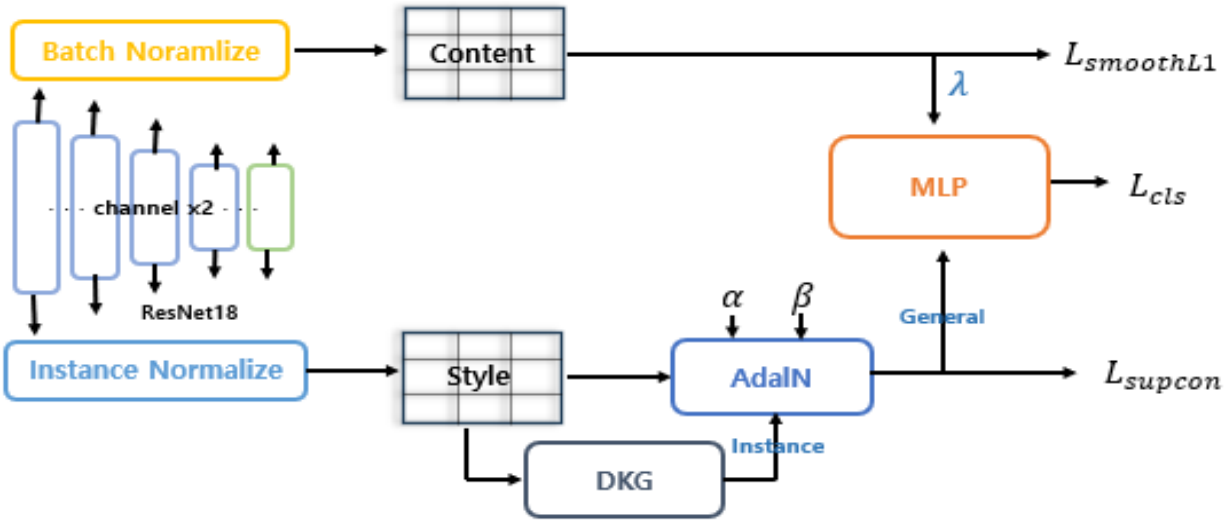

Fig. 1. The overall network architecture comprises three main components. Firstly, a ResNet18 backbone extracts style information from the input image. Subsequently, a Dynamic Kernel module is employed to identify instance features with domain-specific characteristics, ensuring adaptability across domains. Finally, the AdalN module is utilized to remove mean and variance from the features, yielding general features robust to domain variations.

conventional convolution layers to find common features, we can employ uniquely different kernels for each image style. This implies that each image can have its own kernel tailored to its specific style, allowing us to capture the instance style unique to each image. Removing this instance information from the original style effectively eliminates domain-specific details such as the type of present attack or the camera used, facilitating a more straightforward and accurate domain generalization process. Consequently, this approach enables the classification of spoof and live images based on information present in all domains, eliminating the need to separately consider network information for each domain or contemplate the number of hyperplanes required for domain generalization The final step involves utilizing the general features obtained by removing domain-specific instance features from the style information extracted by the backbone network for classification

C. Loss

In our original task, we utilize a Classification Loss for distinguishing between Live and Spoof images. Additionally, we incorporate a smooth L1 Loss to extract consistent content information, namely facial features, present in all images. Lastly, to obtain general features that are similar across all styles and more unique instance features, we employ the SupCon loss

1) Classification Loss: We differentiate between Live and Spoof images using an MLP with the difference between content and general features as input. The reason for excluding content from the general features is that the general features encompass common attributes of the images, which may contain a slight mixture of content information. Thus, by removing content information from the general features, we aim to prevent the

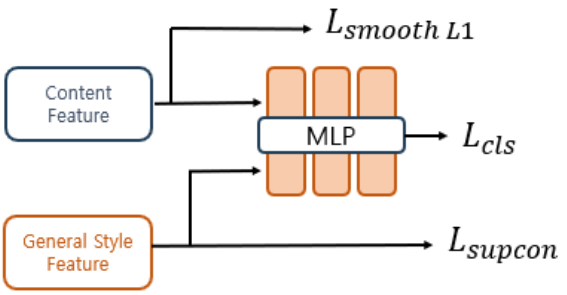

Fig. 2. Diagram illustrating the structured features and losses utilized at the end of the model architecture

classifier from being confused by the content information. Furthermore, to prevent significant loss of general feature information, we apply a λ weight to the content information. We set the $\lambda = 0.2$ to balance the influence of content and general features. The loss function L_{cls} is defined as follows:

$$\hat{y} = \sigma(MLP(f_{general} - \lambda f_{content}))$$
$$L_{cls} = L_{BCE}(\hat{y}, y) \quad (1)$$

2) Smooth L1 Loss: As mentioned earlier, the content feature represents facial information, which should remain consistent for both spoof and live images. To ensure this consistency across spoof and live samples, we utilize the smooth L1 loss to enforce similarity in the content features. This loss function enables us to make the content features identical for both spoof and live images The loss function $L_{smoothL1}$ is defined as follows:

$$L_{smoothL1} = ||C_{live} - C_{spoof}|| \quad (2)$$

3) Supervised Contrastive Loss: To enhance the distinguishability between instance and general features, It is valuable to apply loss functions for style extraction as

a helpful hint, in addition to assigning the task solely to dynamic kernels. Therefore, we utilize the supervised contrastive Loss [4] to encourage the extraction of instance features that are distant from each other, thereby promoting the discovery of more image-specific styles. Meanwhile, for general features, we aim to extract features that are close within the same label and distant across different labels, ensuring the extraction of distinct features suitable for classification. Thus, we adapt the conventional Supervised Contrastive Loss to achieve this objective

Let $s \in S = \{1, ..., N\}, l \in L = \{1, ..., N\}$ denote the index of a spoof and live samples. We define the general feature as z^g and the instance feature as z^i Utilizing these definitions, the equation for the short distance can be formulated as follows

$$D_{pos} = \sum_{s \in S} exp(z_s^g \cdot z_{j(s)}^g / \tau) + \sum_{l \in L} exp(z_s^g \cdot z_{j(s)}^g / \tau) \quad (3)$$

Subsequently, to facilitate the differentiation of general features, we ensure that spoof and live features are distanced from each other, while also arranging for instance features to diverge across all labels.

$$D_{neg} = \sum_{s \in S} \sum_{l \in L} exp(z_s^g \cdot z_l^g / \tau) + \sum_{k \in 2N} exp(z_k^i \cdot z_{j(k)}^i / \tau) \quad (4)$$

So The loss function L_{supcon} is defined as follows:

$$L_{supcon} = -\frac{1}{2N} log(\frac{D_{positive}}{D_{negative}}) \quad (5)$$

The goal of this paper is to enhance the performance of Face Anti-spoofing without being constrained by domains, utilizing a simple structure. Previous works on domain generalization tasks have employed the number of hyperplanes [12] as a hyperparameter or have been limited to specific domains to achieve domain generalization. However, we propose a novel approach where any domain or image input is processed through the Backbone, employing distinct normalization techniques to separate Content and Style. Through the use of Dynamic Kernel and AdalN, we identify features that are invariant to the domain, thereby improving the performance of Classification

Algorithm 1 Domain-Agnostic Face Anti-spoofing

Input: Image I, Backbone Network B, Dynamic Kernel D, Style Transfer AdaLN A
Output: Spoofing prediction P
procedure FAS(I, B, D, A)
    $C, S \leftarrow B(I)$ ▷ Extract Content and Style features using Backbone
    $C_{norm} \leftarrow$ Normalize(C)
    $S_{norm} \leftarrow$ Normalize(S)
    $F_{var} \leftarrow D(C_{norm}, S_{norm})$ ▷ Apply DKG
    $F_{inv} \leftarrow A(S_{norm}, F_{var})$ ▷ AdalN
    $P \leftarrow$ Classify(F_{inv}) ▷ Invariant feature classification
    Compute L_{supcon} using S_{norm} and F and F_{inv} ▷ Supervised contrastive loss
    Compute $L_{smoothL1}$ using C ▷ Smooth L1 loss
    return P
end procedure

IV. Experiments

A. Datasets

To evaluate our proposed method's effectiveness in face anti-spoofing, we utilize four public datasets: CASIA MFSD (C), Idiap Replay-Attack (I), MSU-MFSD (M), and OULU NPU (O). These datasets encompass a wide range of variations in capture devices, attack types, illumination conditions, backgrounds, and demographic diversity, presenting a comprehensive challenge for domain generalization. Adopting a leave-one-out testing protocol, similar to previous domain generalization (DG) FAS approaches, we train our model on three datasets and test on the fourth to assess cross-domain generalization capabilities. $e.g$, under the protocol OCI→M, we train on O, C, and I datasets and evaluate the model's performance on M, ensuring a rigorous and fair comparison with existing methodologies in the field.

B. Experimental Settings

we use ResNet-18 backbone. For train we set the number of epochs to 80. We adopt an SGD optimizer with a learning rate of 1e-2 and reduce the learning rate to 1e-5 through the cosine annealing scheduler. Our input image size is set to 256×256, which cropped using MTCNN and batch size is set to 64. For Content weight λ set 0.2

C. Result

In the evaluation of our Face Anti-Spoofing (FAS) model, we present a comparison against the state-of-the-art (SOTA) methods across four testing domains. As illustrated in Tab. I, our method outperforms the other approaches in ICM to O task. For the ICM to O domain, our model achieved an HTER of 8.60 and an AUC of 97.44, which signifies a notable improvement over SA-FAS model. In the domain transfer from OCI to M, While there was a slight decrease in the Half Total Error Rate (HTER), there was a notable increase of 1% in the AUC.

Upon examining both Fig. 3 and Tab. I, it becomes apparent that our model achieves performance comparable to, or slightly better than, existing state-of-the-art methods. However, a distinct advantage of our approach is the ability to perform binary classification without the need to predetermine the number of domains or deliberate on the types of present attacks. As depicted

Methods	ICM to O		OMI to C		OCI to M		OCM to I	
	HTER(%)↓	AUC(%)↑	HTER(%)↓	AUC(%)↑	HTER(%)↓	AUC(%)↑	HTER(%)↓	AUC(%)↑
MMD-AAE [13]	40.98	63.08	40.98	63.08	40.98	63.08	31.58	75.18
SSDG-M [14]	25.17	81.83	23.11	85.45	16.67	90.47	18.21	94.61
DR-MD-Net [15]	25.02	81.47	19.68	87.43	17.02	90.10	20.87	86.72
RFMeta [16]	16.45	91.16	20.27	88.16	13.89	93.98	17.30	90.48
NAS-FAS [17]	13.80	93.43	16.54	90.18	19.53	88.63	14.51	93.84
SDA [18]	23.10	84.30	24.50	84.40	15.40	91.80	15.60	90.10
DRDG [19]	15.63	91.75	19.05	88.79	12.43	95.81	15.56	91.79
ANRL [20]	15.67	91.90	17.83	89.26	10.83	96.75	16.03	91.04
SSAN-M [3]	19.51	88.17	16.47	90.81	10.42	94.76	14.00	94.58
SA-FAS [12]	10.00	96.23	**8.78**	**95.37**	**5.95**	96.55	6.58	**97.54**
Ours	**8.60**	**97.44**	12.56	94.44	8.81	**97.55**	**6.51**	96.84

TABLE I
COMPARISON TO THE SOTA FACE ANTI-SPOOFING MODELS ON FOUR TESTING DOMAINS.THE BOLD NUMBERS INDICATE THE BEST PERFORMANCE.

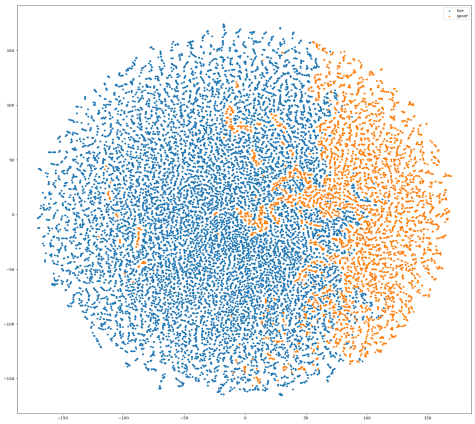

Fig. 3. t-SNE results for OULU NPU (O) from ICM to O. Orange represents spoof while blue represents live.

in Fig. 3, our model operates with a single line, which significant hyperplane that primarily separates spoof and live samples, simplifying the classification process.

In conclusion, the proposed model not only sets a new insight which is not constrained by the quantity of domains in FAS across multiple domains but also exhibits substantial robust to domain shifts

1) Visualize: Fig. 4 serves as a demonstrative example of how our model proficiently segregates content and style. The visualization indicates that, rather than concentrating on the facial features universally, the model is more attentive to the areas beyond the face where the characteristics of Style manifest themselves. This observation substantiates the effectiveness of applying different normalization techniques in the discernment of Style and Content, providing evidence that our approach aids in the separation of these two fundamental aspects.

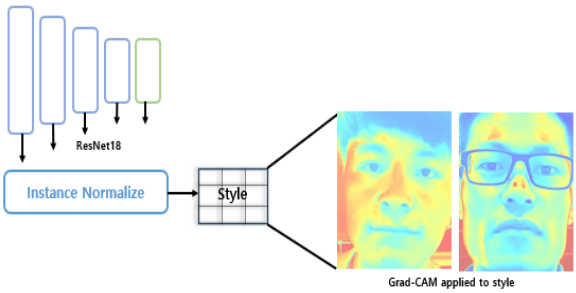

Fig. 4. Applying Grad-CAM to style features on an image sourced from OULU NPU (O) test data to verify focus on non-facial content information

The ability to focus on non-facial attributes suggests that the model could be less susceptible to variations in facial expression or geometry and more robust to changes in environmental conditions, which are represented in the Style features

V. CONCLUSION

In this work, we have proposed a face anti-spoofing model that is not limited by the number of domains. We have proposed a model that, diverging slightly from conventional FAS tasks, diminishes the influence of facial features and focuses more on the textures and other intrinsic characteristics within the image, thereby reducing the impact of facial attributes. Furthermore, we have made concerted efforts to identify and eliminate domain-specific features, reducing the domain influence. Consequently, our model demonstrates strong performance across testing domains with the use of a single and simple hyperplane. we propose offers optimized insights for the field of Domain Generalized Face Anti-Spoofing (DG-FAS) [1], [3], [12]

ACKNOWLEDGMENT

This work was supported by MX Division at Samsung Electronics Co., Ltd.

REFERENCES

[1] Q. Zhou, K.-Y. Zhang, T. Yao, X. Lu, R. Yi, S. Ding, and L. Ma, "Instance-aware domain generalization for face anti-spoofing," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20 453–20 463, 2023.

[2] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11 030–11 039, 2020.

[3] Z. Wang, Z. Wang, Z. Yu, W. Deng, J. Li, T. Gao, and Z. Wang, "Domain generalization via shuffled style assembly for face anti-spoofing," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4123–4133, 2022.

[4] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," *Advances in Neural Information Processing Systems*, pp. 18 661–18 673, 2020.

[5] T. de Freitas Pereira, A. Anjos, J. ´e Mario De Mar-tino, and S. Marce, "Lbp-top based countermeasure against face spoofing attacks," *In Asian Conference on Computer Vision*, p. 121–132, 2012.

[6] J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection with component dependent descriptor," *In 2013 International Conference on Biometrics (ICB)*, p. 1–6, 2013.

[7] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE transactions on information forensics and security*, pp. 2268–2283, 2016.

[8] D. Deb, X. Liu, and A. Jain, "Unified detection of digital and physical face attacks," *In Proceedings of the 17th International Conference on Automatic Face and Gesture Recognition (FG)*, 2023.

[9] X. Guo, Y. Liu, A. Jain, and X. Liu, "Multi-domain learning for updating face anti-spoofing models," *In Computer Vision–ECCV 2022: 17th European Conference*, pp. 23–27, 2022.

[10] H. Li, W. Li, H. Cao, S. Wang, F. Huang, and A. C. Kot, "Unsupervised domain adaptation for face anti-spoofing," *IEEE Transactions on Information Forensics and Security*, pp. 1794–1809, 2018.

[11] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1501–1510, 2017.

[12] Y. Sun, Y. Liu, X. Liu, Y. Li, and W.-S. Chu, "Rethinking domain generalization for face anti-spoofing: Separability and alignment," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24 563–24 574, 2023.

[13] H. Li, S. J. Pan, S. Wang, and A. C. Kot, "Domain generalization with adversarial feature learning," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5400–5409, 2018.

[14] Y. Jia, J. Zhang, S. Shan, and X. Chen, "Single-side domain generalization for face anti-spoofing," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8484–8493, 2020.

[15] G. Wang, H. Han, S. Shan, and X. Chen, "Cross-domain face presentation attack detection via multidomain disentangled representation learning," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6678–6687, 2020.

[16] R. Shao, X. Lan, and P. C. Yuen, "Regularized fine-grained meta face anti-spoofing," *In Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 11 974–11 981, 2020.

[17] Z. Yu, J. Wan, Y. Qin, X. Li, S. Z. Li, and G. Zhao, "Nas-fas: Static-dynamic central difference network search for face anti-spoofing," *IEEE transactions on pattern analysis and machine intelligence*, pp. 3005–3023, 2020.

[18] J. Wang, J. Zhang, Y. Bian, Y. Cai, C. Wang, and S. Pu, "Self-domain adaptation for face anti-spoofing," *In Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 2746–2754, 2021.

[19] S. Liu, K.-Y. Zhang, T. Yao, K. Sheng, S. Ding, Y. Tai, J. Li, Y. Xie, and L. Ma, "Dual reweighting domain generalization for face presentation attack detection," *arXiv preprint arXiv:2106.16128*, 2021.

[20] S. Liu, K.-Y. Zhang, T. Yao, M. Bi, S. Ding, J. Li, F. Huang, and L. Ma, "Adaptive normalized representation learning for generalizable face anti-spoofing," *In Proceedings of the 29th ACM International Conference on Multimedia*, pp. 1469–1477, 2021.