

# DOMAIN-GENERALIZED FACE ANTI-SPOOFING WITH UNKNOWN ATTACKS

Zong-Wei Hong<sup>1</sup>, Yu-Chen Lin<sup>1</sup>, Hsuan-Tung Liu<sup>2</sup>, Yi-Ren Yeh<sup>3</sup>, Chu-Song Chen<sup>1</sup>

<sup>1</sup>National Taiwan University

<sup>2</sup>E.SUN Financial Holding Co., Ltd.

<sup>3</sup>National Kaohsiung Normal University

## ABSTRACT

Although face anti-spoofing (FAS) methods have achieved remarkable performance on specific domains or attack types, few studies have focused on the simultaneous presence of domain changes and unknown attacks, which is closer to real application scenarios. To handle domain-generalized unknown attacks, we introduce a new method, DGUA-FAS<sup>1</sup>, which consists of a Transformer-based feature extractor and a synthetic unknown attack sample generator (SUASG). The SUASG network simulates unknown attack samples to assist the training of the feature extractor. Experimental results show that our method achieves superior performance on domain generalization FAS with known or unknown attacks.

**Index Terms**— face anti-spoofing, open set recognition.

## 1. INTRODUCTION

With the widespread application of online payment, mobile device, and access control, FAS technology has received extensive attention. However, when deploying FAS modules in real scenarios, we often need to tackle domain changes caused by different image sensors and photographing environments. Besides, new attack types can occur and render well-trained systems ineffective. It is thus crucial to conduct FAS methods that are robust to unseen domains and novel attacks.

Although traditional deep learning methods [1, 2] achieve great progress on specific datasets and protocols, they do not perform well on unseen domains and unknown attacks. Domain adaptation [3, 4, 5] and generalization techniques [6, 7, 8, 9, 10, 11] have been developed to handle domain gaps. For handling novel attacks, zero/few-shot learning [12, 13] and anomaly detection [14, 15] are used. However, tackling the simultaneous domain and attack changes is still challenging.

In this paper, we propose a method combining transformer-based architecture and open-set feature generator to simulate unknown attack samples scattered in the feature space. Our approach achieves favorable performance in several unseen domain and unknown attack settings. To our knowledge, this is the first time to introduce an open set feature generator as an attack sample generator for FAS. Our design boosts the performance of the state-of-the-art transformer-based FAS model

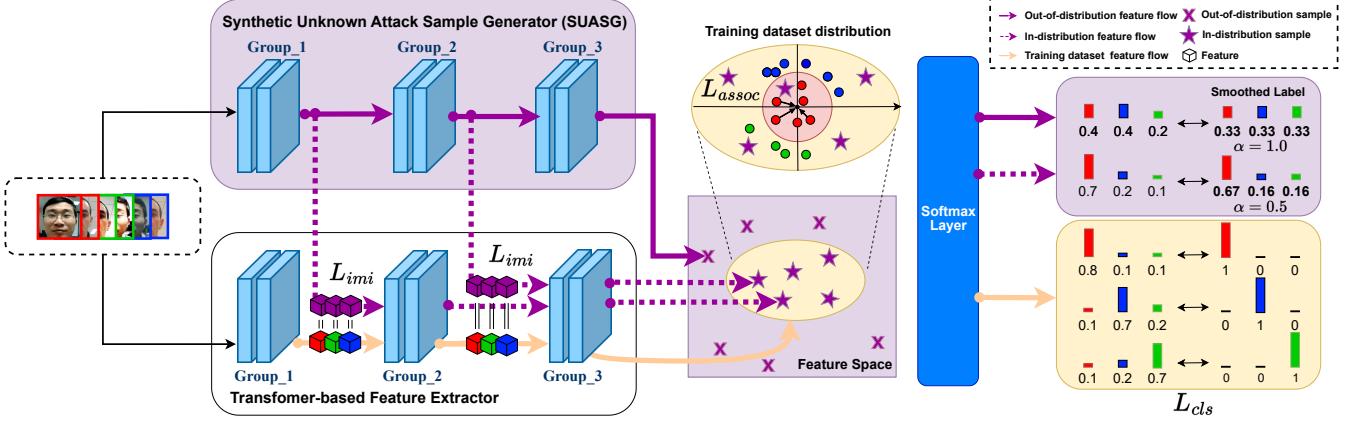
and we demonstrate that it is also effective based on other model backbones, indicating that our design can address the unseen domain and novel attacks for FAS effectively.

## 2. RELATED WORK

**Handling Cross Domain in FAS.** Compared with domain adaptation, domain generalization is more practical since no unseen domain data are needed in the training phase. A typical technique is to extract domain-independent features for classification. The approach in [6] lets the real features from different domains be indistinguishable by single-side adversarial training, while spoofing features from each domain separate individually and all are far from real features through triplet learning. In [7], two feature extractors are trained for content and style respectively, while content features are domain-invariant and the classification relies on style features. The method of [8] formulates FAS as a patch-level classification problem which can utilize local features better. Patch features from the same image after non-distorted augmentation are forced to be invariant by the similarity loss, while the asymmetric angular margin softmax loss is used to enforce a larger margin with real-face features. The approach of [11] extends the ideas of [6]. It uses concentration loss to centralize real face features at the origin and aggregate features of the same attack type from different domains.

**Handling Unknown Attacks in FAS.** To tackle unknown attacks, previous studies introduce zero/few-shot learning and anomaly detection for FAS. In [12], without semantic information of unknown attack types, the approach uses a deep tree structure to describe the known attacks and directly classifies the unknown attacks to the closest known attack type. The method in [16] models the training distribution (including real and known attacks) by Gaussian Mixture Model (GMM), and draws samples out of the training distribution as unknown attack samples. As for anomaly detection, it assumes that real samples share more compact features than spoofs, and thus images can be classified by a one-class classifier instead of a traditional binary classifier, indicating that all the instances that do not belong to the only class are considered as attacks. Following this concept, a pair-wise one-class contrastive loss is used in [15] to build a one-class GMM. In addition, some

<sup>1</sup>[https://github.com/AI-Application-and-Integration-Lab/DGUA\\_FAS](https://github.com/AI-Application-and-Integration-Lab/DGUA_FAS)



**Fig. 1.** Overall architecture of our proposed method (illustrated with 2 known attacks). Each training image with different types (*Red* for real, *Blue* for type-1, and *Green* for type-2 attacks) is fed into both the SUASG and the Transformer-based Feature Extractor. There are two types of synthetic samples generated from each input, in-distribution samples (*Purple Star*) and out-of-distribution samples (*Purple Cross*). In-distribution samples are forced to be similar to those generated from corresponding groups of Feature Extractor by  $L_{imi}$ . Finally, we use **Smoothed Label** as the objective of the output probabilities of synthetic samples; On the other side, real features generated by Feature Extractor are clustered at the origin of feature space by  $L_{assoc}$ , while features with the same attack type in different domains are forced to be in the same class by  $L_{cls}$ .

studies formulate the unknown attack classification as an open set problem. Eg., in [17], Extreme Value Theorem (EVT) is adopted to detect unknown attacks. In our work, rather than using constraints for detection, we directly simulate spoofs as our open set data to provide more favorable results.

### 3. METHODOLOGY

We give the problem definition and then describe our method.

#### 3.1. Problem Formulation

Suppose we have  $M$  domains  $\mathcal{D} = \{\mathcal{D}_{1:M}\}$  of the FAS training datasets; each dataset owns  $|\mathcal{A}_i|$  types of attacks ( $\mathcal{A}_{i=1:M}$ ) and one real face category  $\mathcal{F}_i$ . In the training phase, the FAS predictor is learned by using only the training data of these known domains and types of attacks.

In the inference (testing) phase, the learned predictor is applied to some new domain (dataset) which contains all the original attacks  $\bigcup \mathcal{A}_{i=1:M}$  and new types of attacks  $\mathcal{A}_{\text{unknown}}$ . The purpose is to classify the input face as *real* or *spoof*, no matter the spoof data come from known or unknown attacks.

#### 3.2. Domain Generalized Unknown Attacks

Our method (namely DGUA-FAS) handles both domain generalization and unknown attacks for FAS.

**Cross Domain.** Classification loss ( $L_{cls}$ ) and association loss ( $L_{assoc}$ ) are the two main learning objectives for handling cross domain samples in DGUA-FAS. We enforce the samples of the same attack type to have similar (or different types

to have dissimilar) embeddings regardless of their domains. Unlike the setting in [6] which encourages the same attacks in different domains to be separated either, we simply use the common multi-class cross-entropy loss for  $L_{cls}$ . Suppose we have  $K$  types of attacks in the domain union  $\bigcup \mathcal{A}_{i=1:M}$ , the loss  $L_{cls}$  separates the data into  $K + 1$  classes (Real face and  $K$  attacks), no matter which domains the samples belong to.

In addition to separating the real and attack faces, we further enhance the data association among different domains. Since the embeddings of real faces should be irrelevant to domains variations, we assume that all real samples from different datasets are similar and compact in the feature space. Our real-face association loss follows [11], which is designed as

$$L_{assoc} = \frac{1}{|F_{\text{real}}|} \sum_{f \in F_{\text{real}}} \|f\|_1, \quad \forall f \in F_{\text{real}}, \quad (1)$$

where  $F_{\text{real}}$  is the set of real-face feature embeddings, and they are enforced to concentrate at the origin of feature space.

DGUA-FAS is trained with the loss  $L_{\text{main}} = L_{cls} + \lambda \cdot L_{assoc}$  together with some other complementary loss terms that are introduced below. Without loss of generality, we use a transformer as the backbone network (yet CNNs can be used as well), which is shown as the Transformer-based Feature Extractor in Fig. 1 (lower-left part).

**Unknown Attack.** We have the data only in the known attacks  $\mathcal{A}_{i=1:M}$  for training but do not have the unknown attack data in  $\mathcal{A}_{\text{unknown}}$ . To make our model capable of handling open set samples, an intuitive idea is to add some simulated unknown-attack samples in the training process. In [16], only the samples out of the training dataset distribution are produced as novel attacks. However, some unknown attacks

could also be overlapped with the training distribution. Ignoring the possibility of the within-distribution similarity often leads to unreliable predictions. Since our approach learns a classifier of  $K + 1$  classes for prediction, we hope that the simulated samples of unknown attacks can be evenly distributed over the  $K + 1$  regions. E.g., considering  $K = 2$  attack types in Fig. 1, we would hope the output probability of the unknown attacks to be approximately 0.33 for the 3 classes (upper-right part). Hence, we simulate not only the out-of-distribution but also the in-distribution data.

To achieve this, we leverage the different layers in a network to generate samples of different difficulty levels, so that they can scatter both inside and outside of the known-classes distribution. Inspired by DiAS [18], we complement an architecture called Synthetic Unknown Attack Sample Generator (SUASG) to produce synthetic unknown samples for FAS during training (upper-left of Fig. 1). SUASG shares the same structure as the Transformer-based Feature Extractor but has its own weights. To train SUASG, We first divide both networks into three groups. Each training input image is fed into both networks. Between the corresponding groups, we add the following loss (called *imitation loss*) to constrain the features generated by the two networks:

$$\mathbb{L}_{imi} = \frac{1}{N} \sum_{i=1}^N \sum_{g=1}^2 \|f_{SUASG}^{g,i} - f_{extract}^{g,i}\|_1, \quad (2)$$

with  $N$  the mini-batch size;  $f_{SUASG}^{g,i}$  and  $f_{extract}^{g,i}$  are the features of the  $i$ -th image extracted by the  $g$ -th groups of SUASG and Transformer-based Feature Extractor, respectively. To train the network, an input image  $i$  passes through a total of  $G$  paths ( $G = 3$  groups in our case). One is the path passing through only SUASG. Each of the remaining paths goes through the 1st to the  $g$ -th groups of SUASG and then the  $(g + 1)$ -th to the  $G$ -th groups of the Transformer-based Feature Extractor ( $1 \leq g < G$ ). As for the path through SUASG only, the generated samples are not constrained by their final group, which we consider to be a simulation of out-of-distribution samples. When  $1 \leq g < G$ , the samples produced are constrained by the imitation losses and then share the same final groups of Transformer-based Feature Extractor. We then use them to simulate the in-distribution data. Hence, both the out-of-distribution and in-distribution samples of unknown classes are generated. We empirically find that it achieves favorable results in dealing with unknown attacks more efficiently.

**Table 1.** Datasets used in our evaluation. (i) and (v) denote ‘images’ and ‘videos’, respectively.

Dataset	Number of videos/images	Attack type
OULU-NPU[19]	3600 (v)	print, replay
CASIA-FASD[20]	600 (v)	print, replay
MSU-MFSD[21]	280 (v)	print, replay
Replay-Attack [22]	1200 (v)	print, replay
CelebA-Spoof [23]	625537 (i)	print, replay, paper mask
WMCA [24]	1679 (v)	print, replay, Partial (glasses), Mask (plastic, silicone, and paper, Mannequin)

We implement the training process by iterating the fol-

lowing two steps. In the first step, we fix the Transformer-based Feature Extractor and train SUASG by the imitation loss (Eq. 2) and  $\mathbb{L}_{cls}$ . In the second step, we fix SUASG and train the Transformer-based Feature Extractor by minimizing

$$\mathbb{L}_{extract} = \mathbb{L}_{main} + \mathbb{L}_{cls}(x_{SID}, \hat{y}_{SID}) + \mathbb{L}_{cls}(x_{SOOD}, \hat{y}_{SOOD}), \quad (3)$$

where  $x_{SID}$  and  $x_{SOOD}$  are the output probabilities of the synthetic in-distribution and out-of-distribution samples, respectively, and  $\hat{y}_{SID}$  and  $\hat{y}_{SOOD}$  are their respective target probabilities.  $\mathbb{L}_{cls}$  is the multi-class cross-entropy loss. As mentioned, we hope the unknown-class samples to be equally distributed as  $[1/(K + 1)]$  over the  $K + 1$  classes. On the other hand, as the simulated data are synthesized from the original training data, the original labels could serve as the pseudo labels for the synthesized data. We thus smooth them to form the target probabilities of the synthetic data as suggested by [18]:

$$\hat{y} = (1 - \alpha) \cdot y + \alpha/(K + 1) \cdot \mathbf{u}, \quad (4)$$

with  $y$  the original label of the input image  $x$ , and  $\mathbf{u}$  is the all-one vector. A higher  $\alpha$  is set for the out-of-distribution synthetic data and vice versa. When training is finished, in the inference stage, only the Transformer-based Feature Extractor together with the final classification layer are used for testing.

## 4. EXPERIMENTS

To verify our DGUA-FAS, we first examine its performance on the domain-generalized problem. Then, we present the results on the problem of unseen domains with both known and unknown attacks.

**Datasets and Metrics.** The datasets CASIA-FASD [20] (C), MSU-MFSD [21] (M), Idiap Replay-attack [22] (I), OULU-NPU [19] (O), CelebA-Spoof [23], and WMCA[24] are used to evaluate our method. The first four (C&M&I&O) include only print and replay attacks, while the rest (i.e., CelebA-Spoof & WMCA) contain more diverse attack types such as glasses, silicone masks, and paper masks (Table 1). Following previous works, we utilize the Half Total Error Rate (HTER) and the Area Under Curve (AUC) as the evaluation metrics.

**Implementation detail.** We use MobileViT-S as our backbone network. The version we use is [25] and the model is pre-trained on ImageNet-1K. We choose Adam optimizer and let the learning rate and weight decay parameter to be  $10^{-4}$  and  $10^{-6}$ . We conduct all experiments on a single RTX 3090 GPU and set  $\lambda = 1.0$ ,  $\alpha = 0.5$  and  $\alpha = 1.0$  for in-distribution and out-of-distribution samples respectively. To divide the SUASG network and yield the synthetic samples, we consider  $conv_1$  and  $layer_1$  as the first group,  $layer_2$  and  $layer_3$  as the second group, and  $layer_4$ ,  $layer_5$  as the third group.

**Results on Domain-generalized Settings.** First, we show the results of DGUA-FAS on the traditional domain-generalized settings, where the unseen domain has no new types of attacks. The purpose is to verify that our method, though can

**Table 2.** AUC (%) of the proposal method and previous domain-generalized methods on leave-one-out Setting. The best results are **bolded**, and the second best is underlined.

Method	O & C & I to M	O & M & I to C	O & C & M to I	I & C & M to O	Average AUC (%)
SDDG-R [6]	97.17	95.94	96.59	91.54	95.31
SSAN-R [7]	<u>98.75</u>	96.67	96.79	93.63	96.46
PatchNet [8]	98.46	94.58	95.67	<u>95.07</u>	95.945
HFN+MP [9]	97.28	96.09	90.67	94.26	94.575
CIFAS [10]	96.32	95.30	97.24	93.44	95.575
DiVT-M [11]	<b>99.14</b>	<u>96.92</u>	<b>99.29</b>	94.04	<u>97.347</u>
Proposed approach	98.156	<b>97.0</b>	<u>99.187</u>	<b>96.369</b>	<b>97.678</b>

handle new attacks on new domains, still perform well on the standard domain-generalized scenario without sacrificing the performance. As shown in Table 2 (leave-one-out setting) and Table 3 (limited-source setting) on the C&M&I&O datasets, our method is competitive with the SOTA methods even though we have used further synthetic unknown attack samples during training. Table 3 shows that our performance even exceeds those of the SOTA methods. We speculate that, in this case, synthetic samples enrich limited training data by augmenting the training distribution to better fit the testing distribution, thus leading to more favorable results.

In addition, we also change the backbone of our approach to CNNs and examine its performance. We use ResNet18 (also used in SSDG [6]) to build our model, and the main losses follow those in the settings of [6]. We use layers 1-5 as the first, 6-13 as the second, and the rest as the final groups in ResNet18. As shown in Table 3, our method significantly improves SSDG in the limited-source setting, revealing that our framework can work well with both transformer and CNN-based models.

**Table 3.** The experimental results on the comparison between our method and SOTA methods on the limited-source setting.

Method	M&I to C		M&I to O	
	HTER (%)	AUC (%)	HTER (%)	AUC (%)
SSAN-R [7]	25.56	83.89	24.44	82.86
HFN+MP [9]	30.89	72.48	20.94	86.71
CIFAS [10]	22.67	83.89	24.63	81.48
DiVT-M [11]	20.11	86.71	23.61	85.73
SDDG-R [6]	19.86	86.46	27.92	78.72
Proposed approach w/ setting of SSDG [6]	<b>18.667</b>	<b>87.089</b>	<u>20.139</u>	<u>87.523</u>
Proposed approach	19.222	86.806	<b>20.052</b>	<b>88.746</b>

**Results on Unseen Domain with Unknown Attacks.** Previous works mainly tackle the scenario of new attacks on the same domain. There are still very few studies ([4, 10]) which have reported the results on the cross-domain FAS with both known and unknown types of attacks in the testing phase. Following the setting in [4, 10], we compare our method with them on training with M&C&O and testing on CelebA-Spoof datasets. We also do the experiments on training with O&M&I&C and testing on WMCA datasets, and reproduce the recent domain-generalized method [11] on both settings for comparison. Both setups of M&C&O to CelebA-Spoof and O&M&I&C to WMCA contain only print and replay attacks during training, while unknown attacks like glasses and masks appear in the testing phase additionally.

The results are shown in Table 4. As can be seen, our

**Table 4.** Results on new domain with novel attacks. (\*) the method uses further the unlabeled data in the new domain.

Method	M&C&O to CelebA-Spoof		I&M&C&O to WMCA	
	HTER (%)	AUC (%)	HTER (%)	AUC (%)
*SDA-FAS [4]	<b>18.9</b>	<b>90.9</b>	-	-
CIFAS [10]	24.6	83.2	-	-
DiVT-M [11]	25.075	82.335	<u>22.364</u>	86.816
Proposed approach	<u>21.442</u>	<u>86.351</u>	<b>20.624</b>	<b>88.071</b>

**Table 5.** Ablation study on using different levels of synthetic samples on the leave-one-out domain generalized setting.

In-distribution samples	Out-of-distribution samples	Average	
		HTER (%)	AUC (%)
-	-	9.28325	96.48425
v	-	7.97525	97.3725
-	v	8.4375	97.07925
v	v	<b>7.1</b>	<b>97.678</b>

method outperforms CIFAS [10] on the M&C&O to CelebA-Spoof setting, but performs worse than [4]. It is because [4] is a domain-adaptation rather than a domain-generalized approach; a lot of unlabeled data in the new domain are collected and allowed for training in [4]. However, in real scenarios, we could not gather so much new-attack data for building our model in practice. As for the comparison on domain generalization with unknown attacks, our method performs more favorably than [10] and [11], revealing that the proposed solution is effective to tackle the domain generalized FAS containing both known and unknown attacks.

**Ablation Study.** We conduct the experiments on the effectiveness of in-distribution samples and out-of-distribution samples. The results of using (or not using) in-distribution and out-of-distribution synthetic samples are reported in Table 5. The results reveal that the simulated samples can effectively improve the performance alone, and the best performance is achieved when using both.

## 5. CONCLUSION

We introduce DGUA-FAS, a simple yet effective approach to handle FAS problems with both known and unknown attacks in a new domain. Our method centralizes the real face and separates the attack types in the embedding space regardless of the domains. It leverages a simulated-data generator to produce in-distribution and out-of-distribution samples of the unknown classes for training. Experimental results show that DGUA-FAS is not only effective for traditional domain-generalized FAS but can also achieve the most favorable performance on the FAS tasks with additional unknown attacks in the new domain. In the future, we plan to extend our method to video-based approaches.

**Acknowledgement.** This work was supported in part by E.SUN Financial Holding, and National Science and Technology Council in Taiwan (NSTC 111-2634-F-006-012, NSTC 111-2634-F-002-023). Computational and storage resources are partly supported from NCHC of NARLabs in Taiwan.

## 6. REFERENCES

- [1] Haonan Chen, Guosheng Hu, Zhen Lei, Yaowu Chen, Neil M Robertson, and Stan Z Li, “Attention-based two-stream convolutional networks for face spoofing detection,” *IEEE TIFS*, vol. 15, pp. 578–593, 2019.
- [2] Zitong Yu, Chenxu Zhao, Zezheng Wang, Yunxiao Qin, Zhuo Su, Xiaobai Li, Feng Zhou, and Guoying Zhao, “Searching central difference convolutional networks for face anti-spoofing,” in *CVPR*, 2020, pp. 5295–5305.
- [3] Haoliang Li, Wen Li, Hong Cao, Shiqi Wang, Feiyue Huang, and Alex C Kot, “Unsupervised domain adaptation for face anti-spoofing,” *IEEE TIFS*, vol. 13, no. 7, pp. 1794–1809, 2018.
- [4] Yuchen Liu, Yabo Chen, Wenrui Dai, Mengran Gou, Chun-Ting Huang, and Hongkai Xiong, “Source-free domain adaptation with contrastive domain alignment and self-supervised exploration for face anti-spoofing,” in *ECCV*. Springer, 2022, pp. 511–528.
- [5] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Ran Yi, Kekai Sheng, Shouhong Ding, and Lizhuang Ma, “Generative domain adaptation for face anti-spoofing,” in *ECCV*. Springer, 2022, pp. 335–356.
- [6] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen, “Single-side domain generalization for face anti-spoofing,” in *CVPR*, 2020, pp. 8484–8493.
- [7] Zhuo Wang, Zezheng Wang, Zitong Yu, Weihong Deng, Jiahong Li, Tingting Gao, and Zhongyuan Wang, “Domain generalization via shuffled style assembly for face anti-spoofing,” in *CVPR*, 2022, pp. 4123–4133.
- [8] Chien-Yi Wang, Yu-Ding Lu, Shang-Ta Yang, and Shang-Hong Lai, “Patchnet: A simple face anti-spoofing framework via fine-grained patch recognition,” in *CVPR*, 2022, pp. 20281–20290.
- [9] Rizhao Cai, Zhi Li, Renjie Wan, Haoliang Li, Yongjian Hu, and Alex C Kot, “Learning meta pattern for face anti-spoofing,” *IEEE TIFS*, vol. 17, pp. 1201–1213, 2022.
- [10] Yuchen Liu, Yabo Chen, Wenrui Dai, Chenglin Li, Junni Zou, and Hongkai Xiong, “Causal intervention for generalizable face anti-spoofing,” in *ICME*. IEEE, 2022, pp. 01–06.
- [11] Chen-Hao Liao, Wen-Cheng Chen, Hsuan-Tung Liu, Yi-Ren Yeh, Min-Chun Hu, and Chu-Song Chen, “Domain invariant vision transformer learning for face anti-spoofing,” in *WACV*, 2023, pp. 6098–6107.
- [12] Yaojie Liu, Joel Stehouwer, Amin Jourabloo, and Xiaoming Liu, “Deep tree learning for zero-shot face anti-spoofing,” in *CVPR*, 2019, pp. 4680–4689.
- [13] Yunxiao Qin, Chenxu Zhao, Xiangyu Zhu, Zezheng Wang, Zitong Yu, Tianyu Fu, Feng Zhou, Jingping Shi, and Zhen Lei, “Learning meta model for zero-and few-shot face anti-spoofing,” in *AAAI*, 2020, vol. 34, pp. 11916–11923.
- [14] Shervin Rahimzadeh Arashloo, Josef Kittler, and William Christmas, “An anomaly detection approach to face spoofing detection: A new formulation and evaluation protocol,” *IEEE access*, vol. 5, pp. 13868–13882, 2017.
- [15] Anjith George and Sébastien Marcel, “Learning one class representations for face presentation attack detection using multi-channel convolutional neural networks,” *IEEE TIFS*, vol. 16, pp. 361–375, 2020.
- [16] Mohammad Rostami, Leonidas Spinoulas, Mohamed Hussein, Joe Mathai, and Wael Abd-Almageed, “Detection and continual learning of novel face presentation attacks,” in *ICCV*, 2021, pp. 14851–14860.
- [17] Xin Dong, Hao Liu, Weiwei Cai, Pengyuan Lv, and Zekuan Yu, “Open set face anti-spoofing in unseen attacks,” in *ACM MM*, 2021, pp. 4082–4090.
- [18] WonJun Moon, Junho Park, Hyun Seok Seong, Cheol-Ho Cho, and Jae-Pil Heo, “Difficulty-aware simulator for open set recognition,” in *ECCV*. Springer, 2022, pp. 365–381.
- [19] Zinelabidine Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid, “Oulu-npu: A mobile face presentation attack database with real-world variations,” in *FG 2017*. IEEE, 2017, pp. 612–618.
- [20] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z. Li, “A face antispoofing database with diverse attacks,” in *ICB*, March 2012, pp. 26–31.
- [21] Di Wen, Hu Han, and Anil K Jain, “Face spoof detection with image distortion analysis,” *IEEE TIFS*, vol. 10, no. 4, pp. 746–761, 2015.
- [22] Ivana Chingovska, André Anjos, and Sébastien Marcel, “On the effectiveness of local binary patterns in face anti-spoofing,” in *BIOSIG*. IEEE, 2012, pp. 1–7.
- [23] Yuanhan Zhang, Zhenfei Yin, Yidong Li, Guojun Yin, Junjie Yan, Jing Shao, and Ziwei Liu, “Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations,” in *ECCV*, 2020.
- [24] Anjith George, Zohreh Mostaani, David Geissenbuhler, Olegs Nikisins, André Anjos, and Sébastien Marcel, “Biometric face presentation attack detection with multi-channel convolutional neural network,” *IEEE TIFS*, vol. 15, pp. 42–55, 2019.
- [25] Sachin Mehta and Mohammad Rastegari, “Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer,” in *ICLR*, 2022.