A Personalized Benchmark for Face Anti-spoofing

Davide Belli Debasmit Das Bence Major Fatih Porikli
Qualcomm AI Research[*]
{dbelli, debadas, bence, fporikli}@qti.qualcomm.com

Abstract

Thanks to their ease-of-use and effectiveness, face authentication systems are nowadays ubiquitous in electronic devices to control access to protected data. However, the widespread adoption of such systems comes with security and reliability issues. This is because spoofs of face images can be easily fabricated to deceive the recognition systems. Hence, there is a need to integrate the user identification system with a robust face anti-spoofing element, which has the goal to detect whether a queried face image is a spoof or live. Most contemporary face anti-spoofing systems only rely on the query image to accept or reject tentative access. In real-world scenarios, however, face authentication systems often have an initial enrollment step where a few live images of the user are recorded and stored for identification purposes [23, 18, 33]. In this paper, we present a complementary approach to augment existing face anti-spoofing benchmarks to account for enrollment images associated with each query image. We apply this strategy on two recently introduced datasets: CelebA-Spoof [53] and SiW [29]. We showcase how existing anti-spoofing models can be easily personalized using the subject's enrollment data, and we evaluate the effectiveness of the enhanced methods on the newly proposed datasets splits CelebA-Spoof-Enroll and SiW-Enroll.

1. Introduction

Face images as biometric signals are commonly being used in our daily lives for access control and authentication. With the widespread use of social networks and search engines, face images of individuals can be easily located and downloaded by anyone and anywhere in the world. Perpetrators can create copies (or spoofs) of an individual's facial images, use them to break into supposedly safeguarded systems, and then access the content in an unauthorized manner. Therefore, most recent face recognition systems are paired with face-antispoofing components to ensure the security of users' privacy and property.

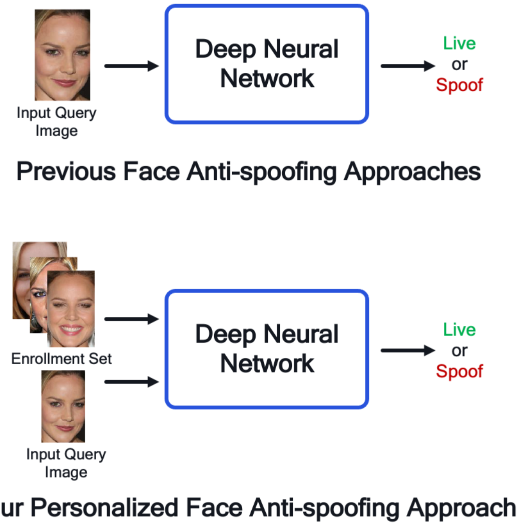

Figure 1. Existing methods use standalone query images to output live or spoof predictions. We personalize our anti-spoofing solution by providing a subject-specific reference, the enrollment images, as an additional input to the model.

Most research efforts on face anti-spoofing is focused on developing sophisticated strategies [29, 21, 28, 39, 19], establishing new and challenging benchmarks [29, 30, 53, 36] and sometimes using additional modalities [52, 10, 40] to aid spoof detection. As depicted in Fig. 1, existing methods rely on training a binary classification network that takes a query image as input and predicts whether it is live or spoof. Existing benchmark datasets mainly contain live and spoof labeled images captured under different conditions. Certain types of spoofs are difficult to distinguish from live ones. Hence, modalities like depth [29] and reflection maps [21] have been introduced in recent datasets to assist the design of effective anti-spoofing systems.

Nevertheless, additional modalities like depth and reflection maps can only be captured using specific sensors, and their usability in low-power mobile devices is limited due to cost and power constraints. Moreover, predictors using

[*]Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

depth and reflection maps require accurate calibration and might render inaccurate results if the sensor configuration at test time differs from training data. This limitation also applies to models trained on RGB images since any training bias in the distribution of subject, sensor, environmental and facial factors might limit the model performance once evaluated on different unseen data.

Systems that support anti-spoofing capabilities are commonly paired with a user identification module before allowing or denying access to protected content. Every time the user requests access, a *query* image is captured and examined by both the anti-spoofing and user identification routines. If only both checks are satisfied, the request is granted for access. The identification module relies on an *enrollment* set which is a collection of live face images that a user records and stores on the device to enable user identification [23, 18, 33]. To the best of our knowledge, there has been no previous work that uses enrollment information as a complementary input to face anti-spoofing models. Since the enrollment set is unique to each subject and sensor, it provides valuable personalized information for the anti-spoofing task. Unlike depth and reflection maps, user-specific enrolled images are readily available and easy to obtain. Enrollment information has therefore the potential to improve anti-spoofing by acting as a live reference against which the system can compare the query. We call this method *personalized* face anti-spoofing.

In this paper, we first devise a methodology to convert an existing face anti-spoofing benchmark into its personalized version. In the original version of the benchmark, each query image is standalone and solely used to predict live or spoof classes. In the personalized version, each query image has an associated enrollment set, and both the query image and the enrolment set are used for prediction. The enrollment set for a query sample is found by first identifying the user. Then, a fixed number of enrollment images are set aside from the live images of that user. This process is repeated for all users who can meet the predefined number of enrollment images in both the training and testing sets.

We also propose multiple personalized baselines, using different modules like attention, GRUs, and GNNs, to provide initial benchmarks on the new face anti-spoofing datasets. We investigate alternative methods to extract information from the enrollment images and then aggregate it into a single representation. This representation is used as a reference against the query for the anti-spoofing task. We use experimental evaluation under different conditions to prove the effectiveness of personalization for face anti-spoofing. We also conduct ablation studies to investigate which methods are most effective for encoding enrollment information and what type of features the personalized model learns.

To summarize, our contributions are as follows:

- We introduce a method to convert existing anti-spoofing datasets into a personalized version where each query image has an associated enrollment set unique for each user;
- We introduce multiple personalized baselines to showcase how enrollment and query information can be jointly used for face anti-spoofing
- We experiment with different configurations and ablation studies to analyze the importance of enrollment information for the anti-spoofing task. We anticipate our benchmark and results to inspire and catalyze further research on personalized anti-spoofing.

The remaining of the paper is structured as follows. In Section 2 we present existing face anti-spoofing datasets and discuss related work on generic and personalized face anti-spoofing. Section 3 explains how an existing dataset can be converted into its personalized version, with concrete examples for CelebA-Spoof and SiW. Section 4 introduces a selection of simple methods to personalize face anti-spoofing backbones, and Section 5 evaluates through experimental results the impact of personalization for the face anti-spoofing task. We finish in Section 6 with conclusions and directions for future work.

2. Related Work

Face Anti-spoofing Datasets. In the latest decade, multiple face anti-spoofing datasets have been introduced. Most of these datasets vary in quantity and quality of spoof attacks, subjects and environmental conditions. Commonly adopted image-based datasets are Replay Attack [7], MSU-MFSD [46], MSU-USSA [35] and CASIA-MFSD [54]. Since images in these datasets have been acquired using low-quality sensors, new benchmarks have been later introduced using high-quality sensors for image capturing. These more recent datasets include Oulu-NPU [5] and HKBU [28]. On top of these, multi-modal datasets like CASIA-SURF [52], 3DMAD [10], MSSpoof [8] and CS-MAD [2] have been proposed, introducing extra modalities like depth and IR in addition to RGB images. Datasets mentioned so far are limited in the variety of subjects, environments and spoof attack types, which limits the generalization capabilities of models trained on them. To address these limitations, two new datasets have been recently introduced: SiW [29] and CelebA-Spoof [53]. These datasets contain a large variety of spoofs, sensors, illumination and environmental conditions with non-overlapping splits between training and testing sets. For this reason, they serve as better benchmarks to evaluate the generalization capability of anti-spoofing models. In this paper, we introduce new versions of CelebA-Spoof and SiW datasets by defining enrollment sets for each subject. We name the new datasets

CelebA-Spoof-Enroll and SiW-Enroll, and in Section 3 we describe the approach we propose to add enrollment samples to these datasets. We further show how enrollment data can be used to personalize the anti-spoofing model.

Generic Face Anti-spoofing Techniques. Face anti-spoofing has been the subject of research for a long time, but the field has recently garnered increased attention within the deep-learning community due to the widespread use of authentication systems based on face biometrics. Before the rise of deep-learning, people proposed binary classifiers on top of handcrafted features like SURF [3], HoG [31, 47, 38], LBP [7, 31, 32, 47], etc. In addition to these features, people explored other robust input spaces like color [3, 4] and frequency [25] spectrum. Even additional temporal cues from eyes [34, 43] and lips [22] were used to improve spoof detection performance. With the advent of deep-learning, convolutional neural networks [15, 26, 14, 50] became a common choice as feature extractors paired with dense classifiers. Sometimes additional supervision is used in a multi-task learning setting. This includes predicting depth maps [29, 21], reflection maps [21] and rPPG signals [28, 27] in addition to binary live or spoof labels. Alternative approaches to face anti-spoofing have also been proposed to improve generalization. For example, Shao et al. [39] use adversarial approaches to learn domain-invariant representations from multiple face anti-spoofing datasets that generalize to a novel face anti-spoofing dataset. In [19], the authors propose a method to decompose a spoof image into its corresponding live image and a residual map capturing spoof traits. Yang et al. [49] propose to use both spatial and temporal information while attending to discriminative regions in consecutive video frames to improve generalization.

Personalized Face Anti-spoofing Techniques While the general trend in face anti-spoofing research is to employ domain adaptation methods and use additional data sources to improve generalization, research investigating personalization of face anti-spoofing methods using enrollment data is scarce. However, personalization of neural networks has already been proven effective for the tasks of speaker identification and anti-spoofing using audio data [20, 16, 24], which suggests similar improvements can be achieved for face anti-spoofing. Some of the existing literature investigated personalized face anti-spoofing through anomaly detection methods. In [11], the authors use person-specific thresholds for anomaly detection calculated from score distributions for each subject, while [13, 12] use a person-specific stacked ensemble optimized with a genetic algorithm. Both solutions rely on indirect person-specific information like thresholds and pruning coefficients to guide learning while our method makes use of direct personal information in the form of enrollment images.

Other person-specific face anti-spoofing methods focus instead on domain adaptation. [55] proposed augmenting the training data by generating fake spoofs based on personal features. [48] proposes instead to train a classifier for each subject, and use iterative matching to synthesize spoofs of subjects for which only live information is available. Both methods assume that the generative models can accurately synthesize meaningful spoofs and that the distribution of spoof types does not change during test time. Moreover, all the personalized approaches we mentioned so far require access to both source and target data, either for the personalization of the model, or to optimize the generation of person-specific spoofs. Hence, these methods cannot be applied out of the box to unseen test data or new subjects.

Finally, [1] suggested using the distribution of model predictions on the enrollment images to tune the anti-spoofing threshold at test time, but did not consider using this data as an additional training signal for the anti-spoofing model.

In this paper, we instead propose to improve the anti-spoofing solution by conditioning the model predictions with user-specific enrollment data. We call this technique *personalization*, as the model is provided with additional reference information which always comes from the user's live samples. In contrast with previous personalization techniques for face anti-spoofing [11, 12, 13, 55, 48], our method does not need any additional information about the test data distribution during training time. To the best of our knowledge, there is no existing work exploring the usage of enrollment data for the personalization of a deep learning face anti-spoofing model.

3. Personalized Benchmarks

In this section, we describe how to convert a given anti-spoofing dataset into its personalized version. A personalized version of a dataset contains an associated enrollment set for each query image. Specifically, we describe our approach to convert the CelebA-Spoof and the SiW dataset into their personalized version. Note, however, this approach can be easily applied to any other datasets in which subject metadata is available.

The task of converting an existing original dataset D_o into its personalized version D_p can be formulated as follows. The input to the conversion process is the original dataset, containing a number of data points:

$$d_i = (I_q^{(i)}, t_q^{(i)}) \quad \text{with} \quad 0 \leq i \leq |D_o|. \tag{1}$$

Each data point d_i is represented by its query image $I_q^{(i)}$ and the binary label $t_q^{(i)} \in \{live, spoof\}$ describing the query's class. The result of the conversion is a personalized dataset where each data point contains an enrollment set in addition to query and label data:

$$d_i = (I_q^{(i)}, t_q^{(i)}, \mathbf{e}^{(i)}) \quad \text{with} \quad 0 \leq i \leq |D_p|. \tag{2}$$

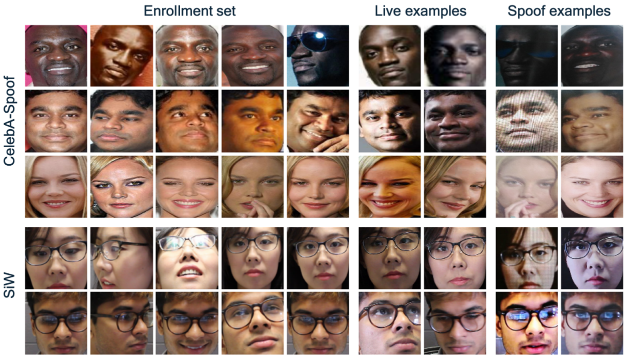

Figure 2. Examples from the enrollment-enabled CelebA-Spoof-Enroll5 and SiW-Enroll5 datasets. Images in SiW vary in pose and illumination, while CelebA-Spoof samples additionally show changes in resolution, color distribution and spoof quality.

Here, $\mathbf{e}^{(i)} = ({I_e}^1, ..., {I_e}^N)$ is the enrollment set for data point i, which includes N live images from that subject. All data points from the same subject share the same enrollment set. Notice how this process must transfer part of the original queries into live enrollment images, thus reducing the number of data points in the personalized dataset to $|D_p| < |D_o|$. In the rest of this section, we discuss in detail how we applied this conversion process to CelebA-Spoof and SiW datasets.

3.1. CelebA-Spoof-Enroll

CelebA-Spoof is a large-scale face anti-spoofing dataset recently introduced in [53]. The dataset contains 625,537 images of 10,177 celebrities captured under different spoof mediums, environments and illumination conditions. The original dataset proposes three different evaluation protocols. For our experimentation, we focus on the most general ”intra” protocol, in which different spoof types, environments and illumination conditions are used for both training and testing.

To generate **CelebA-Spoof-Enroll**, the personalized version of the CelebA-Spoof anti-spoofing dataset, we start by setting the desired enrollment set size N. We decide for a constant number of enrollment images per user to be consistent with the implementation in typical commercial applications and to simplify the dataset definition. For both the training and test split, we count the number of live samples per user and discard those users having $\leq N$ live samples. So, if we desire a higher value of N, a larger number of users would get rejected and hence the number of training samples would be less. Note that it is not possible to include all original data, as a number of users in CelebA-Spoof are missing live samples. Nevertheless, only a very small percentage of training and test data is discarded through this process when choosing $N < 10$.

For each accepted user, the first N live samples (ordered according to the ascending alphanumeric ordering of the original filenames) are chosen to define its *enrollment* set. The rest of the live samples and the spoof samples are marked as *query* samples. It is important to note that using this deterministic method of obtaining enrollment samples does not introduce unwanted bias as most of the CelebA-Spoof images are randomly crawled from the internet and are not ordered according to specific criteria. With this setting, we associate a user identifier with each query sample such that queries can be easily mapped to the correct enrollment set. This method of filtering users for a desired enrollment size N is performed for both training and test split. In the rest of the paper, we refer to CelebA-Spoof-EnrollN or in short CASp-EnrollN to describe the personalized version of the dataset with N enrollment images.

3.2. SiW-Enroll

SiW (Spoofing in the Wild) is a face anti-spoofing dataset recently introduced in [29] where images are extracted from short videos captured at high resolution and 30 frames per second. In total, 4,478 videos are collected from 165 subjects including variations in spoof type, recording device, illumination condition, pose and facial expression. To define train and test sets, we start by following the splitting system described as Protocol 1 in [29]. We further subsample train and test sets by extracting 1 every 10 frames (or every 0.33 seconds) in each video, since consecutive frames are almost identical. Finally, since we observe that simple models can reach very high accuracy, we further make the task harder by creating 2 separate training and evaluation folds in a way that the model is trained and tested on different spoof types. We provide additional details in the supplementary material.

To convert the original SiW into **SiW-EnrollN**, we consider the same enrollment set size N as for CASp-EnrollN. Since the original dataset is collected using two different capturing devices, we can use this information for the definition of enrollment sets. Indeed, in a real scenario, the subject would have to enroll twice when using two different devices, meaning that a unique enrollment set will be generated for each combination of subject and capturing devices. Capturing the sensor bias in the enrollment can be particularly beneficial, for example to detect if the resolution of a face changes, which can indicate a replay attack. To extract enrollment data we choose a single video from each subject and sensor among the available ones. We arbitrarily pick the video without illumination changes and with variation in subject pose, as it resembles the enrollment conditions required in real devices. We then equidistantly sample N frames over the video to construct the enrollment set. As shown in Fig. 2, this allows capturing different poses and facial expressions from the subject. Finally, we exclude all the frames in this video from training and test data to avoid information leakage and match all the remaining queries from the same subject and sensor to their unique enrollment set.

Table 1. Data statistics for the personalized benchmarks CASp-Enroll5 and SiW-Enroll5.

	CASp-Enroll5		SiW-Enroll5	
	Train	Test	Train	Test
# data points	427,696	59,481	13,314	107,835
# subjects	7,500	920	90	75

In Table 1 we report a summary of the dataset population for CASp-Enroll5 and SiW-Enroll5. We plan to release soon the code to generate personalized datasets from CelebA-Spoof and SiW for any arbitrary number of enrollment images per set.

4. Methods

In this section, we introduce a few baseline methods to aggregate the enrollment features, combine them with the query features and use them for classification. By using simple aggregation modules in our baselines, we show how existing anti-spoofing models can be easily adapted for personalization, without the need for complicated backbones or training algorithms. We leave as future work the development of optimized modules for personalization.

In Fig. 3 we show a high-level diagram of our personalized anti-spoofing method using enrollment images. The model takes as input a query image $I_q \in \mathbb{R}^{C \times H \times W}$ and N enrollment images $I_e^i \in \mathbb{R}^{C \times H \times W}$ with $i \in \{1, 2, ...N\}$. For generality we consider two independent CNN-based feature extractors $\phi_q(\cdot)$ and $\phi_e(\cdot)$ which encode the query and enrollment images into latent features, respectively $f_q = \phi_q(I_q)$ and $f_e^i = \phi_e(I_e^i)$ with $f_q \in \mathbb{R}^D$ and $f_e^i \in \mathbb{R}^D$. After the extraction of latent features, an aggregation layer is used to combine the enrollment features $f_e^1, f_e^2, ...f_e^N$ into a single feature f_e^{agg}. Different ways to implement the aggregation layer using parametric and non-parametric methods are described in the rest of this section. Finally, the query features and aggregated enrollment features are concatenated and passed as input to an MLP classifier to obtain the multi-class predictions: $y_q = MLP(f_q, f_e^{agg})$. We trained the model end-to-end by minimizing the cross-entropy between neural network predictions and ground-truth labels. During inference time, the query images along with the enrollment images are fed into the network in the same way as in training to output the class probability of the query image.

In this paper, we apply our personalization approach on top of different backbones using RGB inputs, but it is important to notice how the proposed method can be also applied on top of any other face anti-spoofing model, even with multi-modal input. This makes personalization a viable candidate to improve the performance of a wide variety of anti-spoofing systems.

Since the encoding of enrollment images is not dependent on the query, this operation can be executed once during the user enrollment procedure. This addresses both privacy and efficiency concerns since only latent features are required for the personalization and their extraction only happens once.

Personalization methods based on fine-tuning using enrollment data might be possible. However, fine-tuning a model with as few as 5 enrollment images is a challenging task, while the proposed solution based on model conditioning does not require any additional dependency once the anti-spoofing model has been trained.

Enrollment images could also be used as additional training samples. We argue against this approach as it yields a minor increase in training data at the cost of requiring enrollment and query images to be of the same quality. This is often not true in real systems, where enrollments sets are commonly stored as compressed templates instead of raw features [6, 44] to preserve privacy.

4.1. Concatenation and Mean

We evaluate two non-parametric operators to aggregate enrollment features: vector concatenation and arithmetic mean.

In the first case, we simply concatenate the enrollment features along the dimension axis to obtain a 1-dimensional vector with $N * D$ values. This method retains all information from the original features but is not invariant to the features' ordering.

For the second method, the enrollment features are aggregated through the non-parametric operation of the arithmetic mean in the latent space. This is similar to [42], where the authors take the arithmetic mean of the support samples and compute its distance to the query features. Mathematically, we obtain the aggregated feature vector f_e^{agg} as $f_e^{agg} = \frac{1}{N} \sum_{i=1}^{N} f_e^i$. In this case, the enrollment features are compacted into only D values. Differently from [42] we use a multi-layer perceptron to model the relation between enrollment and query samples instead of relying on Euclidean distance.

4.2. Gated Recurrent Unit

Depending on how enrollment sets are defined, they can be represented as sequential data. For example, in SiW, the enrollment images are sampled from the frames of a short video, which always starts with the subject's face in a frontal, neutral pose and then continues with changes in perspective or facial expression. We propose the usage of Gated Recurrent Unit (GRU) [9] for the aggregation of enrollment features, as it is can easily model relations in sequential data. At each time step in the sequence, GRU takes the input signal at the same time step and the activation from the previous time step to output the activation for the current

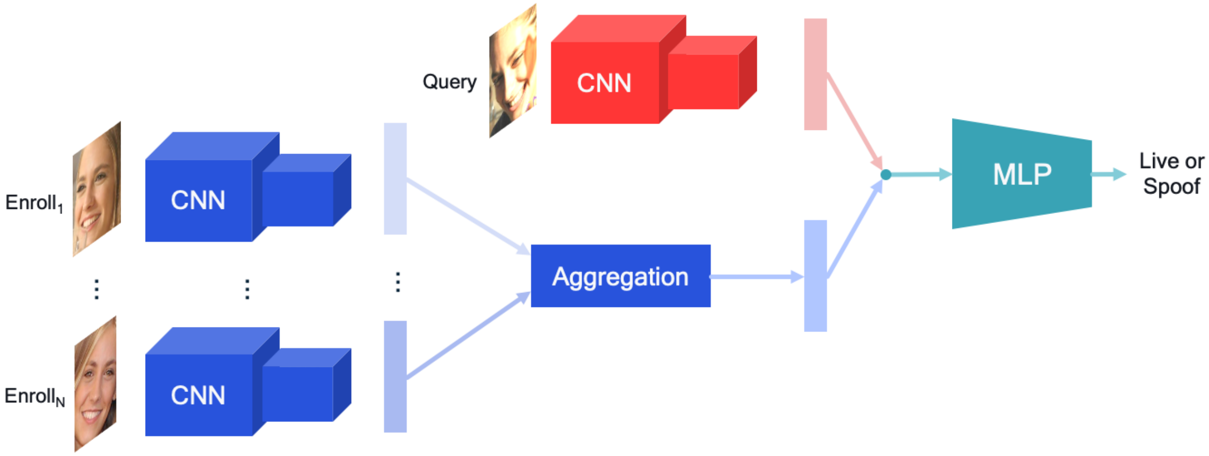

Figure 3. High-level schema of the proposed framework for face anti-spoofing personalization. Compact features are extracted from query and enrollment images using different CNNs. The enrollment features from multiple images are summarized into a single latent vector. Query and enrollment features are finally concatenated and input to a classifier to predict the query's label.

time step. In our setting, each enrollment feature represents the input signal at a particular time step. Formally, inputs and outputs of the GRU module are:

$$h_l^i = GRU(f_l^i, h_l^{i-1}). \tag{3}$$

where, f_l^i is the i^{th} latent feature at layer l and h_l^{i-1} and h_l^i are previous and current activations for the same layer. The input to the first layer are the enrollment features $f_0^i = f_e^i$. The last activation of the final GRU layer is selected as the aggregated feature for the enrollment set: $f_e^{agg} = f_L^N$.

4.3. Attention

To obtain a parametric aggregation which is instead invariant to feature ordering, we propose using the key-query-value attention mechanism [45] between query and enrollment features to compute the aggregated features. The attention layer allows the model to learn the importance of each enrollment image given a specific query image. We obtain the attention query from the query image as $Q = A_Q(f_q)$ and the attention keys and values from the enrollment images as $K_i = A_K(f_e^i)$ and $V_i = A_V(f_e^i)$. Here A_Q, A_K and A_V are linear layers that map from a D-dimensional feature space to an M-dimensional feature space. The attention weights obtained from $Q \in \mathbb{R}^{1\times M}$ and $K \in \mathbb{R}^{N\times M}$ are then applied to the value vectors $V \in \mathbb{R}^{N\times M}$ to obtain the aggregated feature f_e^{agg} as described in:

$$f_e^{agg} = \text{Softmax}(\frac{QK^T}{\sqrt{M}})V \tag{4}$$

4.4. Graph Neural Network

We finally consider a deeper aggregation method based on GNNs which can learn to model more complex relations between enrollment and query features. The proposed GNN is similar to attention in that it learns to compare query and enrollment samples but in addition, it can as well perform message passing across enrollment features. For the implementation of the GNN, we follow the architecture described in [37], and accordingly replace the anti-spoofing classifier with predictions obtained directly from the GNN. In this formulation, the enrollment and query features are represented as nodes of the graph. Each GNN layer consists of two alternating steps: taking the node features to compute multiple adjacency matrices and then applying them for the graph convolution operations. The elements of each adjacency matrix are obtained using a distance function $\psi_l(\cdot)$ between two node features f_l^i, f_l^j such that $A_l^{ij} = \psi_l(f_l^i, f_l^j)$. A neural network is used to parametrize the function ψ_l. The adjacency matrices are used in the graph convolution operation as follows:

$$f_{l+1}^i = GConv(f_l^i) = \rho\Big(\sum_{A_l \in \mathcal{A}_l} A_l f_l^i W_l\Big). \tag{5}$$

Here $A_l \in \mathbb{R}^{(N+1)\times(N+1)}$ is the learned adjacency matrix from the set of adjacency matrices $\mathcal{A}_l$, $f_l^i \in \mathbb{R}^{(N+1)\times d_l}$ is the feature matrix of the l^{th} GNN layer. The feature matrix consists of N enrollment features and 1 query feature of dimension d_l. $W_l \in \mathbb{R}^{d_l \times d_{l+1}}$ is the mapping matrix associated with layer l that maps from d_l to d_{l+1} dimensional feature space and ρ is a non-linearity. The inputs to the first layer of the GNN are $N+1$ nodes consisting of N enrollment features $f_e^1, f_e^2, ...f_e^N$ and the query feature f_q; while the output features for the query node are used as final predictions. Fig. 4 shows an example of GNN for $N = 3$.

Table 2. Comparison between baseline and personalized model for the backbones VGG16, ResNet18 and FeatherNet on CASp-Enroll5 and SiW-Enroll5. For AUC and AUC10 higher is better, for EER lower is better.

	CASp-Enroll5									SiW-Enroll5								
	VGG16			ResNet18			FeatherNet			VGG16			ResNet18			FeatherNet		
Method	AUC	AUC10	EER	AUC	AUC10	EER	AUC	AUC10	EER	AUC	AUC10	EER	AUC	AUC10	EER	AUC	AUC10	EER
Baseline	98.0	92.4	7.2	98.3	94.1	5.9	97.1	89.3	8.8	97.8	92.0	6.8	99.1	96.7	4.3	98.9	95.8	4.8
Personalized	**98.6**	**94.1**	**5.9**	**99.2**	**96.4**	**4.3**	**97.8**	**91.5**	**7.5**	**98.1**	**93.3**	**6.2**	**99.2**	**97.0**	**3.9**	**99.0**	**96.2**	**4.6**

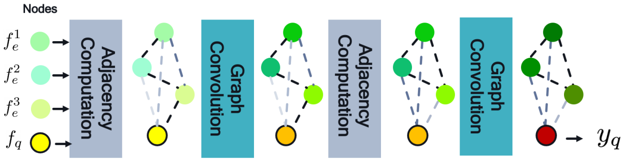

Figure 4. The GNN aggregates neighborhood information from enrollment features to infer the class of the query feature. In each layer, the network first computes the adjacency matrix and then uses it to convolve over the node features.

5. Experiments

To evaluate the importance of enrollment data for the anti-spoofing task, we use VGG16 [41], ResNet18 [17] and FeatherNet [51] architectures as backbones and evaluate various aggregation methods on the two proposed benchmarks. While the first two networks are general-purpose backbones for images, FeatherNet has been recently proposed specifically for the task of face anti-spoofing. We also argue that it is straightforward to apply our personalization techniques on top of other backbones and anti-spoofing models. We leave for future work the optimization of the backbone and personalization module for face anti-spoofing. We provide additional details on implementation, training procedure and model hyper-parameters in the supplementary material.

5.1. Evaluation Metrics

We evaluate our methods using the following three metrics: (a) Area Under Curve (AUC), (b) Area Under Curve till False Negative Rate of 10% (AUC10) and (c) Equal Error Rate (EER). These metrics were chosen as they are agnostic to threshold choices unlike False Negative Rate and False Positive Rate which depend on a specific operating point. AUC is the area under the Receiving Operating Characteristics (ROC) of True Positive Rate versus False Positive Rate as the classification threshold is varied. For anti-spoofing, the spoof class is chosen as the positive label. AUC10 is the area under the same ROC until the False Negative Rate of 10% operating point. The cut-off at 10% allows for evaluation of the anti-spoofing performance in the regimes where user experience is not significantly penalized (low False Negative Rate), which is required for many practical applications. Finally, EER is the point on the ROC where the False Negative Rate and the False Positive Rate are equal. Results are reported over 5 seeds.

We include in the supplementary material additional results and a statistical hypothesis testing study to confirm the significance of our results. Notice also that the absolute scores on CASp-Enroll and SiW-Enroll should not be compared to results on CelebA-Spoof and SiW datasets presented in previous papers, as the conversion to personalized benchmarks inevitably changes training and test sets.

5.2. Results and Ablation Studies

We conduct experiments to investigate: (a) whether personalization can improve anti-spoofing performance over different datasets and backbones, (b) which of the proposed aggregation methods is best to combine enrollment features, (c) the impact of using larger enrollment sets, (d) the effect of using weight-shared feature extractors for query and enrollment images.

5.2.1 Personalization of anti-spoofing models

We firstly investigate whether using enrollment data improves the anti-spoofing capabilities of neural networks. Specifically, we compare two methods: (a) a baseline, which does not use any enrollment and (b) a personalized approach, which uses enrollment sets of size 5 and 'mean' aggregation to condition the classifier. The experiments are carried out on both SiW-Enroll5 and CASp-Enroll5 datasets across three feature extracting backbones: (i) VGG16, (ii) ResNet18 and (iii) FeatherNet. We report the results in Table 2, comparing baseline and personalized methods for all combinations of dataset and backbones. We observe a consistent improvement using personalized methods, with the largest AUC10 gaps reaching 2.2% for CASp-Enroll5, on top of already strong baselines achieving 90% AUC10 or more. The limited improvement on SiW-Enroll5 could be caused by the smaller variety in enrollment images compared to the CASp-Enroll5 dataset. A significance test on these results is reported in the supplementary material.

5.2.2 Aggregation methods

In this experiment, we study the effect of using different methods for aggregation of enrollment features, as described in Section 4. For the evaluation, we use the VGG16 backbone and enrollment set size of 5. The results are shown in Table 3. Overall mean, GRU and GNN methods consistently outperform the baseline. Surprisingly, aggregation through the mean operator is the best or second-best performer in all cases. This suggests that more expressive methods to aggregate enrollment features (e.g.: through GNN, attention or GRU) are not necessary for these datasets and that the model might be learning to extract some shared information (like identity) that is not distorted through the mean operator. While concatenation works well on SiW-Enroll5, the low results on CASp-Enroll5 might be induced by the larger variations in image quality in the latter dataset. Lastly, the attention-based aggregation method performs poorer than the baseline for both datasets, implying that different information is captured in query and enrollment features, which should be aggregated independently. Due to the simplicity and consistent performance of the mean operator, we choose it as the aggregation method for the rest of the experiments in this paper.

Table 3. Comparison between baseline and different aggregation methods with VGG16 architecture on CASp-Enroll5 and SiW-Enroll5. Top performance are highlighted as: **First**, Second.

	CASp-Enroll5			SiW-Enroll5		
Method	AUC	AUC10	EER	AUC	AUC10	EER
Baseline	98.0	92.4	7.2	97.8	92.0	6.8
Concatenation	97.7	91.3	7.5	**98.2**	**94.3**	**6.0**
Mean	**98.6**	**94.1**	**5.9**	98.1	93.3	6.2
Attention	97.9	91.7	7.3	97.6	92.0	7.1
GRU	98.4	93.0	6.2	97.8	92.7	6.8
GNN	98.1	92.8	6.9	97.9	93.4	6.5

5.2.3 Number of enrollment images

With these experiments, we study the effect of changing the number of enrollment images input to the network. The results are shown for both AUC10 and EER metrics in Fig. 5 using the VGG16 backbone with mean aggregation method on the CASp-Enroll8 and SiW-Enroll8 datasets. These versions of the personalized datasets are generated in the same way as the ones with $N = 5$ enrollment images, but using $N = 8$ to allow for more variation in this ablation study. Out of the 8 enrollment samples available per user, we select 1, 2, 4 and 8 of them as input to the network for this study. Choosing 0 enrollment images implies that we use the baseline model without any personalization. The results show that any number of enrollment images provides a significant boost in KPIs over the non-personalized method. Moreover, the best scores are obtained for 2 and 4 enrollment images. We hypothesize that having more than one enrollment image might help to obtain a robust representation of the enrollment set, improving in cases where the first image is noisy or not informative of the subject.

5.2.4 Weight-shared feature extractors

In this study, we investigate the kind of information encoded in enrollment features. We originally hypothesized that query and enrollment features encode comparable information and that the MLP learns to compare the two sources. To this end, we try an ablation study in which feature extractors for query and enrollment have shared weights.

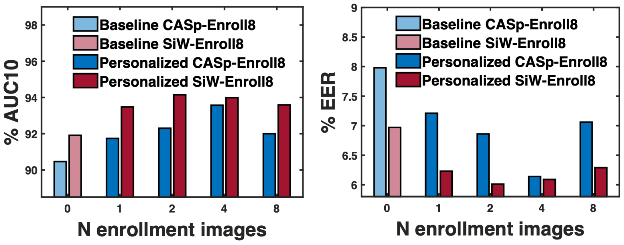

Figure 5. Effect of varying enrollment set size on AUC10 (left) and EER (right) for CASp-Enroll8 and SiW-Enroll8 datasets.

Table 4. Effect of using shared and separated weights with VGG16 backbone on CASp-Enroll5 and SiW-Enroll5 datasets.

	CASp-Enroll5			SiW-Enroll5		
Method	AUC	AUC10	EER	AUC	AUC10	EER
Baseline	98.0	92.4	7.2	97.8	92.0	6.8
Shared	97.7	90.8	13.7	96.1	87.3	20.6
Separated	**98.6**	**94.1**	**5.9**	**98.1**	**93.3**	**6.2**

Table 4 results show that feature extraction with shared weights is inferior to having independent CNN weights. This suggests that the information extracted from query and enrollment images is of different types. In fact, performance using shared weights is even poorer than the one of the non-personalized baseline. This might be due to a conflict in the network's optimization as the encoder tries to learn to represent different information for enrollment and query images.

6. Conclusion

We introduced a new personalized benchmark for face anti-spoofing. We developed a method to personalize existing face anti-spoofing datasets by holding out an enrollment set for each user that can be associated with each query sample. We demonstrated this by converting two recent face anti-spoofing datasets into their personalized versions. The proposed conversion method can be applied to any other dataset. Furthermore, we introduced a simple but effective personalized baseline by conditioning the anti-spoofing prediction on the enrollment set. We then proposed a suite of modules to aggregate information across multiple enrollment images for the conditioning.

Results confirmed that using enrollment samples generally improves performance over the non-personalized baseline for both datasets and multiple backbones. We further analyzed how different aggregation methods, sizes of enrollment sets and extraction strategies affect personalization performance. With our proposed benchmark we aim to incentivize the research community to develop new personalized face anti-spoofing methods exploiting the availability of enrollment data in real-world systems.

References

[1] Waldir R Almeida, Fernanda A Andaló, Rafael Padilha, Gabriel Bertocco, William Dias, Ricardo da S Torres, Jacques Wainer, and Anderson Rocha. Detecting face presentation attacks in mobile devices with a patch-based cnn and a sensor-aware loss function. *PloS one*, 15(9):e0238058, 2020.

[2] Sushil Bhattacharjee, Amir Mohammadi, and Sébastien Marcel. Spoofing deep face recognition with custom silicone masks. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7. IEEE, 2018.

[3] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face antispoofing using speeded-up robust features and fisher vector encoding. *IEEE Signal Processing Letters*, 24(2):141–145, 2016.

[4] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security*, 11(8):1818–1830, 2016.

[5] 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.

[6] Michael Braithwaite, U Cahn von Seelen, James Cambier, John Daugman, Randy Glass, Russ Moore, and Ian Scott. Application-specific biometric templates. In *IEEE Workshop on Automatic Identification Advanced Technologies, Tarrytown, NY, March*, pages 14–15. Citeseer, 2002.

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

[8] Ivana Chingovska, Nesli Erdogmus, André Anjos, and Sébastien Marcel. Face recognition systems under spoofing attacks. In *Face Recognition Across the Imaging Spectrum*, pages 165–194. Springer, 2016.

[9] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder–decoder approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111, 2014.

[10] Nesli Erdogmus and Sébastien Marcel. Spoofing 2d face recognition systems with 3d masks. In *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG)*, pages 1–8. IEEE, 2013.

[11] Soroush Fatemifar, Shervin Rahimzadeh Arashloo, Muhammad Awais, and Josef Kittler. Client-specific anomaly detection for face presentation attack detection. *Pattern Recognition*, 112:107696, 2021.

[12] Soroush Fatemifar, Muhammad Awais, Ali Akbari, and Josef Kittler. A stacking ensemble for anomaly based client-specific face spoofing detection. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1371–1375. IEEE, 2020.

[13] Soroush Fatemifar, Muhammad Awais, Shervin Rahimzadeh Arashloo, and Josef Kittler. Combining multiple one-class classifiers for anomaly based face spoofing attack detection. In *2019 International Conference on Biometrics (ICB)*, pages 1–7. IEEE, 2019.

[14] 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.

[15] Litong Feng, Lai-Man Po, Yuming Li, Xuyuan Xu, Fang Yuan, Terence Chun-Ho Cheung, and Kwok-Wai Cheung. Integration of image quality and motion cues for face antispoofing: A neural network approach. *Journal of Visual Communication and Image Representation*, 38:451–460, 2016.

[16] Alejandro Gomez-Alanis, Jose A Gonzalez-Lopez, S Pavankumar Dubagunta, Antonio M Peinado, and Mathew Magimai Doss. On joint optimization of automatic speaker verification and anti-spoofing in the embedding space. *IEEE Transactions on Information Forensics and Security*, 16:1579–1593, 2020.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[18] Javier Hernandez-Ortega, Javier Galbally, Julian Fierrez, Rudolf Haraksim, and Laurent Beslay. Faceqnet: Quality assessment for face recognition based on deep learning. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019.

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

[20] Jangho Kim, Simyung Chang, Sungrack Yun, and Nojun Kwak. Prototype-based personalized pruning. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3925–3929. IEEE, 2021.

[21] 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.

[22] Klaus Kollreider, Hartwig Fronthaler, Maycel Isaac Faraj, and Josef Bigun. Real-time face detection and motion analysis with application in "liveness" assessment. *IEEE Transactions on Information Forensics and Security*, 2(3):548–558, 2007.

[23] Yassin Kortli, Maher Jridi, Ayman Al Falou, and Mohamed Atri. Face recognition systems: A survey. *Sensors*, 20(2):342, 2020.

[24] Jiakang Li, Meng Sun, Xiongwei Zhang, and Yimin Wang. Joint decision of anti-spoofing and automatic speaker verification by multi-task learning with contrastive loss. *IEEE Access*, 8:7907–7915, 2020.

[25] Jiangwei Li, Yunhong Wang, Tieniu Tan, and Anil K Jain. Live face detection based on the analysis of fourier spec-

tra. In *Biometric technology for human identification*, volume 5404, pages 296–303. International Society for Optics and Photonics, 2004.

[26] Lei Li, Xiaoyi Feng, Zinelabidine Boulkenafet, Zhaoqiang Xia, Mingming Li, and Abdenour Hadid. An original face anti-spoofing approach using partial convolutional neural network. In *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2016.

[27] 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.

[28] 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.

[29] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 389–398, 2018.

[30] Yaojie Liu, Joel Stehouwer, Amin Jourabloo, and Xiaoming Liu. Deep tree learning for zero-shot face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4680–4689, 2019.

[31] Jukka Määttä, Abdenour Hadid, and Matti Pietikäinen. Face spoofing detection from single images using texture and local shape analysis. *IET biometrics*, 1(1):3–10, 2012.

[32] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[33] Dailé Osorio-Roig, Christian Rathgeb, Pawel Drozdowski, and Christoph Busch. Stable hash generation for efficient privacy-preserving face identification. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2021.

[34] Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *2007 IEEE 11th international conference on computer vision*, pages 1–8. IEEE, 2007.

[35] Keyurkumar Patel, Hu Han, and Anil K Jain. Secure face unlock: Spoof detection on smartphones. *IEEE transactions on information forensics and security*, 11(10):2268–2283, 2016.

[36] Sandip Purnapatra, Nic Smalt, Keivan Bahmani, Priyanka Das, David Yambay, Amir Mohammadi, Anjith George, Thirimachos Bourlai, Sébastien Marcel, Stephanie Schuckers, et al. Face liveness detection competition (livdet-face)-2021. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2021.

[37] Victor Garcia Satorras and Joan Bruna Estrach. Few-shot learning with graph neural networks. In *International Conference on Learning Representations*, 2018.

[38] William Robson Schwartz, Anderson Rocha, and Helio Pedrini. Face spoofing detection through partial least squares and low-level descriptors. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2011.

[39] 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.

[40] Tao Shen, Yuyu Huang, and Zhijun Tong. Facebagnet: Bag-of-local-features model for multi-modal face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[41] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

[42] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*, 30:4077–4087, 2017.

[43] Lin Sun, Gang Pan, Zhaohui Wu, and Shihong Lao. Blinking-based live face detection using conditional random fields. In *International Conference on Biometrics*, pages 252–260. Springer, 2007.

[44] Yagiz Sutcu, Qiming Li, and Nasir Memon. Secure biometric templates from fingerprint-face features. In *2007 IEEE Conference on computer vision and pattern recognition*, pages 1–6. IEEE, 2007.

[45] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[46] 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.

[47] Jianwei Yang, Zhen Lei, Shengcai Liao, and Stan Z Li. Face liveness detection with component dependent descriptor. In *2013 International Conference on Biometrics (ICB)*, pages 1–6. IEEE, 2013.

[48] Jianwei Yang, Zhen Lei, Dong Yi, and Stan Z Li. Person-specific face antispoofing with subject domain adaptation. *IEEE Transactions on Information Forensics and Security*, 10(4):797–809, 2015.

[49] Xiao Yang, Wenhan Luo, Linchao Bao, Yuan Gao, Dihong Gong, Shibao Zheng, Zhifeng Li, and Wei Liu. Face anti-spoofing: Model matters, so does data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3507–3516, 2019.

[50] 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.

[51] Peng Zhang, Fuhao Zou, Zhiwen Wu, Nengli Dai, Skarpness Mark, Michael Fu, Juan Zhao, and Kai Li. Feathernets: convolutional neural networks as light as feather for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[52] Shifeng Zhang, Xiaobo Wang, Ajian Liu, Chenxu Zhao, Jun Wan, Sergio Escalera, Hailin Shi, Zezheng Wang, and Stan Z

Li. A dataset and benchmark for large-scale multi-modal face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 919–928, 2019.

[53] 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.

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

[55] Junwei Zhou, Ke Shu, Dongdong Zhao, and Zhe Xia. Domain adaptation based person-specific face anti-spoofing using color texture features. In *Proceedings of the 2020 5th International Conference on Machine Learning Technologies*, pages 79–85, 2020.