Multi-domain Learning for Updating Face Anti-spoofing Models

Xiao Guo(✉), Yaojie Liu, Anil Jain, and Xiaoming Liu

Michigan State University, Michigan, USA
{guoxia11,liuyaoj1,jain,liuxm}@cse.msu.edu

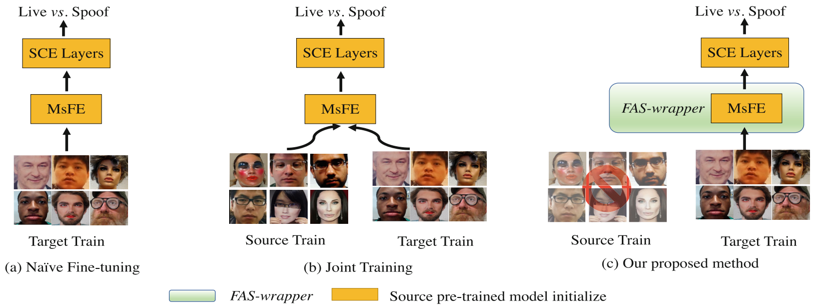
Abstract. In this work, we study multi-domain learning for face anti-spoofing (MD-FAS), where a pre-trained FAS model needs to be updated to perform equally well on both source and target domains while only using target domain data for updating. We present a new model for MD-FAS, which addresses the forgetting issue when learning new domain data, while possessing a high level of adaptability. First, we devise a simple yet effective module, called spoof region estimator (SRE), to identify spoof traces in the spoof image. Such spoof traces reflect the source pre-trained model's responses that help upgraded models combat catastrophic forgetting during updating. Unlike prior works that estimate spoof traces which generate multiple outputs or a low-resolution binary mask, SRE produces one single, detailed pixel-wise estimate in an unsupervised manner. Secondly, we propose a novel framework, named FAS-wrapper, which transfers knowledge from the pre-trained models and seamlessly integrates with different FAS models. Lastly, to help the community further advance MD-FAS, we construct a new benchmark based on SIW, SIW-Mv2 and Oulu-NPU, and introduce four distinct protocols for evaluation, where source and target domains are different in terms of spoof type, age, ethnicity, and illumination. Our proposed method achieves superior performance on the MD-FAS benchmark than previous methods. Our code is available at https://github.com/CHELSEA234/Multi-domain-learning-FAS.

1 Introduction

Face anti-spoofing (FAS) comprises techniques that distinguish genuine human faces and faces on spoof mediums [6], such as printed photographs, screen replay, and 3D masks. FAS is a critical component of the face recognition pipeline that ensures only genuine faces are being matched. As face recognition systems are widely deployed in real world applications, a laboratory-trained FAS model is often required to deploy in a new target domain with face images from novel camera sensors, ethnicities, ages, types of spoof attacks, *etc.*, which differ from the source domain training data in the laboratory.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-19778-9_14.

S. Avidan et al. (Eds.): ECCV 2022, LNCS 13673, pp. 230–249, 2022.
https://doi.org/10.1007/978-3-031-19778-9_14



(a) Naïve Fine-tuning (b) Joint Training (c) Our proposed method

Fig. 1. We study multi-domain learning face anti-spoofing (MD-FAS), in which the model is trained only using target domain data. We first derive the general formulation of FAS models, which contains Spoof Cue Estimate Layers (SCE layers) and multi-scale feature extractor (MsFE). Based on these two components, we propose *FAS-wrapper* that can be adopted for any FAS models, as depicted in (c). (a) and (b) represent the naive fine-tuning and joint training.

In the presence of a large domain-shift [23,50,58] between the source and target domain, it is necessary to employ new target domain data for updating the pre-trained FAS model, in order to perform well in the new test environment. Meanwhile, the source domain data might be inaccessible during updating, due to data privacy issues, which happens more and more frequently for Personally Identifiable Information (PII). Secondly, the FAS model needs to be evaluated jointly on source and target domains, as spoof attacks should be detected regardless of which domain they originate from. Motivated by these challenges, the goal of this paper is to answer the following question:

How can we update a FAS model using only target domain data, so that the upgraded model can perform well in both the source and target domains?

We define this problem as multi-domain learning face anti-spoofing (MD-FAS), as depicted in Fig. 1. Notably, Domain Adaptation (DA) works [14,27,31,38,51] mainly evaluate on the target domain, whereas MD-FAS requires a joint evaluation. Also, MD-FAS is related to Multiple Domain Learning (MDL) [17,44,45], which aims to learn a universal representation for images in many generic image domains, based on one unchanged model. In contrast, MD-FAS algorithm needs to be model-agnostic for the deployment, which means the MD-FAS algorithm can be tasked to update FAS models with various architectures or loss functions. Lastly, the source domain data is unavailable during the training in MD-FAS, which is different from previous domain generalization methods in FAS [20,36,42,53] or related manipulation detection problems [5].

There are two main challenges in MD-FAS. First, the source domain data is unavailable during the updating. As a result, MD-FAS easily suffers from the long-standing *catastrophic forgetting* [25] in learning new tasks, gradually degrading source domain performance. The most common solution [12,22,29] to such a forgetting issue is to use logits and class activation map (grad-CAM) [52] restoring prior model responses when processing the new data. However, due to the increasingly sophisticated spoof image, using logits and grad-CAM empirically fail to precisely pinpoint spatial pixel locations where spoofness occurs, unable to uncover the decision making behind the FAS model. To this end, we propose a simple yet effective module, namely *spoof region estimator* (*SRE*), to identify the spoof regions given an input spoof image. Such spoof traces serve as responses of the pre-trained model, or better replacement to logits and activation maps in the MD-FAS scenario. Notably, unlike using multiple traces to pinpoint spoofness or manipulation in image [31,69], or low-resolution binary mask as manipulation indicator [10,33,64], our *SRE* offers a single and high-resolution detailed binary mask representing pixel-wise spatial locations of spoofness. Also, many anti-forgetting algorithms [8,13,40,46,49,54] usually require extra memory for restoring exemplar samples or expanding the model size, which makes them inefficient in real-world situations.

Secondly, to develop an algorithm with a high level of adaptability, it is desirable to keep original FAS models intact for the seamless deployment while changing the network parameters. Unlike methods proposed in [44,45] that specialize on the certain architecture (*e.g.*, ResNet), we first derive the general formulation after studying FAS models [30,34,37,53,63,65], then based on such a formulation we propose a novel architecture, named *FAS-wrapper* (depicted in Fig. 2), which can be deployed for FAS models with minimum changes on the architecture.

In summary, this paper makes the following contributions:

◇ Driven by the deployment in real-world applications, we define a new problem of MD-FAS, which requires to update a pre-trained FAS model only using target domain data, yet evaluate on both source and target domains. To facilitate the MD-FAS study, we construct the FASMD benchmark, based on existing FAS datasets [7,34,36], with four evaluation protocols.

◇ We propose a *spoof region estimator* (*SRE*) module to identify spoof traces in the input image. Such spoof traces serve as the prior model's responses to help tackle the *catastrophic forgetting* during the FAS model updating.

◇ We propose a novel method, *FAS-wrapper*, which can be adopted by any FAS models for adapting to target domains while preserving the source domain performance.

◇ Our method demonstrates superior performance over prior works, on both source and target domains in the FASMD benchmark. Moreover, our method also generalizes well in the cross-dataset scenario.

Table 1. We study the multi-domain learning face anti-spoofing, which is different to prior works.

Paradigm	Method	Source free	Learning new domain	Joint evaluation	Model agnostic	Anti-forgetting mechanism
Face anti-spoofing domain learning	SSDG [20]	✘	✔	✔	✔	N/A
	MADDoG [53]	✘	✔	✔	✔	N/A
	FSDE-FAS [61]	✘	✔	✔	✔	N/A
Anti-forgetting learning	EWC [25]	✔	✘	✔	✔	Prior-driven
	iCaRL [46]	✔	✘	✔	✔	Replay
	MAS [4]	✔	✘	✔	✔	Prior-driven
	LwF [29]	✔	✘	✔	✔	Data-driven (class prob.)
	LwM [12]	✔	✘	✔	✔	Data-driven (feat. map)
Multi-domain learning	DAN [14]	✘	✔	✘	✔	N/A
	OSBP [51]	✘	✔	✘	✔	N/A
	STA [31]	✘	✔	✘	✔	N/A
	CIDA [27]	✔	✔	✘	✘	N/A
	Seri. Adapter [44]	✔	✔	✔	✘	N/A
	Para. Adapter [45]	✔	✔	✔	✘	N/A
Multi-domain learning face anti-spoofing	*FAS-wrapper* (Ours)	✔	✔	✔	✔	Data-driven (spoof region)

2 Related Works

Face Anti-spoofing Domain Adaptation. In Domain Adaption (DA) [14, 27, 31, 38, 51], many prior works assume the source data is accessible, but in our setup, source domain data is unavailable. The DA performance evaluation is biased towards the target domain data, as source domain performance may deteriorate, whereas FAS models need to excel on both source and target domain data. There are some FAS works that study the cross-domain scenario [20, 36, 43, 53, 56, 59, 61]. [61] is proposed for the scenario where source and a few labeled new domain data are available, with the idea to augment target data by style transfer [62]. [53] learns a shared, indiscriminative feature space without the target domain data. Besides, [20] constructs a generalized feature space that has a compact real faces feature distribution in different domains. [36] also works on unseen domain generalization. But the same as the other works, the new domain is not based on bio-metric patterns (*i.e.*, age). Being orthogonal to prior works, the source domain data in our study is unavailable, which is a more challenging setting, as shown in Table 1.

Anti-forgetting Learning. The main challenge in MD-FAS is the long-studied *catastrophic forgetting* [25]. According to [11], there exist four solutions: replay [8, 46, 54], parameter isolation [13, 40, 49], prior-driven [4, 25, 28] and data-driven [12, 22, 29]. The replay method requires to restore a fraction of training data which breaks our source-free constraint, *e.g.*, [47] needs to store the exemplar training data. Parameter isolation methods [13, 40, 49] dynamically expand the network, which is also discouraged due to the memory expense. The prior-driven methods [4, 25, 28] are proposed based on the assumption that model parameters obey the Gaussian distribution, which is not always the case. The data-driven method [3, 15, 16, 19] is always more favored in the community, due to its effectiveness and low computation cost. However, the development of data-driven methods is dampened in the FAS, since the commonly-used pre-trained model responses (*e.g.*, class probabilities [29] and grad-CAM [12]) fail to capture spoof regions. In this context, our SRE is a simple yet effect way of estimating the spoof trace in the image, which serves as the responses of the pre-train model.

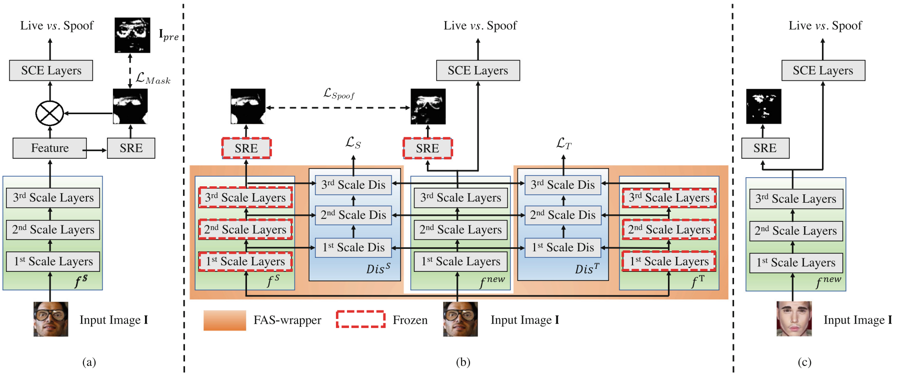


Fig. 2. (a) Given the source pre-trained model that contains feature extractor f^S, we fine-tune it with the proposed *spoof region estimator* (SRE) on the target domain data, in which we use preliminary mask ($\mathbf{I}_{pre}$) to assist the learning (see Sect. 3.2). Then, we obtain a well-trained *SRE* and a new feature extractor f^T which specializes in the target domain. (b) In *FAS-wrapper*, SRE helps f^S and updated model (f^{new}) generate binary masks indicating spoof cues, which serve as model responses given an input image (**I**). $\mathcal{L}_{Spoof}$ prevents the divergence between estimated spoof traces, to combat *catastrophic forgetting*. Meanwhile, using two multi-scale discriminators (Dis^S and Dis^T), *FAS-wrapper* transfers the knowledge from two teacher models (f^S and f^T) to f^{new} via the adversarial training. (c) The update model f^{new} and SRE can be used for the inference.

Multi-domain Learning. Mostly recently, many large-scale FAS datasets with rich annotations have been collected [30,67,68] in the community, among which [30] studies cross-ethnicity and cross-gender FAS. However they work on multi-modal datasets, whereas our input is a single RGB image. In the literature, our work is similar to the multi-domain learning (MDL) [41,44,45], where a re-trained model is required to perform well on both source and target domain data. The common approaches are proposed from [44,45] based on ResNet [18], which, compared to [26,55], has advantages in increasing the abstraction by convoluation operations. In contrast, an ideal MD-FAS algorithm, such as *FAS-wrapper*, should work in a model-agnostic fashion.

3 Proposed Method

This section is organized as follows. Section 3.1 summarizes the general formulation of recent FAS models. Sectons 3.2 and 3.3 introduce the *spoof region estimator* and overall *FAS-wrapper* architecture. Training and inference procedures are reported in Sect. 3.4.

3.1 FAS Models Study

We investigate the recently proposed FAS methods (see Table 2) and observe that these FAS models have two shared characteristics. **Spoof Cue Estimate.**

Beyond treating FAS as a binary classification problem, many SOTA works emphasize on estimating spoof clues from a given image. Such spoof clues are detected in two ways: (a) optimizing the model to predict auxiliary signals such as depth map or rPPG signals [34,63,65]; (b) interpreting the spoofness from different perspectives: the method in [21] aims to disentangle the spoof noise, including color distortions and different types of artifacts, and spoof traces are interpreted in [35,37] as multi-scale and physical-based traces.

Multi-scale Feature Extractor. Majority of previous FAS methods adopt the multi-scale feature. We believe such a multi-scale structure assists in learning information at different frequency levels. This is also demonstrated in [37] that low-frequency traces (*e.g.*, makeup strokes and specular highlights) and high-frequency content (*e.g.*, Moiré patterns) are equally important for the FAS models' success.

Table 2. Summary of recent FAS models.

Method	Year	Number of scale	Spoof cue estimate
Auxiliary [34]	2018	3	Depth and rPPG signal
Despoofing [21]	2018	3	Color distortions, and display artifacts
MADD [53]	2019	3	Depth
CDCN [65]	2020	3	Depth
STDN [37]	2020	3	Color range bias, content and texture pattern, and depth
BCN [63]	2020	3	Patch, reflection and depth
PSMM-Net [30]	2021	4	Depth, RGB and infrared image
PhySTD [35]	2022	4	Additive and inpainting trace, and depth

As a result, we formalize the generic FAS model using two components: feature extractor f and spoof cue estimate (SCE) layers (or decoders) g. When f takes an input face image, denoted as $\mathbf{I}$, the output feature map at t-th layer of the feature extractor f is $f_t(\mathbf{I})$. The size of $f_t(\mathbf{I})$ is $C_t \times H_t \times W_t$, where C_t is the channel number, and H_t and W_t are respectively the height and width of feature maps.

3.2 Spoof Region Estimator

Motivation. Apart from the importance of identifying spoof cues for FAS performance, we observe that spoof trace also serves as a key reflection of how different models make the binary decision, namely, different models' activations on the input image. In other words, although different models might unanimously classify the same image as spoof, they in fact could make decisions based on distinct spatial regions, as depicted in Fig. 6. Thus, we attempt to prevent the divergence between spoof regions estimated from the new model (*i.e.*, f^{new}) and source domain pretrained model (*i.e.*, f^S), such that we can enable f^{new} to perceive spoof cues from the perspective of f^S, thereby combating the *catastrophic forgetting* issue. To this end, we propose a *spoof region estimator* (*SRE*) to localize spatial pixel positions with spoof artifacts or covered by spoof materials.

Formulation. Let us formulate the spoof region estimate task. We denote the pixel collection in an image as $D_{\mathbf{I}} = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, the proposed method aims to predict the region where the area of presentation attack can be represented as a binary mask, denoted as $D_{pred} = \{(x_1, y_1'), (x_2, y_2'), ..., (x_n, y_n')\}$, where x_i, y_i and y_i' respectively represent the pixel, ground truth pixel label, and predicted label at i th pixel. Also, the spoof region estimate task can be regarded

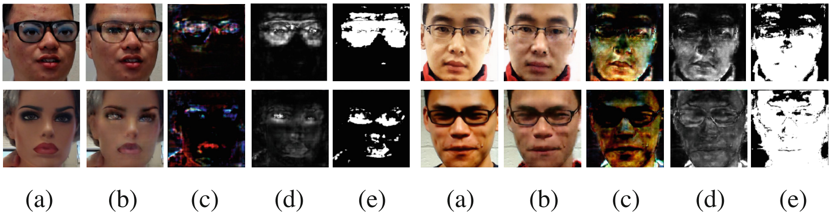


Fig. 3. The preliminary mask generation process: (a) the spoof image, (b) the live reconstruction, (c) and (d) are difference image in RGB and gray format, and (e) is the preliminary spoof mask.

as a pixel-level binary classification problem, namely pixel being live or spoof, thus we have $y_i \in \{o^{Live}, o^{Spoof}\}$. Note that $i \in \{1, 2, 3..., n\}$ and n is the total number of pixels in the image.

Method. As depicted in Fig. 2, we insert a *SRE* module in the source pre-trained model, between the feature extractor f^S and spoof cue estimate layers g^S. The region estimator converts $f^s(\mathbf{I})$ to a binary mask $\mathbf{M}$ with the size $H_{t'} \times W_{t'}$. In the beginning of the training, we create the preliminary mask to supervise *SRE* for generating the spoof region. The preliminary mask generation is based on the reconstruction method proposed in [37], as illustrated in Fig. 3. In particular, we denote input spoof image as $\mathbf{I}_{spoof}$ and use the method in [37] to reconstruct its live counterpart $\hat{I}_{live}$. By subtracting $\mathbf{I}_{spoof}$ from $\hat{\mathbf{I}}_{live}$, and taking the absolute value of the resulting image, we obtain the different image $\mathbf{I}_d$, whose size is $C_0 \times H_0 \times W_0$ where C_0 is 3. We convert $\mathbf{I}_d$ to a gray image $\hat{\mathbf{I}}_d$, by summing along with its channel dimension. Apparently, $\hat{\mathbf{I}}_d$ has the size as $C_1 \times H_0 \times W_0$ where C_1 is 1. We assign each pixel value in the preliminary mask by applying a predefined threshold T,

$$p'_{ij} = \begin{cases} 0 & p_{ij} < T \\ 1 & p_{ij} \geq T, \end{cases} \tag{1}$$

where pixels in $\hat{\mathbf{I}}_d$ and $\mathbf{I}_{pre}$ are p_{ij} and p'_{ij} respectively.

Evidently, the supervisory signal $\mathbf{I}_{pre}$ is not the ground truth. Inspired by [10] that a model can generate the manipulation mask by itself during training procedure, we only use $\mathbf{I}_{pre}$ as the supervision at the first a few training epochs, then steer the model itself to find the optimal spoof region by optimizing towards a higher classification accuracy. More details are in Sect. 3.4.

Discussion. Firstly, we discuss the difference to prior spoof region estimate works. The previous methods [35,37] use various traces to help live or spoof image reconstruction, while our goal is to pinpoint the region with spoof artifacts, which serves as pre-trained model's responses to help the new model behave similar to the pre-trained one(s), alleviating the forgetting issue. [10] offers low-resolution binary masks as the supervisory signal, but our self-generated $\mathbf{I}_{pre}$ can only bootstrap the system. Also, [69] proposes an architecture for producing multiple masks, which is not practical in our scenario. Thus our mask generation method is different from theirs. Finally, SRE can be a plug-in module for any given FAS model, and details are in Sect. 5.4.

3.3 *FAS-Wrapper* Architecture

Motivation. We aim to deliver an update algorithm that can be effortlessly deployed to different FAS models. Thus, it is important to design a model agnostic algorithm that allows the FAS model to remain intact, thereby maintaining the original FAS model performance. Our *FAS-wrapper* operates in a model-agnostic way where only external expansions are made, largely maintaining the original FAS model's ability.

As depicted in Fig. 2, we denote the source pre-trained feature extractor f^S as *source teacher*, and the feature extractor after the fine-tuning procedure as *target teacher* (f^T). Instead of using one single teacher model like [20], we use f^S and f^T to regularize the training, offering the more informative and instructive supervision for the newly upgraded model, denoted as f^{new}. Lastly, unlike prior FAS works [20,53,61] which apply the indiscriminative loss on the final output embedding or logits from f^S, we construct multi-scale discriminators that operate at the feature-map level for aligning intermediate feature distributions of f^{new} to those of teacher models (*i.e.*, f^T and f^S). Motivations of the multi-scale discriminators are: (a) the multi-scale features, as a common FAS model attribute (Sect. 3.1), should be considered; (b) the adversarial learning can be used at the feature-map level which contains the richer information than final output logits.

Method. We construct two multi-scale discriminators, Dis^S and Dis^T, for transferring semantic knowledge from f^S and f^T to f^{new} respectively, via an adversarial learning loss. Specifically, at l-th scale, Dis_l^S and Dis_l^T take the previous discriminator output and the l-th scale feature generated from feature extractors. We use $\mathbf{d}_l^S$ and $\mathbf{d}_l^T$ to represent two discriminators' outputs at l-th level while taking teacher generated features (*i.e.*, $f_l^S(\mathbf{I})$ and $f_l^T(\mathbf{I})$), and $\mathbf{d}_l'^S$ and $\mathbf{d}_l'^T$ while taking upgraded model generated feature, $f_l^{new}(\mathbf{I})$. Therefore, the first-level discriminator output are:

$$\mathbf{d}_1^S = Dis_1^S(f_1^S(\mathbf{I})), \quad \mathbf{d}_1^T = Dis_1^T(f_1^T(\mathbf{I})), \tag{2}$$

$$\mathbf{d}_1'^S = Dis_1^S(f_1^{new}(\mathbf{I})), \quad \mathbf{d}_1'^T = Dis_1^T(f_1^{new}(\mathbf{I})), \tag{3}$$

and discriminators at following levels take the l-th ($l > 1$) backbone layer output feature and the previous level discriminator output, so we have:

$$\mathbf{d}_l^S = Dis_l^S(f_l^S(\mathbf{I})) \oplus \mathbf{d}_{l-1}^S), \quad \mathbf{d}_l^T = Dis_l^T(f_l^T(\mathbf{I})) \oplus \mathbf{d}_{l-1}^T), \tag{4}$$

$$\mathbf{d}_l'^S = Dis_l^S(f_l^{new}(\mathbf{I})) \oplus \mathbf{d}_{l-1}'^S), \quad \mathbf{d}_l'^T = Dis_l^T(f_l^{new}(\mathbf{I})) \oplus \mathbf{d}_{l-1}'^T). \tag{5}$$

After obtaining the output from the last-level discriminator, we define $\mathcal{L}_{D_S}$ and $\mathcal{L}_{D_T}$ to train Dis_s and Dis_t, and $\mathcal{L}_S$ and $\mathcal{L}_T$ to supervise f^{new}.

$$\mathcal{L}_S = -\mathbb{E}_{x_p \sim P_s}[log(\mathbf{d}_l^S)] - \mathbb{E}_{x_f \sim P_{new}}[log(1 - \mathbf{d}_l'^S)], \tag{6}$$

$$\mathcal{L}_T = -\mathbb{E}_{x_p \sim P_t}[log(\mathbf{d}_l^T)] - \mathbb{E}_{x_f \sim P_{new}}[log(1 - \mathbf{d}_l'^T)], \tag{7}$$

$$\mathcal{L}_{D_s} = -\mathbb{E}_{x_p \sim P_s}[log(1 - \mathbf{d}_l^S)] - \mathbb{E}_{x_f \sim P_{new}}[log(\mathbf{d}_l'^S)], \tag{8}$$

$$\mathcal{L}_{D_t} = -\mathbb{E}_{x_p \sim P_t}[log(1 - \mathbf{d}_l^T)] - \mathbb{E}_{x_f \sim P_{new}}[log(\mathbf{d}_l'^T)]. \tag{9}$$

Discussion. The idea of adopting adversarial training on the feature map for knowledge transfer is similar to [9]. However, the method in [9] is for the online task and transferring knowledge from two models specialized in the same domain. Conversely, our case is to learn from heterogeneous models which specialize in different domains. Additionally, using two regularization terms with symmetry based on the two pre-trained models, is similar to work in [66] on the knowledge distillation topic that is different to FAS. However, the same is the effect of alleviating the imbalance between classification loss and regularization terms, as reported in [24,66].

3.4 Training and Inference

Our training procedure contains two stages, as depicted in Fig. 2. Firstly, we fine-tune given any source pre-trained FAS model with the proposed SRE, on the target dataset. We optimize the model by minimizing the ℓ_1 distance (denoted as $\mathcal{L}_{Mask}$) between the predicted binary mask $\mathbf{M}$ and $\mathbf{I}_{pre}$, and the original loss $\mathcal{L}_{Orig}$ that is used in the training procedure of original FAS models. After the fine-tuning process, we obtain well-trained *SRE* and a feature extractor (f^T) that is able to work reasonably well on target domain data. Secondly, we integrate the well-trained *SRE* with the updated model (f^{new}) and the source pre-trained model (f^S), such that we can obtain estimated spoof cues from perspectives of two models. We use $\mathcal{L}_{Spoof}$ to prevent the divergence between spoof regions estimated from f^{new} and f^S. Lastly, we use $\mathcal{L}_S$ and $\mathcal{L}_T$ as introduced in Sect. 3.3 for transferring knowledge from the f^S and f^T to f^{new}, respectively. Therefore, the overall objective function in the training is denoted as $\mathcal{L}_{total}$:

$$\mathcal{L}_{total} = \lambda_1\mathcal{L}_{Orig} + \lambda_2\mathcal{L}_{Spoof} + \lambda_3\mathcal{L}_S + \lambda_4\mathcal{L}_T, \tag{10}$$

where λ_1-λ_4 are the weights to balance the multiple terms. In inference, we only keep new feature extract f^{new} and *SRE*, as depcited in Fig. 2 (c).

4 FASMD Dataset

We construct a new benchmark for MD-FAS, termed FASMD, based on SiW [34], SiW-Mv2 [36][1] and Oulu-NPU [7]. MD-FAS consists of five sub-datasets: dataset A is the source domain dataset, and B, C, D and E are four target domain datasets, which introduce unseen spoof type, new ethnicity distribution, age distribution and novel illumination, respectively. The statistics of the FASMD benchmark are reported in Table 3.

[1] We release SiW-Mv2 on CVLab website.

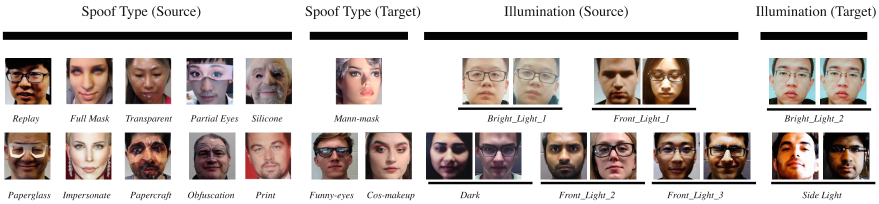


Fig. 4. Representative examples in source and target domain for spoof and illumination protocols.

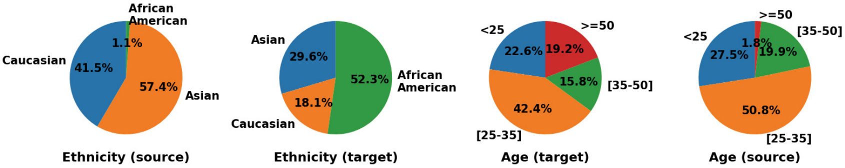


Fig. 5. The distribution of ethnicity and age in source and target domain subsets.

Table 3. The FASMD benchmark. [Keys: eth.= ethnicity, illu.= illumination.]

Video Num/Subject Num		
Dataset ID	Train	Test
A (Source)	4, 983/603	2, 149/180
B (New spoof type)	1, 392/301	383/71
C (New eth. distribution)	1, 024/360	360/27
D (New age distribution)	892/157	411/43
E (New illu. distribution)	1, 696/260	476/40

New Spoof Type. As illustrated in Fig. 4, target domain dataset B has novel spoof types that are excluded from the source domain dataset (A). The motivation for this design is, compared with the *print* and *replay* that are prevalent nowadays, other new spoof types are more likely to emerge and cause threats. As a result, given the fact that, five macro spoof types are introduced in SIW-Mv2 (*print*, *replay*, *3D mask*, *makeup* and *partial manipulation attack*), we select one micro spoof type from other three macro spoof types besides *print* and *replay* to constitute the dataset B, which are *Mannequin mask*, *Cosmetic makeup* and *Funny eyes*.

New Ethnicity Distribution. In reality, pre-trained FAS models can be deployed to organizations with certain ethnicity distribution (*e.g.*, African American sports club). Therefore, we manually annotate the ethnicity information of each subject in three datasets, then devise the ethnicity protocol where dataset A has only 1.1% African American samples, but this proportion increases to 52.3% in dataset C, as depicted in Fig. 5.

New Age Distribution, Likewise, a FAS model that is trained on source domain data full of college students needs to be deployed to the group with a different age distribution, such as a senior care or kindergartens. We estimate the age information by the off-the-shelf tool [48], and construct dataset D to have a large portion of subjects over 50 years old, as seen in Fig. 5.

New Illumination. Oulu-NPU dataset has three different illumination sessions, and we use methods proposed in [71] to estimate the lighting condition for each sample in SIW and SIW-Mv2 datasets. Then we apply K-means [39] to cluster them into K groups. For the best clustering performance, we use "eblow method" [57] to decide the value of K. We annotate different illumination sessions as *Dark*, three *Front Light*, *Side Light*, and two *Bright Light* (Fig. 4), then dataset E introduces the new illumination distribution.

5 Experimental Evaluations

5.1 Experiment Setup

We evaluate our proposed method on the FASMD dataset. In Sect. 5.3, we report *FAS-wrapper* performance with different FAS models, and we choose PhySTD [35] as the FAS model for analysis in Sect. 5.2, because PhySTD has demonstrated competitive empirical FAS results. Firstly, we compare to anti-forgetting methods (*e.g.*, LwF [29], MAS [4] and LwM [12]). Specifically, based on the architecture of PhySTD, we concatenate feature maps generated by last convolution layers in different branches, then employ Global Average Pooling and fully connected (FC) layers to convert concatenated features into a 2-dimensional vector. We fix the source pre-trained model weights and only train added FC layers in the original FAS task, as a binary classifier. In this way, we can apply methods in [4,12,29] to this binary classifier. For multi-domain learning methods (*e.g.*, Serial and Parallel Res-Adapter [44,45]), we choose the 1×1 kernel size convolution filter as the adapter and incorporate into the PhySTD as described in original works (see details in the supplementary material). We use standard FAS metrics to measure the performance, which are Attack Presentation Classification Error Rate (APCER), Bona Fide Presentation Classification Error Rate (BPCER), and Average Classification Error Rate ACER [1], Receiver Operating Characteristic (ROC) curve.

Implementation Details. We use Tensorflow [2] in implementation, and we run experiments on a single NVIDIA TITAN X GPU. In the source pre-train stage, we use a learning rate 3e–4 with a decay rate 0.99 for every epoch and the total epoch number is 180. We set the mini-batch size as 8, where each mini-batch contains 4 live images and 4 spoof images (*e.g.*, 2 SIW-Mv2 images, 1 image in SIW and OULU-NPU, respectively). Secondly, we keep the same hyper-parameter setting as the pre-train stage, fine-tune the source domain pre-trained model with *SRE* at a learning rate 1e–6. The overall FAS-wrapper is trained with a learning rate 1e–7.

5.2 Main Results

Table 4 reports the detailed performance from different models on all four protocols. Overall, our method surpasses the previous best method on source and target domain evaluation in *all categories*, with the only exception of the target domain performance in the illumination protocol (0.3% worse than [12]).

Table 4. The main performance reported in TPR@FPR=0.5%. Scores before and after "/" are performance on the source and target domains respectively. [Key: **Best**, **Second Best**, except for two teacher models and upper bound performance in the first three rows (🟦)].

Method	Training data	Spoof	Age	Ethnicity	Illumination	Average
Upper Bound	Source + Target	89.5/52.5	86.7/82.3	87.7/62.8	89.0/74.4	88.2/68.0
Source Teacher	Source	84.2/39.8	84.2/72.8	84.2/59.2	84.2/64.4	84.2/59.1
Target Teacher	Target	73.5/51.5	67.9/80.8	77.2/61.9	65.0/71.8	70.9/66.3
LwF [29]	Target	74.8/50.8	71.7/77.9	71.0/59.8	65.3/69.2	70.7/64.4
LwM [12]	Target	76.5/51.0	71.5/80.0	76.0/62.0	71.0/**71.8**	73.8/**65.9**
MAS [4]	Target	73.4/48.8	68.3/78.6	73.5/60.9	66.0/65.9	71.4/63.5
Seri. RA [44]	Target	74.3/**51.4**	72.6/79.8	72.0/61.7	67.0/70.4	71.5/65.8
Para. RA [45]	Target	75.5/51.2	73.0/79.7	72.0/61.5	68.0/69.3	72.1/65.4
Ours - ($\mathcal{L}_T + \mathcal{L}_S$)	Target	80.3/50.5	77.1/79.0	75.1/61.3	77.2/69.4	77.4/65.1
Ours - $\mathcal{L}_T$	Target	**80.5**/50.8	**79.0**/79.4	**76.1**/61.5	**78.3**/70.2	**78.5**/77.4
Ours - $\mathcal{L}_{Spoof}$	Target	75.5/51.0	70.4/79.3	74.9/**62.1**	70.1/70.0	72.7/65.6
Ours	Target	**81.8**/**51.5**	**79.5**/**80.6**	**76.8**/**62.3**	**79.6**/**71.5**	**79.4**/**66.4**

More importantly, regarding performance on source domain data, it is impressive that our method surpasses the best previous method in all protocols by a large margin (*e.g.*, 5.3%, 6.5%, 0.8% and 8.6%, and 5.6% on average). We believe that, the proposed SRE can largely alleviate the *catastrophic forgetting* as mentioned above, thereby yielding the superior source domain performance than prior works. However, the improvement diminishes on the new ethnicity protocol. One possible reason is that the *print* and *replay* attacks account for a large portion of data in new ethnicity distribution, and different methods, performance on these two common presentation attacks are similar.

Additionally, Table 5 reports the average performance on four protocols in terms of ACPER, BCPER and ACER. Our method still remains the best, besides BPCER on the target domain performance. It is worth mentioning that we have 4.2% APCER on source domain data and 8.9% ACER on target domain data, which are better than best results from prior works, namely 5.6% APCER in [12] and 11.0% ACER in [45]. Furthermore, in Sect. 5.4, we examine the adaptability of our proposed method, by incorporating it with different FAS methods.

Table 5. The average performance of the different methods in four protocols. The scores before and after "/" are performance on source and target domains. [Key: **Best**, **Second Best**, except for two teacher models and upper bound performance in first three rows (🟦)].

Method	APCER (%)	BPCER (%)	ACER (%)
Upper Bound	3.7/4.5	5.8/13.3	5.2/13.3
Source Teacher	4.1/4.8	7.4/23.0	6.1/23.0
Target Teacher	6.6/4.6	6.4/14.2	5.5/10.0
LwF [29]	6.4/5.6	8.1/15.9	6.8/11.2
LwM [12]	5.6/**4.3**	8.0/14.0	**6.3**/11.1
MAS [4]	6.7/5.3	8.4/13.9	6.8/11.2
Seri. RA [44]	6.3/5.0	8.4/15.0	6.7/11.3
Para. RA [45]	6.2/7.6	8.4/**13.4**	6.7/11.0
Ours - ($\mathcal{L}_S + \mathcal{L}_T$)	4.8/6.5	9.1/15.6	6.9/11.1
Ours - $\mathcal{L}_T$	**4.5**/5.2	8.3/14.9	6.4/**10.1**
Ours - $\mathcal{L}_{Spoof}$	6.0/8.8	**8.0**/14.2	7.0/11.5
Ours	**4.2**/**4.2**	**7.8**/**13.5**	**6.0**/**8.9**

Table 6. (a) The *FAS-wrapper* performance with different FAS models; (b) Performance of adopting different architecture design choices.

TPR@FPR=0.5% (Source/Target)	PhySTD [35]	CDCN [65]	Auxi.-CNN [34]
Naive Fine.	70.9/66.3	69.2/63.2	63.3/61.3
Full (Ours)	**79.4/66.4**	**74.8/62.7**	**70.3**/61.3
Full - $\mathcal{L}_{Spoof}$	72.7/65.6	74.1/62.5	69.0/**61.4**
Full - $\mathcal{L}_{D_t}$	78.5/64.4	73.1/62.3	69.1/61.3
Full - $(\mathcal{L}_T + \mathcal{L}_S)$	77.3/65.1	71.6/62.1	65.6/61.3

	TPR@FPR=0.5%
$\mathcal{L}_{Spoof}$ + Multi-disc. (Ours)	**79.4/66.4**
$\mathcal{L}_{Spoof}$ + Multi-disc. (same weights)	75.0/65.8
$\mathcal{L}_{Spoof}$ + Single disc. (concat.)	74.4/66.0
$\mathcal{L}_{Spoof}$ + [60]	65.2/63.2

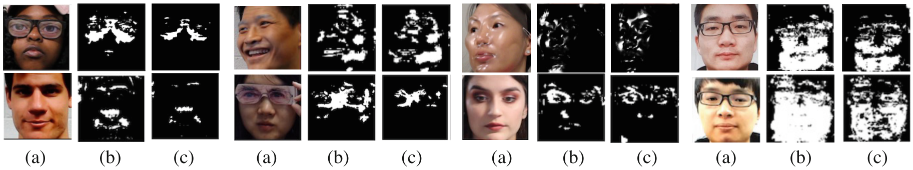


Fig. 6. Spoof region estimated from different models. Given input image (a), (b) and (c) are model responses from [35] and [34], respectively. Detailed analyses in Sect. 5.4.

Ablation Study Using $\mathcal{L}_{Spoof}$. *SRE* plays a key role in *FAS-wrapper* for learning the new spoof type, as ablating the $\mathcal{L}_{Spoof}$ largely decreases the source domain performance, namely from 79.4% to 72.7% on TPR@FPR = 0.5% (Table 4) and 1.8% on APCER (Table 5). Such a performance degradation supports our statement that, $\mathcal{L}_{Spoof}$ prevents divergence between spoof traces estimated from the source teacher and the upgraded model, which helps to combat the *catastrophic forgetting* issue, and maintain the source domain performance.

Ablation Study Using $\mathcal{L}_S$ and $\mathcal{L}_T$. Without the adversarial learning loss ($\mathcal{L}_S$ + $\mathcal{L}_T$), the model performance constantly decreases, according to Table 4, although such impacts are less than removal of $\mathcal{L}_{Spoof}$, which still causes 2.0% and 1.3% average performance drop on source and target domains. Finally, we have a regularization term $\mathcal{L}_T$ which also contributes to performance. That is, removing $\mathcal{L}_T$ hinders the FAS performance (*e.g.*, 1.0% ACER on target domain performance), as reported in Table 5.

5.3 Adaptability Analysis

We apply *FAS-wrapper* on three different FAS methods: Auxi.-CNN [34], CDCN [65] and PhySTD [35]. CDCN uses a special convolution (*i.e.*, Central Difference Convolution) and Auxi.-CNN is the flagship work that learns FAS via auxiliary supervisions. As shown in Table 6 (a), *FAS-wrapper* can consistently improve the performance of naive fine-tuning. When ablating the $\mathcal{L}_{Spoof}$, PhySTD [35] experiences the large performance drop (6.7%) on the source domain, indicating the importance of SRE in the learning the new domain. Likewise, the removal of adversarial learning loss (*e.g.*, $\mathcal{L}_T + \mathcal{L}_S$) leads to difficulty in preserving the source domain performance, which can be shown from, on the source domain, CDCN [65] decreases 3.2% and Auxi-CNN [34] decreases 4.7%.

This means dual teacher models, in the *FAS-wrapper*, trained with adversarial learning benefit the overall FAS performance. Also, we visualize the spoof region generated from *SRE* with [34,35] in Fig. 6. We can see the spoof cues are different, which supports our hypothesis that, although FAS models make the same final binary prediction, they internally identify spoofness in different areas.

5.4 Algorithm Analysis

Spoof Region Visualization. We feed output features from different models (*i.e.*, f^S, f^T and f^{new}) to a well-trained *SRE* to generate the spoof region, as depicted in Fig. 7. In general, the f^S produces more accurate activated spoof regions on the source domain images. For example, two source images in new spoof category have detected makeup spoofness on eyebrows and mouth (first row) and more intensive activation on the funny eye region (second row). f^T has the better spoof cues estimated on the target domain image. For example, two target images in the new spoof category, where spoofness estimated from f^T is stronger and more comprehensive; in the novel ethnicity category, the spoofness covers the larger region. With $\mathcal{L}_{Spoof}$, the updated model (f^{new}) identifies the spoof traces in a more accurate way.

Explanability. We compare *SRE* with the work which generate binary masks indicating the spoofness [10], and works which explain how a model makes a binary classification decision [52,70]. In Fig. 8, we can observe that our generated spoof traces can better capture the manipulation area, regardless of spoof types. For example, in the first *print* attack image, the entire face is captured as spoof in our method but other three methods fail to achieve so. Also, our binary mask is more detailed and of higher resolution than that of [10], and more accurate and robust than [52,70]. Notably, we do not include works in [35,37,69] which use many outputs to identify spoof cues.

Architecture Design. We compare to some other architecture design choices, such as all multi-scale discriminators with the same weights, concatenation of

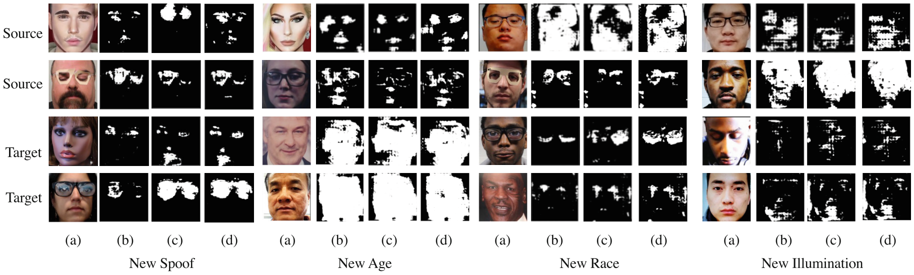


Fig. 7. Given the input spoof image (a), spoof regions generated by *SRE* with two teacher models (*i.e.*, f^S and f^T) in (b) and (c), and the new upgraded model (f^{new}) in (d), for different protocols. Detailed analyses are in Sect. 5.4.

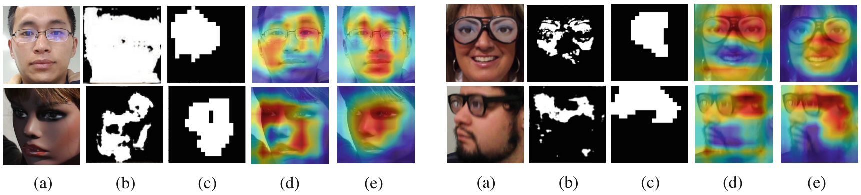


Fig. 8. Different spoof estimate methods. Given input image (a), (b) and (c) are the spoof regions estimated from ours and [10]. (d) and (e) are the activated map from methods in [52,70].

different scale features and one single discriminator. Moreover, we use correlation similarity table in [60] instead of multi-scale discriminators for transfering knowledge from f^S and f^T to f^{new}. Table 6(b) demonstrates the superiority of our architectural design.

5.5 Cross-Dataset Study

We evaluate our methods in the cross-dataset scenario and compare to SSDG [20] and MADDG [53]. Specifically, we denote OULU-NPU [7] as O, SIW [34] as S, SIW-Mv2 [36] as M, and HKBU-MARs [32] as H. We use three datasets as source domains for training and one remaining dataset for testing. We train three individual source domain teacher models on three source datasets respectively. Then, as depicted in Fig. 9, inside *FAS-wrapper*, three multi-scale discriminators are employed to transfer knowledge from three teacher models to the updated model f^{new} which is then evaluated on the target domain. Notably, we remove proposed *SRE* in this cross-dataset scenario, as there is no need to restore the prior model responses.

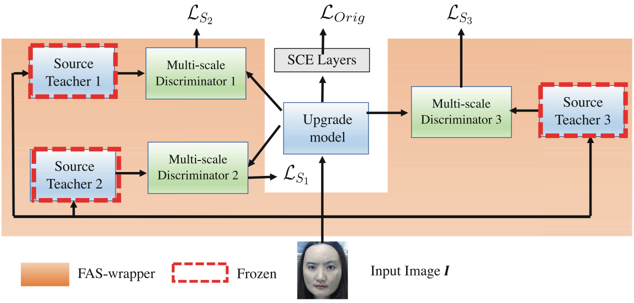


Fig. 9. We adapt *FAS-wrapper* for the cross-dataset scenario.

Table 7. The cross-dataset comparison.

	O&M&H to S		O&W&H to M		M&S&H to O		M&S&O to H	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
MADDG [53]	16.7	90.5	50.3	60.7	17.6	73.0	33.2	73.5
SSDG-M [20]	**11.1**	93.4	29.6	67.1	**12.1**	**89.0**	**25.0**	82.5
SSDG-R [20]	13.3	93.4	29.3	**69.5**	13.3	83.4	28.9	81.0
Ours	15.4	**93.6**	**28.1**	68.4	14.8	85.6	27.1	**83.8**

The results are reported in Table 7, indicating that our *FAS-wrapper* also exhibits a comparable performance on the cross-dataset scenario as prior works.

6 Conclusion

We study the multi-domain learning face anti-spoofing (MD-FAS), which requires the model perform well on both source and novel target domains, after updating the source domain pre-trained FAS model only with target domain data. We first summarize the general form of FAS models, then based on which we develop a new architecture, *FAS-wrapper*. *FAS-wrapper* contains spoof region estimator which identifies the spoof traces that help combat *catastrophic forgetting* while learning new domain knowledge, and the *FAS-wrapper* exhibits a high level of flexibility, as it can be adopted by different FAS models. The performance is evaluated on our newly-constructed FASMD benchmark, which is also the first MD-FAS dataset in the community.

Acknowledgement. This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2017-17020200004. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

1. international organization for standardization. Iso/iec jtc 1/sc 37 biometrics: Information technology biometric presentation attack detection part 1: Framework. https://www.iso.org/obp/ui/iso. Accessed 3 Mar 2022
2. Abadi, M., et al.: TensorFlow: a system for large-scale machine learning. In: OSDI (2016)
3. AbdAlmageed, W., et al.: Assessment of facial morphologic features in patients with congenital adrenal hyperplasia using deep learning. JAMA Netw. Open **3** (2020)

4. Aljundi, R., Babiloni, F., Elhoseiny, M., Rohrbach, M., Tuytelaars, T.: Memory aware synapses: learning what (not) to forget. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11207, pp. 144–161. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01219-9_9
5. Asnani, V., Yin, X., Hassner, T., Liu, S., Liu, X.: Proactive image manipulation detection. In: CVPR (2022)
6. Atoum, Y., Liu, Y., Jourabloo, A., Liu, X.: Face anti-spoofing using patch and depth-based CNNS. In: IJCB (2017)
7. Boulkenafet, Z., Komulainen, J., Li, L., Feng, X., Hadid, A.: OULU-NPU: a mobile face presentation attack database with real-world variations. In: IEEE International Conference on Automatic Face and Gesture Recognition (2017)
8. Chaudhry, A., Ranzato, M., Rohrbach, M., Elhoseiny, M.: Efficient lifelong learning with a-gem. ICLR (2019)
9. Chung, I., Park, S., Kim, J., Kwak, N.: Feature-map-level online adversarial knowledge distillation. In: ICML (2020)
10. Dang*, H., Liu*, F., Stehouwer*, J., Liu, X., Jain, A.: On the detection of digital face manipulation. In: CVPR (2020)
11. Delange, M., et al.: A continual learning survey: Defying forgetting in classification tasks. In: TPAMI (2021)
12. Dhar, P., Singh, R.V., Peng, K.C., Wu, Z., Chellappa, R.: Learning without memorizing. In: CVPR (2019)
13. Fernando, C., et al.: PathNet: evolution channels gradient descent in super neural networks. arXiv preprint arXiv:1701.08734 (2017)
14. Ganin, Y., et al.: Domain-adversarial training of neural networks. J. Mach. Learn. Res. **17**, 2096–2030 (2016)
15. Guo, X., Choi, J.: Human motion prediction via learning local structure representations and temporal dependencies. In: AAAI (2019)
16. Guo, X., Mirzaalian, H., Sabir, E., Jaiswal, A., Abd-Almageed, W.: Cord19sts: Covid-19 semantic textual similarity dataset. arXiv preprint arXiv:2007.02461 (2020)
17. Guo, Y., Li, Y., Wang, L., Rosing, T.: Depthwise convolution is all you need for learning multiple visual domains. In: AAAI (2019)
18. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
19. Hsu, I., et al.: Discourse-level relation extraction via graph pooling. arXiv preprint arXiv:2101.00124 (2021)
20. Jia, Y., Zhang, J., Shan, S., Chen, X.: Single-side domain generalization for face anti-spoofing. In: CVPR (2020)
21. Jourabloo, A., Liu, Y., Liu, X.: Face de-spoofing: Anti-spoofing via noise modeling. In: ECCV (2018)
22. Jung, H., Ju, J., Jung, M., Kim, J.: Less-forgetting learning in deep neural networks. arXiv preprint arXiv:1607.00122 (2016)
23. Khosla, A., Zhou, T., Malisiewicz, T., Efros, A.A., Torralba, A.: Undoing the damage of dataset bias. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7572, pp. 158–171. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33718-5_12
24. Kim, J.Y., Choi, D.W.: Split-and-bridge: Adaptable class incremental learning within a single neural network. In: AAAI (2021)
25. Kirkpatrick, J., et al.: Overcoming catastrophic forgetting in neural networks. In: Proceedings of the National Academy of Sciences (2017)

26. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: NeuriPS (2012)
27. Kundu, J.N., Venkatesh, R.M., Venkat, N., Revanur, A., Babu, R.V.: Class-incremental domain adaptation. In: ECCV (2020)
28. Lee, S.W., Kim, J.H., Jun, J., Ha, J.W., Zhang, B.T.: Overcoming catastrophic forgetting by incremental moment matching. In: NeurIps (2017)
29. Li, Z., Hoiem, D.: Learning without forgetting. In: TPAMI (2017)
30. Liu, A., Tan, Z., Wan, J., Escalera, S., Guo, G., Li, S.Z.: URF CeFA: a benchmark for multi-modal cross-ethnicity face anti-spoofing. In: WACV (2021)
31. Liu, H., Cao, Z., Long, M., Wang, J., Yang, Q.: Separate to adapt: open set domain adaptation via progressive separation. In: CVPR (2019)
32. Liu, S., Yang, B., Yuen, P.C., Zhao, G.: A 3d mask face anti-spoofing database with real world variations. In: CVPR Workshop (2016)
33. Liu, X., Liu, Y., Chen, J., Liu, X.: PSSC-Net: progressive spatio-channel correlation network for image manipulation detection and localization. In: T-CSVT (2022)
34. Liu, Y., Jourabloo, A., Liu, X.: Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In: CVPR (2018)
35. Liu, Y., Liu, X.: Physics-guided spoof trace disentanglement for generic face anti-spoofing. In: TPAMI (2022)
36. Liu, Y., Stehouwer, J., Jourabloo, A., Liu, X.: Deep tree learning for zero-shot face anti-spoofing. In: CVPR (2019)
37. Liu, Y., Stehouwer, J., Liu, X.: On disentangling spoof trace for generic face anti-spoofing. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12363, pp. 406–422. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58523-5_24
38. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: ICML (2015)
39. MacQueen, J., et al.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (1967)
40. Mallya, A., Lazebnik, S.: Packnet: Adding multiple tasks to a single network by iterative pruning. In: CVPR (2018)
41. Mancini, M., Ricci, E., Caputo, B., Bulò, S.R.: Adding new tasks to a single network with weight transformations using binary masks. In: Leal-Taixé, L., Roth, S. (eds.) ECCV 2018. LNCS, vol. 11130, pp. 180–189. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11012-3_14
42. Qin, Y., et al.: Learning meta model for zero-and few-shot face anti-spoofing. In: AAAI (2020)
43. Quan, R., Wu, Y., Yu, X., Yang, Y.: Progressive transfer learning for face anti-spoofing. In: TIP (2021)
44. Rebuffi, S.A., Bilen, H., Vedaldi, A.: Learning multiple visual domains with residual adapters. In: NeurIPS (2017)
45. Rebuffi, S.A., Bilen, H., Vedaldi, A.: Efficient parametrization of multi-domain deep neural networks. In: CVPR (2018)
46. Rebuffi, S.A., Kolesnikov, A., Sperl, G., Lampert, C.H.: ICARL: incremental classifier and representation learning. In: CVPR (2017)
47. Rostami, M., Spinoulas, L., Hussein, M., Mathai, J., Abd-Almageed, W.: Detection and continual learning of novel face presentation attacks. In: ICCV (2021)
48. Rothe, R., Timofte, R., Van Gool, L.: DEX: deep expectation of apparent age from a single image. In: ICCV Workshops (2015)D

49. Rusu, A.A., et al.: Progressive neural networks. arXiv preprint arXiv:1606.04671 (2016)
50. Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting Visual category models to new domains. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 213–226. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15561-1_16
51. Saito, K., Yamamoto, S., Ushiku, Y., Harada, T.: Open set domain adaptation by backpropagation. In: ECCV (2018)
52. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision **128**(2), 336–359 (2019). https://doi.org/10.1007/s11263-019-01228-7
53. Shao, R., Lan, X., Li, J., Yuen, P.C.: Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In: CVPR (2019)
54. Shin, H., Lee, J.K., Kim, J., Kim, J.: Continual learning with deep generative replay. NeurIPS (2017)
55. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
56. Stehouwer, J., Jourabloo, A., Liu, Y., Liu, X.: Noise modeling, synthesis and classification for generic object anti-spoofing. In: CVPR (2020)
57. Thorndike, R.L.: Who belongs in the family? Psychometrika **18**, 267–276 (1953)
58. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: CVPR (2011)
59. Tu, X., Ma, Z., Zhao, J., Du, G., Xie, M., Feng, J.: Learning generalizable and identity-discriminative representations for face anti-spoofing. In: TIST (2020)
60. Tung, F., Mori, G.: Similarity-preserving knowledge distillation. In: ICCV (2019)
61. Yang, B., Zhang, J., Yin, Z., Shao, J.: Few-shot domain expansion for face anti-spoofing. arXiv preprint arXiv:2106.14162 (2021)
62. Yoo, J., Uh, Y., Chun, S., Kang, B., Ha, J.W.: Photorealistic style transfer via wavelet transforms. In: ICCV (2019)
63. Yu, Z., Li, X., Niu, X., Shi, J., Zhao, G.: Face anti-spoofing with human material perception. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12352, pp. 557–575. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58571-6_33
64. Yu, Z., Li, X., Shi, J., Xia, Z., Zhao, G.: Revisiting pixel-wise supervision for face anti-spoofing. Behavior, and Identity Science, IEEE Trans. Biomet. 3, 285–295 (2021)
65. Yu, Z., et al.: Searching central difference convolutional networks for face anti-spoofing. In: CVPR (2020)
66. Zhang, J., et al.: Class-incremental learning via deep model consolidation. In: WACV (2020)
67. Zhang, S., et al.: CASIA-SURF: a large-scale multi-modal benchmark for face anti-spoofing. Behavior, and Identity Science, IEEE Trans. Biomet. **2**, 182–193 (2020)
68. Zhang, Y., et al.: CelebA-Spoof: large-scale face anti-spoofing dataset with rich annotations. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12357, pp. 70–85. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58610-2_5

69. Zhao, H., Zhou, W., Chen, D., Wei, T., Zhang, W., Yu, N.: Multi-attentional deepfake detection. In: CVFPR (2021)
70. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: CVPR (2016)
71. Zhou, H., Hadap, S., Sunkavalli, K., Jacobs, D.W.: Deep single-image portrait relighting. In: ICCV (2019)