Test-Time Adaptation for Robust Face Anti-Spoofing

Pei-Kai Huang
alwayswithme@gapp.nthu.edu.tw
Chen-Yu Lu
vanyar0409@gapp.nthu.edu.tw
Shu-Jung Chang
vuc802@gapp.nthu.edu.tw
Jun-Xiong Chong
jxchong@gapp.nthu.edu.tw
Chiou-Ting Hsu
cthsu@cs.nthu.edu.tw

Department of Computer Science,
National Tsing Hua University,
Taiwan

Abstract

Face anti-spoofing (FAS) aims to defend face recognition systems from various presentation attacks. To deal with cross-domain testing scenarios, many FAS methods adopted domain generalization or domain adaptation approaches by using all the available source domain data to adapt the model in the offline training stage. However, as there exist ever-growing and ever-evolving attacks, attempting to simulate unseen attacks by offline adaptation techniques is extremely difficult if not impossible. Test-Time Adaptation (TTA), which focuses on on-line adapting an off-the-shelf model to unlabeled target data without referring to any source data, has been successfully adopted in image classification but is still unexplored in FAS methods. In this paper, our goal is to address the TTA issues for robust face anti-spoofing. We first propose a novel TTA benchmark covering different domains and various attacks to simulate the challenges of FAS when facing new domain data and unseen attacks. Next, we develop a novel framework 3A-TTA, including three main components: activation-based pseudo-labeling, anti-forgetting feature learning, and asymmetric prototype contrastive learning to tackle the issues of TTA in FAS. Our extensive experiments on the proposed benchmark show that the proposed 3A-TTA achieves superior performance for on-line detecting both seen and unseen types of face presentation attacks from new domains.

1 Introduction

Face recognition techniques have been widely adopted in our daily life to facilitate efficient authentication or mobile payments. Despite the convenience of face recognition systems, they are susceptible to many facial presentation attacks, such as Print Attack (i.e., printing a face on paper), Replay Attack (i.e., replaying a face video on digital devices), and 3D Mask Attack (i.e., wearing a face mask). Therefore, many face anti-spoofing (FAS) methods [1, 4,

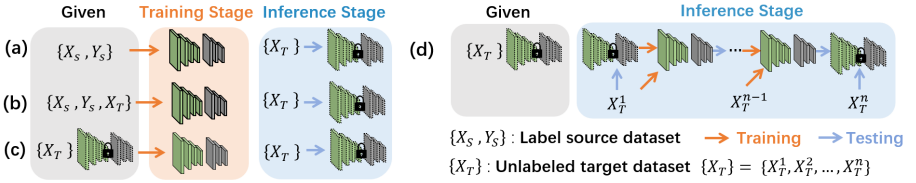

Figure 1: Illustration of different cross-domain settings: (a) domain generalization (DG), (b) domain adaptation (DA), (c) source-free domain adaptation (SFDA), and (d) test-time adaptation. Note that the inference stage in test-time adaptation consists of two sub-stages: an adaptation process (Training) and an inference process (Testing).

[5, 7, 9, 11, 12, 13, 16, 17, 18, 19, 22, 23, 24, 25, 26, 27, 29, 32, 34, 36, 37, 38, 39, 41, 43, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56] have been developed to strengthen the security of face recognition systems. To counter various attacks from different domains, as shown in Figure 1 (a), some of previous methods have investigated using domain generalization (DG) [11, 15, 17, 19, 22, 26, 32, 33, 36, 37, 43, 49] to learn a generalized model from multiple source domains. If the target data are also available in the training stage, then domain adaptation (DA) [7, 14, 44, 58] has been adopted to adapt the source knowledge into the target domain during the model training, as shown in Figure 1 (b). Moreover, once the source data is unavailable for training because of data privacy, the idea of source-free domain adaptation (SFDA) [30], as shown in Figure 1 (c), has been proposed to directly fine-tune an off-the-shelf model on the target domain. Among the three settings in Figure 1, both DA and SFDA require the target data for adaptation during the offline training. However, because collection of all possible attack types is impossible in the offline training stage, the pre-adapted model may still fail to detect unseen attack types and need to adapt again during the online inference stage.

Unlike DA and SFDA, Test-Time Adaptation (TTA) considers a more realistic scenario that only an off-the-shelf model is available but the source domain data are inaccessible or no longer available. Hence, the goal of TTA is to online adapt this off-the-shelf model directly to the unlabeled target data in the inference stage. Many TTA methods [6, 20, 21, 31, 42] have been widely investigated in image classification and these methods generally fall into two categories: score-based [6, 31, 42] and class prototype-based [20, 21] approaches for pseudo-labeling the target data. Score-based methods [6, 31, 42] mainly used the prediction score of each class to determine the pseudo labels. Class prototype-based methods [20, 21] mostly assigned pseudo-labels to the target samples in terms of the similarity between the target samples and the class centers, which are initialized with the weights from the source model classifier and then are updated during the adaptation. Compared with the image classification task, FAS deals with highly similar visual characteristics between live and spoof faces and faces more challenges in TTA setting.

There are three major challenges in TTA for FAS. The first one is the noisy pseudo-label problem. When adapting a model to unlabeled target data, TTA faces the same challenge as DA on determining pseudo-labels for the target data to enable model adaptation. However, because live and spoof faces are visually similar, existing TTA methods are insufficient to offer reliable pseudo-live and pseudo-spoof labels to guide the online adaptation. Next, the second challenge concerns the data imbalance problem. Because real-world target data generally consist of only spoof or only live images, the two classes (i.e., live and spoof classes)

are not evenly distributed within a batch of target data. Moreover, in the realistic attack scenario, the attackers usually keep updating the attack types until successfully deceiving the model and thus result in a spoof-only dataset with time-variant characteristics. This data imbalance issue frequently leads the model to overfit to one dominant class and forget the previously acquired knowledge of the other class. Finally, the third challenge stems from the lack of prior knowledge about unseen attack types. Because there exist increasingly developed attacks, a pre-adapted model is totally oblivious to unseen attacks and is thus unable to detect the attacks in a new target domain.

In this paper, we aim to address the above-mentioned three challenges and propose a novel TTA framework 3A-TTA containing three major ideas: **A**ctivation-based pseudo-labeling, **A**nti-forgetting feature learning, and **A**symmetric prototype contrastive learning. First, to address the noisy pseudo-label problem, we found that either the score-based [6, 31, 42] or class prototype-based [20, 21] pseudo-labeling methods are inappropriate for FAS. Although the score-based methods [6, 31, 42] achieved great success in image classification involving multiple classes, the prediction scores for binary live/spoof classes become less informative and even unreliable when facing unseen attack types. On the other hand, the class prototype-based approaches [20, 21], which relied on class centers to determine pseudo labels, are unable to identify the spoof class containing various attack types. Unlike the live class, the spoof class consists of different attack types with their distinct characteristics. For example, print attacks usually exhibit grid artifacts, and replay attacks have visible moiré patterns. The single class center generated by the class prototype-based approaches is thus far from enough to describe the highly complex spoof class. Therefore, in this paper, we propose an activation-based pseudo-labeling by including fine-grained class information captured from class activation map [35] to tackle the noisy pseudo-label problem. Second, to address the data imbalance problem, we refer to the idea mentioned in [3] and propose an anti-forgetting feature learning strategy. Note that, in [3], the authors proposed to store source data to tackle the forgetting issue. However, in our TTA setting, there exist no source data in the online adaptation stage. We therefore propose a feature selection mechanism and preserve only the selected target data to prevent forgetting. Finally, to tackle the issue of unseen attack types, we propose an asymmetric prototype contrastive learning by associating similar characteristics between seen and unseen attacks. Although the spoof class covers various attack types with different characteristics, some attack types still share similarities with others. For example, spoof images of all attack types are supposed to have similar refection artifacts [24, 56]. Therefore, we propose to incorporate both global and nearest-neighbor information to devise the asymmetric contrastive learning.

To evaluate the proposed method 3A-TTA on TTA setting, we further propose a comprehensive FAS benchmark: TTA-FAS, i.e., Test Time Adaptation for Face Anti-Spoofing. In the TTA-FAS benchmark, we generate two difficult real-world cases, including: (1) unseen attack types, and (2) seen attack types from new domains, to simulate the realistic scenario of test-time adaption. Our experimental results on the proposed TTA-FAS benchmark not only verify the effectiveness of 3A-TTA in TTA setting but also indicate future research directions towards countering ever-evolving face presentation attacks.

Our contributions are summarized as follows:

• We propose a new benchmark TTA-FAS covering different domains and various attacks to simulate the real-world scenario. To the best of our knowledge, this is the first work focusing on addressing test-time adaptation for face anti-spoofing.

• We introduce a novel activation-based pseudo-labeling to tackle the noisy pseudo-label problem by including fine-grained class information from the class activation maps.

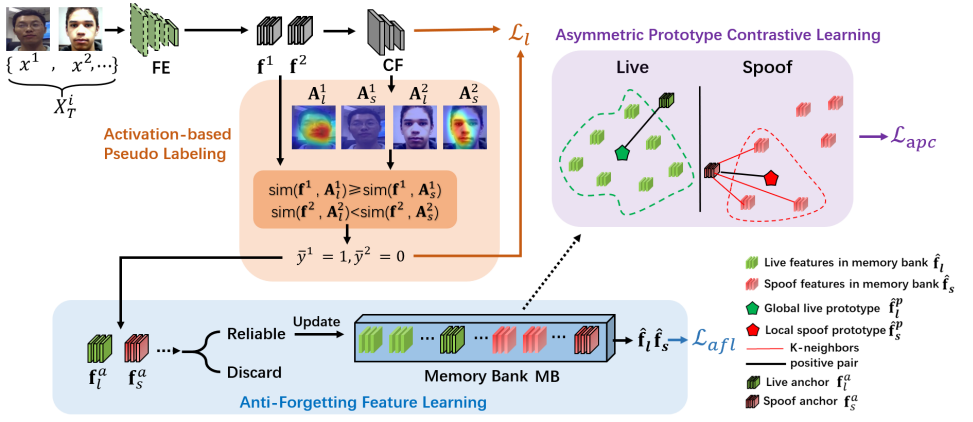

Figure 2: Overview of the proposed 3A-TTA framework, which consists of three main components: activation-based pseudo-labeling, anti-forgetting feature learning, and asymmetric prototype contrastive learning.

- We present an anti-forgetting feature learning to address the data imbalance problem.
- We develop an asymmetric prototype contrastive learning to learn distinctive feature representations by combining global and nearest-neighbor information.
- Extensive experiments demonstrate that 3A-TTA achieves superior performance for online detecting both seen and unseen types from new domains.

2 Proposed Method

In Section 2.1, we first present the problem statement of test-time adaptation in face anti-spoofing and give a brief overview of the proposed 3A-TTA framework. Next, in Sections 2.2-2.4, we present the three main components in 3A-TTA, including activation-based pseudo-labeling, anti-forgetting feature learning, and asymmetric prototype contrastive learning, respectively. Finally, in Section 2.5, we describe the total loss of the proposed method.

2.1 Problem Statement and Overview of 3A-TTA framework

In this paper, we address the issues of test-time adaptation in face anti-spoofing. Given an off-the-shelf anti-spoofing model, our goal is to adapt this model to the incoming batch of target sample $X_T^i = \{x^j\}_{j=1}^B$ (where i indicates the batch index and B is the batch size) and then conduct the inference subsequently. Note that each batch may contain multiple attack types and that some attack types are new to the source model.

As shown in Figure 2, we first use the proposed activation-based pseudo-labeling to determine the pseudo-labels for all the samples within the batch X_T^i. Then, we select reliable samples in terms of pseudo-labels and store their features in the memory bank. Next, we employ the proposed anti-forgetting feature learning to mitigate the risk of forgetting information. Finally, we adopt the proposed asymmetric prototype contrastive learning to increase the feature discriminability for classifying live and spoof images.

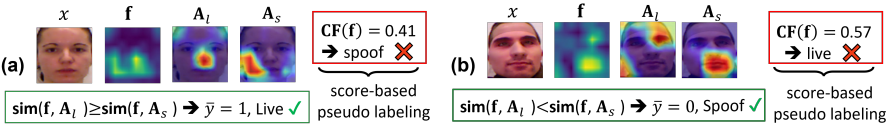

Figure 3: Examples of activation-based pseudo-labeling (green solid box) and score-based pseudo-labeling (red solid box) on (a) a live image and (b) a spoof image.

2.2 Activation-based Pseudo-Labeling

In [17], the authors showed that the class activation maps [35] offer discriminative as well as fine-grained information for identifying live and spoof faces. Inspired by [17], we propose using class activation maps [35] as a labeling criterion to address the noisy pseudo-label problem.

Given a target image x^j in a mini-batch, we first extract its liveness feature $\mathbf{f}$ by the feature extractor $\mathbf{FE}$ and then use the classifier $\mathbf{CF}$ to obtain its live activation map $\mathbf{A}_l$ and spoof activation map $\mathbf{A}_s$ by,

$$\mathbf{A}_l = \textbf{Grad-CAM}(\mathbf{CF(f)}; c=1); \mathbf{A}_s = \textbf{Grad-CAM}(\mathbf{CF(f)}; c=0), \tag{1}$$

where **Grad-CAM** indicates the class activation operation [35], and c indicates the class label. As mentioned in [35], the class activation map is a weighted liveness feature involving gradient information during the backpropagation process. Therefore, the class activation maps $\mathbf{A}_l$ and $\mathbf{A}_s$ are highly correlated with the liveness feature $\mathbf{f}$ and provide valuable information for determining the most possible pseudo labels. Hence, we assign a pseudo-label to each target sample by measuring the similarity between class activation maps and the liveness feature $\mathbf{f}$ by,

$$\bar{y} = \begin{cases} 1, & if \quad \textbf{sim}(\mathbf{f}, \mathbf{A}_l) \geq \textbf{sim}(\mathbf{f}, \mathbf{A}_s); \\ 0, & if \quad \textbf{sim}(\mathbf{f}, \mathbf{A}_l) < \textbf{sim}(\mathbf{f}, \mathbf{A}_s), \end{cases} \tag{2}$$

where $\textbf{sim}(\mathbf{f}, \mathbf{A}) = \frac{\mathbf{f} \cdot \mathbf{A}}{||\mathbf{f}|| ||\mathbf{A}||}$ is the cosine similarity function. Figure 3 gives some examples to demonstrate the effectiveness of the proposed activation-based pseudo-labeling. In particular, the activation-based pseudo-labeling is able to determine correct pseudo-labels for those misclassified by the score-based pseudo-labeling [3]. With the assigned pseudo-labels $\bar{y}$, we define the liveness loss $\mathcal{L}_l$ to adapt the model to new domains and diverse attacks by,

$$\mathcal{L}_l = -\sum \bar{y} log \mathbf{CF}(\mathbf{f}) + (1-\bar{y}) log(1 - \mathbf{CF}(\mathbf{f})), \tag{3}$$

where $\bar{y}$ is assigned pseudo-label, i.e., $\bar{y} = 1$ for live images and $\bar{y} = 0$ for spoof images.

2.3 Anti-Forgetting Feature Learning

To address the data imbalance problem, we propose an effective anti-forgetting feature learning by selecting reliable liveness features from the batch data and storing them in a memory bank **MB** to diminish the impact of noisy pseudo-labels and to facilitate robust feature learning. As shown in Figure 2, after obtaining the pseudo-label $\bar{y}$, we select the reliable features by the selection mechanism defined by:

$$\gamma = \begin{cases} 1, & if \quad \begin{cases} \mathbf{CF}(\mathbf{f}) > \alpha \quad and \quad m_{sim} \geq \beta; \\ \mathbf{CF}(\mathbf{f}) < 1-\alpha \quad and \quad m_{sim} \leq -\beta; \end{cases} \\ 0, & otherwise, \end{cases} \tag{4}$$

where $\alpha = 0.8$, $\beta = 0.2$, and $m_{sim} = \mathbf{sim}(\mathbf{f}, \mathbf{A}_l) - \mathbf{sim}(\mathbf{f}, \mathbf{A}_s)$. That is, we select those features with $\gamma = 1$ and store them in the memory bank **MB**. Specifically, we apply a First In First Out (FIFO) approach to update the memory bank and empirically set the size s of the memory bank for each class. With the stored features $\hat{\mathbf{f}}$ and their corresponding pseudo-labels $\hat{y}$ in **MB**, we define the anti-forgetting liveness loss $\mathcal{L}_{afl}$ by,

$$\mathcal{L}_{afl} = -\sum \hat{y} log \mathbf{CF}(\hat{\mathbf{f}}) + (1-\hat{y}) log(1-\mathbf{CF}(\hat{\mathbf{f}})). \tag{5}$$

2.4 Asymmetric Prototype Contrastive Learning

To address the lack of prior knowledge about unseen attack types, we propose an asymmetric prototype contrastive learning by incorporating local spoof information and global live information. Note that, although different spoof images have distinctive characteristics, we can still associate their similar characteristics between seen and unseen attacks. Therefore, we propose to aggregate local spoof information by referring to the nearest-neighbor spoof features in **MB**. In addition, because live images in different domains have small distribution discrepancies [17, 22], we are also able to cluster all the live images by globally aggregating their live information. As shown in Figure 2, according to the pseudo-label $\bar{y}$, we divide all the samples $\mathbf{f}$ in a mini-batch into a spoof anchor set and a live anchor set. Next, by referring to the features stored in the memory bank **MB**, we define the asymmetric prototype contrastive loss $\mathcal{L}_{apc}$ as follows:

$$\mathcal{L}_{apc} = -log \frac{\mathbf{exp}(\mathbf{sim}(\mathbf{f}_s^a, \hat{\mathbf{f}}_s^p))}{\sum_{j=\{\hat{\mathbf{f}}_s^p \cup N_s\}} \mathbf{exp}(\mathbf{sim}(\mathbf{f}_s^a, \hat{\mathbf{f}}^j))} - log \frac{\mathbf{exp}(\mathbf{sim}(\mathbf{f}_l^a, \hat{\mathbf{f}}_l^p))}{\sum_{i=\{\hat{\mathbf{f}}_l^p \cup N_l\}} \mathbf{exp}(\mathbf{sim}(\mathbf{f}_l^a, \hat{\mathbf{f}}^i))}. \tag{6}$$

where $\mathbf{f}_s^a$ and $\mathbf{f}_l^a$ are the spoof and live anchors, $\hat{\mathbf{f}}_s^p$ and $\hat{\mathbf{f}}_l^p$ are the local spoof prototype and the global live prototype, N_s is the set of negative pair of $\mathbf{f}_s^a$, and N_l is the set of the negative pair of $\mathbf{f}_l^a$.

For each spoof anchor $\mathbf{f}_s^a$, we determine its positive pair by calculating the spoof prototype $\hat{\mathbf{f}}_s^p$ from K nearest spoof features $\hat{\mathbf{f}}_s$, because these locally nearest neighbors usually share similar characteristics to the spoof anchor $\mathbf{f}_s^a$. This K- neighbor mechanism effectively enables the model to handle unseen attack samples through referring to similar features from seen attacks stored in **MB**. As to the negative pairs N_s of spoof anchor $\mathbf{f}_s^a$, we treat all the live features $\hat{\mathbf{f}}_l$ in **MB** as the negative pairs of $\mathbf{f}_s^a$.

For each live anchor $\mathbf{f}_l^a$, we calculate the average of all the live features in **MB** as a live prototype $\hat{\mathbf{f}}_l^p$ and use $\hat{\mathbf{f}}_l^p$ as its positive pair. Moreover, instead of pushing all the spoof features away, we consider only the K nearest spoof features $\hat{\mathbf{f}}_s$ from the live anchor $\mathbf{f}_l^a$ as negative pairs. In all our experiments, we use Euclidean distance to determine nearest neighbors.

2.5 Total Loss

Finally, we include the liveness loss $\mathcal{L}_l$, the anti-forgetting liveness loss $\mathcal{L}_{afl}$, and the asymmetric prototype contrastive loss $\mathcal{L}_{apc}$ to define the total loss $\mathcal{L}_T$ by,

$$\mathcal{L}_T = \mathcal{L}_l + \lambda_1 \mathcal{L}_{afl} + \lambda_2 \mathcal{L}_{apc}. \tag{7}$$

where λ_1 and λ_2 are the weight factors. In all our experiments, we empirically set $\lambda_1 = 0.5$ and $\lambda_2 = 1.5$.

Figure 4: Sample frames from six public FAS datasets: (a) OULU-NPU [4], (b) CASIA-MFSD [57], (c) MSU-MFSD [50], (d) Idiap Replay-Attack [8], (e) 3DMAD [10], and (f) HKBU-MARs [28]. These images consist of live faces (green solid box), print attacks (red solid box), replay attacks (red dotted box), and 3D mask attacks (magenta solid box).

Protocol	Subset	Attack Type	Real data (V/I)	Attack data (V/I)	All data (V/I)
[O,C,I] → [M,D,H]	Source: OCI	print, replay	1280	5110	6390
	Target: MDH	print,replay,3D Mask	339	355	694
[O,M,I] → [C,D,H]	Source: OMI	print, replay	1200	4360	5560
	Target: CDH	print,replay,3D Mask	419	595	1014
[O,C,M] → [I,D,H]	Source: OCM	print, replay	1210	4620	5830
	Target: IDH	print,replay,3D Mask	409	845	1254
[I,C,M] → [O,D,H]	Source: ICM	print, replay	350	1360	1710
	Target: ODH	print,replay,3D Mask	1259	4105	5364

Table 1: The proposed unseen attack testing from the proposed TTA-FAS benchmark.

3 Experiments

3.1 Test-Time Adaptation Benchmark for FAS (TTA-FAS)

In this paper, we propose a new Test-Time Adaptation benchmark for Face Anti-Spoofing (TTA-FAS) to study the challenges associated with the emergence of face presentation attacks. We construct TTA-FAS benchmark based on six publicly available face anti-spoofing (FAS) datasets, including **OULU-NPU** [4] (denoted as O), **CASIA-MFSD** [57] (denoted as C), **MSU-MFSD** [50] (denoted as M), **Idiap Replay-Attack** [8] (denoted as I), **3DMAD** [10] (denoted as D), and **HKBU-MARs** [28] (denoted as H). Figure 4 shows some examples of live and spoof images from these datasets, which include print attacks and replay attacks in Figure 4 (a)-(d), and 3D mask attacks in Figure 4 (e) and (f). In Table 1, we design the unseen attack testing for TTA-FAS benchmark to evaluate the efficacy of anti-spoofing models when encountering new face presentation attacks under test-time adaptation scenario. Note that, because the widely used DG protocols in FAS [17, 22, 36] use only one frame per video for model evaluation, we also adopt similar preprocessing in TTA-FAS benchmark to enable fair experimental comparisons. In addition, we design the leave-one-attack-out testing with limited training attack type for TTA-FAS benchmark to include "print", "replay" and "3D mask" attacks as new attack types, as summarized in Table 2. To obtain the pre-trained models of each protocol, we refer to [6, 20, 21, 31, 42] and pre-train the anti-spoofing models by using only the cross-entropy loss. These pre-trained models are available at https://github.com/Pei-KaiHuang/TTA-FAS for enabling fair comparisons on our proposed TTA-FAS benchmark.

Protocols	Subset	Attack Type	Protocols	Subset	Attack Type	Protocols	Subset	Attack Type
Protocol i	Source: OMI	Print	Protocol ii	Source: OMI	Replay	Protocol iii	Source: CDH	3D Mask
	Target: CDH	Replay + 3D Mask		Target: CDH	Print + 3D Mask		Target: OMI	Print + Replay

Table 2: The proposed leave-one-attack-out testing from the proposed TTA-FAS benchmark.

Method	Total Loss $\mathcal{L}_T$				pseudo-labeling Mechanisms			Feature Selection	[OMI] → D		[OMI] → C	
	$\mathcal{L}_l$	$\mathcal{L}_{afl}$	$\mathcal{L}_{apc}$	$\mathcal{L}_c$	Score based	Class Prototype based	Activation based		HTER	AUC	HTER	AUC
M0									26.86	87.83	28.78	86.26
M1	✓				✓				23.19	88.41	29.78	85.05
M2	✓					✓			27.72	88.28	30.24	86.05
M3	✓						✓		21.87	88.68	26.02	86.44
M4	✓	✓					✓		19.31	88.49	24.94	86.29
M5	✓	✓					✓	✓	18.15	89.47	24.33	87.07
M6	✓	✓		✓			✓	✓	18.20	87.78	26.00	86.56
M7	✓	✓	✓				✓	✓	**17.21**	**90.63**	**23.55**	**87.29**

Table 3: Ablation study on the protocol [O,M,I] → D and [O,M,I] → C, using different combinations of loss terms, different pseudo-labeling mechanisms, and feature selection. The evaluation metrics are HTER(%) ↓ and AUC(%) ↑.

3.2 Evaluation Metrics and Implementation Details

We evaluate our method on the proposed TTA-FAS benchmark and report the results using different evaluation metrics, including Half Total Error Rate (HTER) [2], Area Under Curve (AUC), and the total running time (second) across all target datasets. To have a fair comparison, we continuously conduct the experiment ten times and report the average results as the final outcomes. More details are given in the supplemental materials.

3.3 Ablation Study

3.3.1 Different Combinations of Loss Terms and Modules

In Table 3, we compare using different combinations of loss terms and pseudo-labeling mechanisms to update **FE** and **CF** during the adaptation process of inference stage. Note that, "M0" indicates that the pre-trained model is fixed and does not adapt to the incoming target samples. Here, we test various FAS models on the proposed protocols [**O,M,I**] → **D** and [**O,M,I**] → **C** from Table 1 for covering the 3D mask attack, print attack and replay attack during testing.

First, we compare using different pseudo-labeling mechanisms, including the score-based pseudo-labeling [3, 6, 31, 42], the class prototype-based pseudo-labeling [20, 21], and the proposed activation-based pseudo-labeling, to evaluate the reliability of assigned pseudo-labels. When including $\mathcal{L}_l$ and different pseudo-labeling mechanisms to fine-tune **FE** and **CF**, we show that the proposed activation-based pseudo-labeling (M3) outperforms the score-based pseudo-labeling (M1) and the class prototype-based pseudo-labeling (M2) in terms of reliability of pseudo labels.

Next, when $\mathcal{L}_{afl}$ is included, both M4 (i.e., without feature selection) and M5 (i.e., with feature selection) result in improved performance over M3 and demonstrate the efficacy of our proposed anti-forgetting feature learning. In addition, when including $\mathcal{L}_l + \mathcal{L}_{afl}$ with feature selection in M5, we observe that M5 outperforms M4 because the proposed anti-forgetting feature learning indeed benefits from reliable features and mitigates the noisy pseudo-label issue.

Moreover, if further including the regular supervised contrastive learning loss $\mathcal{L}_c$ [9], i.e., the case of M6, we see degraded performance resulted by the aggregation of different attacks into the spoof class. When replacing $\mathcal{L}_c$ with the proposed $\mathcal{L}_{apc}$, i.e., the case of M7, we show that the proposed asymmetric prototype contrastive learning effectively handles

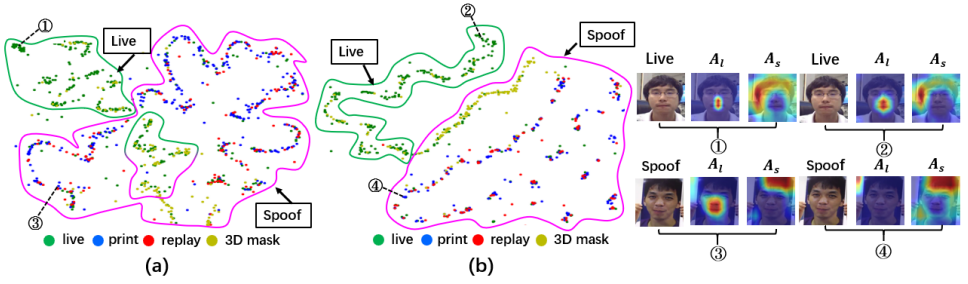

Figure 5: The t-SNE [40] visualization and activation visualization obtained by the anti-spoofing model without adaptation, i.e., (a), ①, and ③, and the proposed 3A-TTA , i.e., (b), ②, and ④, on the protocol [O,M,I] → [C,D,H].

Method	[O,C,I] → [M,D,H]									[O,M,I] → [C,D,H]								
	O,C,I → M		O,C,I → D		O,C,I → H		Average		Time	O,M,I → C		O,M,I → D		O,M,I → H		Average		Time
	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC		HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	
No adaptation	26.67	94.49	19.55	88.11	22.15	84.33	22.79	88.98	0.50	28.78	86.26	26.86	87.83	23.47	84.91	26.37	86.33	0.62
Tent [43]	27.98	94.49	22.67	87.44	22.49	84.55	24.38	88.83	1.061	28.14	79.68	46.10	53.69	28.54	79.36	34.26	70.91	1.36
OAP [8]	26.41	94.49	19.79	88.09	22.15	84.35	22.78	87.35	0.55	29.34	86.03	26.86	87.78	22.95	85.86	25.38	86.55	0.70
3A-TTA	**26.21**	**94.53**	**16.26**	**92.03**	**20.89**	**84.74**	**21.12**	**90.43**	4.35	**23.55**	**87.29**	**17.21**	**90.63**	**20.33**	**86.99**	**20.36**	**88.30**	7.12
Method	[O,C,M] → [I,D,H]									[I,C,M] → [O,D,H]								
	O,C,M → I		O,C,M → D		O,C,M → H		Average		Time	I,C,M → O		I,C,M → D		I,C,M → H		Average		Time
	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC		HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	
No adaptation	30.36	71.22	25.27	83.89	19.93	90.08	25.19	81.73	0.71	37.73	81.95	25.80	**81.79**	34.93	83.88	32.82	**82.54**	2.28
Tent [43]	35.73	70.16	25.43	84.12	22.28	89.81	27.81	81.36	1.57	47.01	64.23	26.43	80.11	42.43	83.40	38.62	75.91	8.72
OAP [8]	29.69	71.15	25.15	83.81	19.93	90.09	24.92	81.68	0.81	31.21	78.50	25.62	81.55	35.41	83.65	30.75	81.23	2.28
3A-TTA	**28.11**	**72.45**	**21.78**	**86.28**	**16.99**	**90.36**	**22.29**	**83.03**	8.72	**25.62**	**82.25**	**24.35**	80.06	**30.71**	**84.41**	**26.89**	82.24	37.06

Table 4: Experimental comparisons on the proposed TTA-FAS benchmark. The evaluation metrics are HTER(%) ↓, AUC(%) ↑, and the total running time (s) across all target datasets.

unseen attack samples by referring to spoof features with similar characteristics stored in **MB** and achieves the best performance.

3.3.2 Activation maps and t-SNE Visualization

In Figure 5, we first use t-SNE [40] to visualize the latent liveness features obtained without adaptation and with the proposed 3A-TTA. When the anti-spoofing model is not adapted to the target data, in Figures 5 (a), the visualization result shows that the live features tend to overlap extensively with the spoof features when encountering unseen attacks from new domains. By contrast, when the anti-spoofing model is adapted by the proposed 3A-TTA, in Figures 5 (b), the live and spoof features obtained by the proposed 3A-TTA are diversely distributed and strongly support the model to adapt to unseen domains. Next, we also visualize some activation maps to demonstrate the efficacy of the proposed 3A-TTA. By comparing the live and spoof activation maps between the anti-spoofing model without adaptation (i.e., Figures 5 ① and ③) and the proposed 3A-TTA (i.e., Figures 5 ② and ④), we see that: (1) the live activation map obtained by the proposed 3A-TTA (Figures 5 ②) mostly locates on the facial regions; and (2) the spoof image has hardly any response on the live activation map obtained by 3A-TTA (Figures 5 ④). These visualization results show excellent ability of the proposed 3A-TTA for encountering unseen attacks from new domains.

Method	Protocol i			Protocol ii			Protocol iii		
	HTER	AUC	Time	HTER	AUC	Time	HTER	AUC	Time
No adaptation	23.97	80.61	1.86	23.15	81.04	1.67	35.80	74.43	3.57
Tent [43]	21.81	81.05	3.70	22.56	80.15	3.32	45.53	67.37	7.09
OAP [3]	24.60	77.58	2.17	23.33	78.89	1.96	33.21	72.39	4.14
3A-TTA	**20.46**	**82.14**	24.29	**21.05**	**83.71**	21.99	**31.53**	**75.17**	48.26

Table 5: Experimental comparisons on new protocols.

3.4 Experimental Comparisons on the proposed TTA-FAS benchmark

In Table 4, we present the evaluation results of our proposed 3A-TTA method and compare with other test-time adaptation methods on TTA-FAS benchmark. We re-implement all the compared methods and remove the source samples in OAP [3] to conduct test-time adaptation. Note that, "No adaptation" indicates that the anti-spoofing model is fixed during the inference stage. First, we observe that Tent [42] yields poorer performance than the baseline "No adaptation" because these methods are not specifically designed for face anti-spoofing. Next, we observe that the FAS method OAP [3] only slightly improves the performance compared to "No adaptation" by simply using the score-based pseudo-labeling to update the classifier during adaptation. In contrast, the proposed 3A-TTA outperforms previous methods and achieves 12.54% improvements in HTER and 2.11% on AUC over OAP [3] in average among all protocols. In addition, we also record the total running time (seconds) across all the target datasets. The results show that the proposed method requires a response time of only 0.006s for each target sample. Next, in Table 5, we evaluate the proposed 3A-TTA method on the proposed leave-one-attack-out testing. Table 5 shows our experimental results on these protocols and clearly verifies the effectiveness of our proposed method in detecting various new attack types.

4 Conclusion

In this paper, we propose a new benchmark TTA-FAS and a novel test-time adaptation approach 3A-TTA for face anti-spoofing. In 3A-TTA, we first design an effective activation-based labeling mechanism to handle the noisy pseudo-label problem and also to stabilize the adaptation process. Next, with the proposed anti-forgetting feature learning, we enable the anti-spoofing model to retain the information of the non-dominant class and to preserve discriminative capability when adapting to a data-imbalanced batch. Furthermore, we devise an asymmetric prototype contrastive learning to increase feature discriminability between live and spoof faces. Our experimental results on the proposed benchmark TTA-FAS demonstrate the effectiveness of the proposed method on handling realistic TTA scenarios and also show its outperformance over existing test-time adaptation methods.

References

[1] Akshay Agarwal, Richa Singh, Mayank Vatsa, and Afzel Noore. Boosting face presentation attack detection in multi-spectral videos through score fusion of wavelet partition images. *Frontiers in big Data*, page 53, 2022.

[2] André Anjos and Sébastien Marcel. Counter-measures to photo attacks in face recognition: a public database and a baseline. In *2011 international joint conference on Biometrics (IJCB)*, pages 1–7. IEEE, 2011.

[3] Davide Belli, Debasmit Das, Bence Major, and Fatih Porikli. Online adaptive personalization for face anti-spoofing. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 351–355. IEEE, 2022.

[4] Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. Oulu-npu: A mobile face presentation attack database with real-world variations. In *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*, pages 612–618. IEEE, 2017.

[5] Baoliang Chen, Wenhan Yang, Haoliang Li, Shiqi Wang, and Sam Kwong. Camera invariant feature learning for generalized face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 16:2477–2492, 2021.

[6] Dian Chen, Dequan Wang, Trevor Darrell, and Sayna Ebrahimi. Contrastive test-time adaptation. In *CVPR*, 2022.

[7] Zhihong Chen, Taiping Yao, Kekai Sheng, Shouhong Ding, Ying Tai, Jilin Li, Feiyue Huang, and Xinyu Jin. Generalizable representation learning for mixture domain face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1132–1139, 2021.

[8] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG)*, pages 1–7. IEEE, 2012.

[9] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 539–546. IEEE, 2005.

[10] Nesli Erdogmus and Sebastien Marcel. Spoofing face recognition with 3d masks. *IEEE transactions on information forensics and security*, 9(7):1084–1097, 2014.

[11] Meiling Fang, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Learnable multi-level frequency decomposition and hierarchical attention mechanism for generalized face presentation attack detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3722–3731, 2022.

[12] Haocheng Feng, Zhibin Hong, Haixiao Yue, Yang Chen, Keyao Wang, Junyu Han, Jingtuo Liu, and Errui Ding. Learning generalized spoof cues for face anti-spoofing. *arXiv preprint arXiv:2005.03922*, 2020.

[13] Anjith George and Sébastien Marcel. On the effectiveness of vision transformers for zero-shot face anti-spoofing. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2021.

[14] Hsin-Ping Huang, Deqing Sun, Yaojie Liu, Wen-Sheng Chu, Taihong Xiao, Jinwei Yuan, Hartwig Adam, and Ming-Hsuan Yang. Adaptive transformers for robust few-shot cross-domain face anti-spoofing. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIII*, pages 37–54. Springer, 2022.

[15] Pei-Kai Huang, Chu-Ling Chang, Hui-Yu Ni, and Chiou-Ting Hsu. Learning to augment face presentation attack dataset via disentangled feature learning from limited spoof data. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2022.

[16] Pei-Kai Huang, Ming-Chieh Chin, and Chiou-Ting Hsu. Face anti-spoofing via robust auxiliary estimation and discriminative feature learning. In *Asian Conference on Pattern Recognition*, pages 443–458. Springer, 2022.

[17] Pei-Kai Huang, Hui-Yu Ni, Yan-Qin Ni, and Chiou-Ting Hsu. Learnable descriptive convolutional network for face anti-spoofing. In *BMVC*, 2022.

[18] Pei-Kai Huang, Cheng-Hsuan Chiang, Jun-Xiong Chong, Tzu-Hsien Chen, Hui-Yu Ni, and Chiou-Ting Hsu. Ldcformer: Incorporating learnable descriptive convolution to vision transformer for face anti-spoofing. In *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023.

[19] Pei-Kai Huang, Jun-Xiong Chong, Hui-Yu Ni, Tzu-Hsien Chen, and Chiou-Ting Hsu. Towards diverse liveness feature representation and domain expansion for cross-domain face anti-spoofing. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2023.

[20] Yusuke Iwasawa and Yutaka Matsuo. Test-time classifier adjustment module for model-agnostic domain generalization. *Advances in Neural Information Processing Systems*, 34:2427–2440, 2021.

[21] Minguk Jang, Sae-Young Chung, and Hye Won Chung. Test-time adaptation via self-training with nearest neighbor information. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=EzLtB4M1SbM.

[22] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Single-side domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8484–8493, 2020.

[23] Amin Jourabloo, Yaojie Liu, and Xiaoming Liu. Face de-spoofing: Anti-spoofing via noise modeling. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 290–306, 2018.

[24] Taewook Kim, YongHyun Kim, Inhan Kim, and Daijin Kim. Basn: Enriching feature representation using bipartite auxiliary supervisions for face anti-spoofing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[25] Xiaobai Li, Jukka Komulainen, Guoying Zhao, Pong-Chi Yuen, and Matti Pietikäinen. Generalized face anti-spoofing by detecting pulse from face videos. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 4244–4249. IEEE, 2016.

[26] Shubao Liu, Ke-Yue Zhang, Taiping Yao, Mingwei Bi, Shouhong Ding, Jilin Li, Feiyue Huang, and Lizhuang Ma. Adaptive normalized representation learning for generalizable face anti-spoofing. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 1469–1477, 2021.

[27] Si-Qi Liu, Xiangyuan Lan, and Pong C Yuen. Remote photoplethysmography correspondence feature for 3d mask face presentation attack detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 558–573, 2018.

[28] Siqi Liu, Pong C Yuen, Shengping Zhang, and Guoying Zhao. 3d mask face anti-spoofing with remote photoplethysmography. In *European Conference on Computer Vision*, pages 85–100. Springer, 2016.

[29] Yaojie Liu, Joel Stehouwer, and Xiaoming Liu. On disentangling spoof trace for generic face anti-spoofing. In *European Conference on Computer Vision*, pages 406–422. Springer, 2020.

[30] 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 *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XII*, pages 511–528. Springer, 2022.

[31] Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Zhiquan Wen, Yaofo Chen, Peilin Zhao, and Mingkui Tan. Towards stable test-time adaptation in dynamic wild world. In *Internetional Conference on Learning Representations*, 2023.

[32] Yunxiao Qin, Zitong Yu, Longbin Yan, Zezheng Wang, Chenxu Zhao, and Zhen Lei. Meta-teacher for face anti-spoofing. *IEEE transactions on pattern analysis and machine intelligence*, 2021.

[33] Ruijie Quan, Yu Wu, Xin Yu, and Yi Yang. Progressive transfer learning for face anti-spoofing. *IEEE Transactions on Image Processing*, 30:3946–3955, 2021.

[34] Nilay Sanghvi, Sushant Kumar Singh, Akshay Agarwal, Mayank Vatsa, and Richa Singh. Mixnet for generalized face presentation attack detection. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 5511–5518. IEEE, 2021.

[35] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.

[36] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C Yuen. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10023–10031, 2019.

[37] Rui Shao, Xiangyuan Lan, and Pong C Yuen. Regularized fine-grained meta face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11974–11981, 2020.

[38] Yiyou Sun, Yaojie Liu, Xiaoming Liu, Yixuan Li, and Wen-Sheng Chu. Rethinking domain generalization for face anti-spoofing: Separability and alignment. *arXiv preprint arXiv:2303.13662*, 2023.

[39] Xiaoguang Tu, Zheng Ma, Jian Zhao, Guodong Du, Mei Xie, and Jiashi Feng. Learning generalizable and identity-discriminative representations for face anti-spoofing. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(5):1–19, 2020.

[40] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[41] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20281–20290, 2022.

[42] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=uXl3bZLkr3c.

[43] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Cross-domain face presentation attack detection via multi-domain disentangled representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6678–6687, 2020.

[44] Jingjing Wang, Jingyi Zhang, Ying Bian, Youyi Cai, Chunmao Wang, and Shiliang Pu. Self-domain adaptation for face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2746–2754, 2021.

[45] Yu-Chun Wang, Chien-Yi Wang, and Shang-Hong Lai. Disentangled representation with dual-stage feature learning for face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1955–1964, 2022.

[46] Zezheng Wang, Zitong Yu, Chenxu Zhao, Xiangyu Zhu, Yunxiao Qin, Qiusheng Zhou, Feng Zhou, and Zhen Lei. Deep spatial gradient and temporal depth learning for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5042–5051, 2020.

[47] Zhuo Wang, Qiangchang Wang, Weihong Deng, and Guo dong Guo. Learning multi-granularity temporal characteristics for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 17:1254–1269, 2022.

[48] Zhuo Wang, Qiangchang Wang, Weihong Deng, and Guodong Guo. Face anti-spoofing using transformers with relation-aware mechanism. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(3):439–450, 2022.

[49] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4123–4133, 2022.

[50] Di Wen, Hu Han, and Anil K Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015.

[51] Zitong Yu, Jun Wan, Yunxiao Qin, Xiaobai Li, Stan Z Li, and Guoying Zhao. Nas-fas: Static-dynamic central difference network search for face anti-spoofing. *IEEE transactions on pattern analysis and machine intelligence*, 43(9):3005–3023, 2020.

[52] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5295–5305, 2020.

[53] Ke-Yue Zhang, Taiping Yao, Jian Zhang, Ying Tai, Shouhong Ding, Jilin Li, Feiyue Huang, Haichuan Song, and Lizhuang Ma. Face anti-spoofing via disentangled representation learning. In *European Conference on Computer Vision*, pages 641–657. Springer, 2020.

[54] Ke-Yue Zhang, Taiping Yao, Jian Zhang, Shice Liu, Bangjie Yin, Shouhong Ding, and Jilin Li. Structure destruction and content combination for face anti-spoofing. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–6. IEEE, 2021.

[55] Shifeng Zhang, Ajian Liu, Jun Wan, Yanyan Liang, Guodong Guo, Sergio Escalera, Hugo Jair Escalante, and Stan Z Li. Casia-surf: A large-scale multi-modal benchmark for face anti-spoofing. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(2):182–193, 2020.

[56] 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 *European Conference on Computer Vision*, pages 70–85. Springer, 2020.

[57] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z Li. A face anti-spoofing database with diverse attacks. In *2012 5th IAPR international conference on Biometrics (ICB)*, pages 26–31. IEEE, 2012.

[58] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Ran Yi, Kekai Sheng, Shouhong Ding, and Lizhuang Ma. Generative domain adaptation for face anti-spoofing. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part V*, pages 335–356. Springer, 2022.