

SynFacePAD 2023: Competition on Face Presentation Attack Detection Based on Privacy-aware Synthetic Training Data

Meiling Fang^{1,2,*}, Marco Huber^{1,2,*}, Julian Fierrez^{3,*}, Raghavendra Ramachandra^{4,*}, Naser Damer^{1,2,*}, Alhasan Alkhaddour^{5,+}, Maksim Kasantcev^{5,+}, Vasiliy Pryadchenko^{5,+}, Ziyuan Yang^{6,+}, Huijie Huangfu^{6,+}, Yingyu Chen^{6,+}, Yi Zhang^{6,+}, Yuchen Pan^{7,+}, Junjun Jiang^{7,+}, Xianming Liu^{7,+}, Xianyun Sun^{8,+}, Caiyong Wang^{8,+}, Xingyu Liu^{8,+}, Zhaohua Chang^{8,+}, Guangzhe Zhao^{8,+}, Juan Tapia^{9,10,+}, Lazaro Gonzalez-Soler^{9,+}, Carlos Aravena^{10,+}, Daniel Schulz^{10,+}

¹Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany

²Department of Computer Science, TU Darmstadt, Darmstadt, Germany

³Biometrics and Data Pattern Analytics Lab, Universidad Autonoma de Madrid, Spain

⁴Norwegian University of Science and Technology (NTNU), Norway

⁵ID R&D, Inc., New York, US

⁶School of Cyber Science and Engineering, Sichuan University, Chengdu, China

⁷School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

⁸School of Electrical and Information Engineering,

Beijing University of Civil Engineering and Architecture, China

⁹Biometrics and Security Research Group, Hochschule Darmstadt, Darmstadt, Germany

¹⁰I+D Vision Center, Santiago, Chile

*Competition Organizers. +Competition participant.

Email: meiling.fang@igd.fraunhofer.de

Abstract

This paper presents a summary of the Competition on Face Presentation Attack Detection Based on Privacy-aware Synthetic Training Data (SynFacePAD 2023) held at the 2023 International Joint Conference on Biometrics (IJCBI 2023). The competition attracted a total of 8 participating teams with valid submissions from academia and industry. The competition aimed to motivate and attract solutions that target detecting face presentation attacks while considering synthetic-based training data motivated by privacy, legal and ethical concerns associated with personal data. To achieve that, the training data used by the participants was limited to synthetic data provided by the organizers. The submitted solutions presented innovations and novel approaches that led to outperforming the considered baseline in the investigated benchmarks.

1. Introduction

Face recognition has been widely deployed in various application scenarios, such as access control, phone unlocking, and mobile payment. Reasons for this include its convenience and outstanding performance [15, 2, 3]. However, face recognition is susceptible to Presentation Attacks (PAs), such as high-resolution photos and videos of an authorized user [56, 10, 1, 20, 24]. Therefore, face Presenta-

tion Attack Detection (PAD) technology [32], which describes the process of identifying whether a face presented to the system is a bona fide (live) or PA, plays an important role to secure recognition from PAs [28]. These PA detectors are often built using authentic biometric data [22], raising ethical and legal challenges. Such challenges have recently been discussed in face recognition [7, 47], face morphing attack detection [13, 51], and face PAD [21]. There was previously a series of competitions on face PAD based on authentic data [46] and a competition targeting face morphing attack detection based on privacy synthetic training data [34]. However, this is the first competition targeting PAs on face recognition while limiting its development data to synthetic data. Given the legal privacy regulations, the collection, use, share, and maintenance of face data for biometric processing is extremely challenging [11]. For example, several large-scale face recognition datasets [9, 27, 38] were withdrawn by their creators with privacy and proper subjects consent issues being the main reason. One of the main solutions for this issue is the use of synthetic data [11]. This has been very recently and successfully proposed for the training of face recognition [47, 5, 6] and morphing attack detection [13, 34, 18, 12], among other processes such as model quantization [4, 40]. Furthermore, a recent work followed this motivation to take advantage of synthetic data

to develop PADs in a privacy-friendly manner [21] and proved the usability of synthetic data for the development of face PADs. The utilized assumption is that learning to detect the differences between bona fide and attack samples of a synthetic origin can be used to detect these differences between authentic bona fide and attacks and thus train PAD without authentic private data.

Driven by the need for the development of face PAD datasets that prioritize the privacy of individuals, promote data sharing within the research community, and ensure the reproducibility and continuity of face PAD research, we conduct the SynFacePAD 2023: Competition on Face Presentation Attack Detection Based on Privacy-aware Synthetic Training Data at the International Joint Conference on Biometrics 2023. The results and observations are summarized in this paper.

2. Dataset, Evaluation Criteria, and Participants

Dataset	Year	# Bona fide/attack	# Sub	Attack types
SynthASpoof [21]	2023	25,000 / 78,800 (I&V)	25,000	1 Print, 3 Replay
CASIA-FASD [56]	2012	150 / 450 (V)	50	1 Print, 1 Replay
Replay-Attack [10]	2012	200 / 1,000 (V)	50	1 Print, 2 Replay
MSU-MFSD [52]	2015	70 / 210 (V)	35	1 Print, 2 Replay
OULU-NPU [1]	2017	1,980 / 3,960 (V)	55	2 Print, 2 Replay

Table 1. Summary of the used face PAD datasets. V and I are shorthand for video and image, respectively. SynthASpoof is a synthetic face PAD, serving as training dataset, and the other four are public available authentic face PAD evaluation benchmarks.

2.1. Dataset

Training Dataset: To promote the development of face PAD on synthetic data, this competition restricts the training data to the provided privacy-friendly synthetic dataset, **SynthASpoof** [21]. The SynthASpoof consists of 25,000 bona fide and 78,800 attack samples and is publicly available. The bona fide samples were generated using StyleGAN2-ADA [36] and the attack samples were collected by presenting these synthetic samples as printed or replayed attacks to three different capture sensors. To ensure that the participants trained their solutions solely by using the training data set provided, the solutions were trained again by the organizers.

Evaluation Benchmarks: For the evaluation, we use four authentic face PAD benchmarks: MSU-MFSD [52] (denoted as M), CASIA-MFSD [56] (denoted as C), Idiap Replay-Attack [10] (denoted as I), and OULU-NPU [1] (denoted as O). We select these four datasets by considering their widely used in generalized face PAD studies [17, 19, 42, 49, 50]. The **MSU-MFSD (M)** [52] dataset comprises 440 videos captured from 35 subjects, utilizing two different resolutions of cameras. The dataset includes two types of attacks: printed photo attacks and replay attacks. The **CASIA-MFSD (C)** [56] dataset consists of 600 videos from 50 subjects and includes three types of attacks:

warped photo attack, cut photo attack, and video replay attack. The **Idiap Replay-Attack (I)** [10] dataset contains 300 videos from 50 subjects captured under different sensors and illumination conditions. The dataset includes two types of attacks: print attacks and replay attacks. The **Oulu-NPU (O)** [1] is a mobile face PAD dataset collected in a realistic mobile scenario. It consists of 5940 video clips from 55 subjects using six different mobile phones.

Samples from the provided training dataset SynthASpoof, as well as four evaluation benchmarks, are shown in Figure 1, and the corresponding information is summarized in Table 1. In addition, we provide participants with a pre-processing implementation that includes face detection and cropping¹. For the evaluation benchmark, the faces were detected and cropped using the MTCNN method. Notably, there are no restrictions on the pre-processing of the training data.

2.2. Baseline Methods

The baseline performance is based on two face PA detectors reported in [21], ResNet and PixBis. ResNet is one of the most popular backbone architectures used in face PAD algorithm design [54, 19, 55, 17]. PixBis [23] employs a binary supervisory strategy at pixel-level to simplify the problem and obviate the need for a computationally intensive synthesis of depth maps.

2.3. Evaluation Criteria

The SynFacePAD competition uses two PAD metrics to evaluate the submitted solutions following the ISO/IEC 30107-3 [35] standard: Bona fide Presentation Classification Error Rate (BPCER) and Attack Presentation Classification Error Rate (APCER). BPCER refers to the proportion of bona fide presentations classified as attack samples, while APCER is the proportion of attack presentations incorrectly classified as bona fide presentation. The submitted solutions are evaluated at two different fixed APCER (and BPCER) values, 10% and 20%, and the corresponding BPCER (and APCER) is reported. To cover diverse operational points and enable a detailed results discussion, we provide a visual evaluation by plotting Receiver Operating Characteristic (ROC) curves, where the x-axis of the ROC is APCER and the y-axis is 1-BPCER. Furthermore, following existing cross-domain face PAD methods [21, 42, 49, 50], the Half Total Error Rate (HTER), which is the mean of BPCER and APCER [35] and Area under the ROC Curve (AUC) value is also reported. The HTER threshold is computed based on the Equal Error Rate (EER) threshold from the targeted evaluating benchmark directly.

The ranking of the submitted solutions on each benchmark is determined by the APCER at a fixed BPCER of 20%, allowing us to analyze the detectability of different

¹https://github.com/meilfang/SynthASpoof/tree/main/data_preprocess

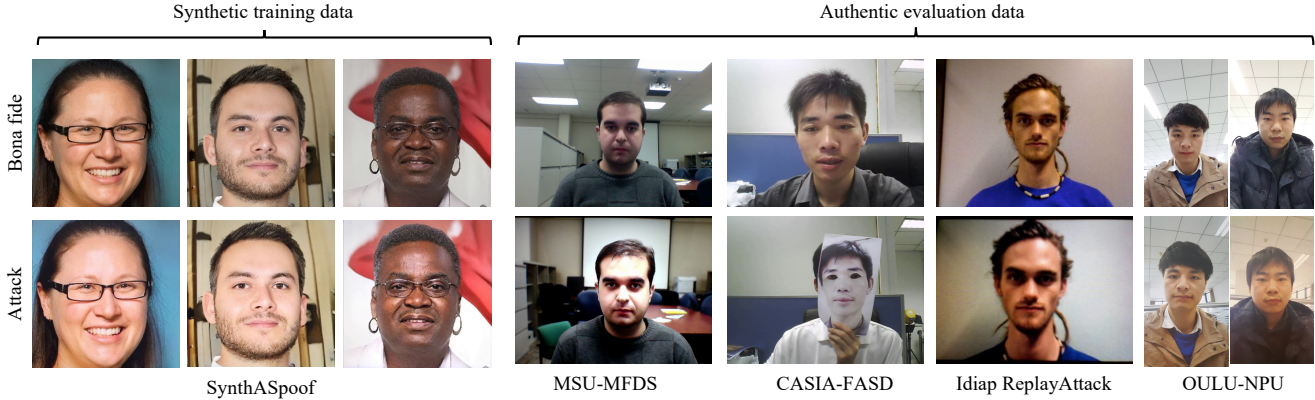


Figure 1. Samples of the synthetic training data from SynthASpoof [21], as well as four authentic evaluation face PAD benchmarks.

Team	Team members	Affiliations	Type	Solution
ID R&D Inc	Alhasan Alkhaddour, Maksim Kasantcev, Vasiliy Pryadchenko	ID R&D, Inc., New York, US	Industry	ViT-SIDE B
SCU-DIG	Ziyuan Yang, Huijie Huangfu, Yingyu Chen, Yi Zhang	School of Cyber Science and Engineering, Sichuan University, Chengdu, China	Academic	SynFace Co-Former B SynFace Co-Former A
HIT	Yuchen Pan, Junjun Jiang, Xianming Liu	School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China	Academic	CoDe-Lc CoDe-Lh
BUCEA	Xianyun Sun, Caiyong Wang, Xingyu Liu, Zhaohua Chang, Guangzhe Zhao	School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, China	Academic	OrthPADNet
hda	Juan Tapia, Lazaro J. Gonzalez-Soler	Biometrics and Security Research Group, Hochschule Darmstadt, Darmstadt, Germany	Academic	hdaFVPAD
idvc	Juan Tapia, Carlos Aravena, Daniel Schulz	I+D Vision Center, Santiago, Chile	Industry	idvcVT
Anonymous-1	-	-	Industry	Saliency-ResNet-CAS Saliency-ResNet-ES
Anonymous-2	-	-	Academic	-

Table 2. A summary of the valid participating teams, team members, affiliations, type of institution, and solutions. More details on the submitted algorithms is provided in Section 3.

solutions on the attack samples. Once the solutions are ranked, the final ranking of the team is based on their best-performing solution if a team submits two solutions.

2.4. Competition Participants

The goal of the SynFacePAD competition is to attract participants from both academia and industry, with a wide geographic and activity variation. The call for participation was shared on the website of the International Joint Conference on Biometrics (IJCB) 2023, the competition’s own website, and various social media platforms. As a result, 14 registered teams, both from academia and industry registered for the competition. Among them, eight teams submitted a total of 11 valid solutions, with each team allowed to submit up to two solutions. These eight teams have affiliations in five different countries, consisting of five teams with academic affiliations and three teams with industry affiliations. Two teams opted to be anonymous. Table 2 provides a summary of the participating teams.

2.5. Submission and Evaluation Process

Each team participating in the SynFacePAD competition registered with a team name and a list of team members with their affiliations for the competition and was then provided access to the synthetic data. The training data was restricted to the use of the synthetic data provided by the organizers which consisted of the SynthASpoof dataset [21]. Only this

dataset was allowed to be used during the training of the PADs. Teams were allowed to use pre-trained weights on non-face data. The organizers provided pre-processing code for the training data, but it was not mandatory for teams to employ it. Each team was then requested to submit either a Win32 or Linux executable or, if wanted, their Python script. To ensure the integrity of the competition, the top three ranked solutions were examined by the organizers, i.e. these models were re-trained to validate that only the provided synthetic data was used for training and no pre-trained weights for faces were used. All solutions were evaluated on a restricted system without an internet connection to prevent any potential data leaks.

3. Submitted Solutions

In total, 14 teams have registered for the competition. Each team was allowed to submit up to two submissions. Eventually, 11 valid submissions from eight different teams were received. Solution names, team members, affiliations, and type of institution (academic or industry) are summarized in Table 2. Two teams opted to keep their names and affiliations anonymized. A condensed summary of the details of the approaches (e.g., base architecture, data augmentation, selection of checkpoints) is listed in Table 3. Anonymous-2 did not provide a detailed description of their approach. In the following, a brief description of the valid

Solution	Base architecture	Augmentation	Init. Weight	Loss function	Hardware	Selec. of model	FLOPs (G)	Param. (M)
ViT-SIDE B	ViT-Tiny	CJ, JPEG compression, rotation, blur, random crop	ImageNet	CE	NVIDIA T4, 16GB	Minmun loss with a maximum epoch of 100 and a patience of 20	1.08	5.524
SynFace Co-Former B	Swin-transformer	CJ, HF, SR, GA, RGB-S	scratch	CE	RTX 3090, 24GB	At 6 epochs	28.09	168.74
SynFace Co-Former A		Proposed reflection simulation augmentation						
CoDe-Lc	AlexNet	CJ, HF, SR, GA, RGB-S	scratch	BCE, MSE, Cosine similarity loss	RTX 3080 Ti, 12GB	At 200 epochs	1.42	115.06
CoDe-Lh		+ Gaussian blur		BCE, MSE, hypersphere loss			1.42	114.53
OrthPADNet	ResNet-18	CJ, HF, SR, GA, RGB-S, JPEG compression	scratch	CE with orthogonal projection loss, BCE, Cosine similarity loss	RTX 3090, 24GB	At 30 epochs	43.81	55.4
hdaFVPAD	Fisher Vector + SVM	-	-	-	-	-	-	0.0015
idvcVT	Swin-Transformer	-	ImageNet	BCE	RTX 2080 Ti, 11GB	Grid Search	4.36	28.29
Saliency-ResNet-CAS	ResNet-50	CJ, random erasing, channel normalization	-	FocalLoss	RTX 3060, 6GB	Best AUC at val set	-	36.92
Saliency-ResNet-ES	ResNet-50							

Table 3. Basic details of the submitted algorithms. SynFace Co-Former B and A refer to Baseline and the proposed reflection simulation Augmentation technique. Saliency-ResNet-CAS indicates use Cosine Annealing Scheduler (CAS) and Saliency-ResNet-ES refers to use Exponential Scheduler (ES). CJ, HF, SR, GA, RGB-S in Augmentation column refer to Color Jittering, Horizontal Flipping, Scale and Rotation, Gamma Adjustment, and RGB-Shift, respectively.

submitted solutions is provided:

ViT-SIDE B: The proposed approach involved fine-tuning a pre-trained ViT model architecture [16], specifically vit-tiny-patch16-224 from the timm library [53], on the provided synthetic data. To optimize the model’s performance, a binary cross-entropy (BCE) loss function was utilized. The Adam [39] optimizer with an initial learning rate of 1e-5 was employed along with step learning rate scheduler, which reduces the learning rate by a factor of 0.7 every 16 epochs until reaching a minimum value of 1e-7. The model was trained for a minimum of 10 epochs, with the checkpoint exhibiting the minimum loss selected as the final model. During training, a weighted sampling strategy was employed to maintain a balanced bona fide (genuine) to attack ratio of 1:1. A batch size of 32 was used, comprising 16 distinct bona fide samples randomly selected. For each bona fide sample, an attack sample with the same identity was paired. The attacks were uniformly sampled from the available sub-attacks. To enhance the model’s robustness, various data augmentations from Albumentation [8] were applied, including JPEG compression, rotation, color jitter, and blurring. These augmentations introduced variations in the training data, enabling the model to handle different types of spoofing attempts.

SynFace Co-Former Base (B) and Aug (A): The proposed Co-Former consists of three Transformer-based branches to extract different-level semantic features, including shallow, normal, and deep semantic features, by using tiny, small, and normal-scale swin-transformer [44]. Then, the extracted features are concatenated and processed by two linear layers to cooperatively predict the final results. In addition to Co-Former B using several conventional data augmentation techniques, Co-Former A is a proposed reflection simulation method to augment the data for imitating the reflective effect caused by the material in practice. Concretely, Co-Former A randomly selects the coordinate

index as the reflection center, and then utilizes a 2-D Gaussian distribution to simulate diffusion reflections. In general, two solutions were given, which were training with the proposed reflection augmentation method (i.e. Co-Former A), and the conventional augmentation methods (horizontal flipping, scaling and rotating, random gamma adjustment, RGB shifting and color jittering) (i.e. Co-Former B), respectively. Two solutions were both trained from scratch, and supervised by the cross-entropy loss function. The Stochastic Gradient Descent (SGD) optimizer with a momentum of 0.9 and weight decay of 5e-4, and an exponential learning scheduler with a gamma of 0.998 was applied during training. The initial learning rate for training the ResNet models was set to 0.001. The training epoch is 6 to avoiding overfitting and the batch size is set to 14, respectively. Co-Former A serves as a data augmentation method which can be easily incorporated in any other models. The implementation codes and pre-trained models are publicly released² for research reproducibility. Despite two drawbacks associated with Co-Former: a large number of parameters and high computational cost, the performance does not significantly degrade when using only one branch as two branches provided more subtle patterns for PAD decision.

CoDe-Lc and CoDe-Lh: CoDe-Lc and CoDe-Lh (as shown in Figure 2) are two ensemble models consisting of dual branches using AlexNet [41] as backbone architecture. Both models were trained from scratch, utilizing a weighted sampling which was performed to ensure a bona fide-attack ratio of 1:1. For CoDe-Lc, the cosine similarity function was employed as the loss function to measure the discrepancy between the feature layers from each branch. Additionally, the Mean Squared Error (MSE) loss was used as an auxiliary metric to evaluate the similarity of features. The BCE loss was computed for the final prediction as well as

²SynFace Co-Former: <https://github.com/Zi-YuanYang/IJCB-SynFacePAD-DIG>

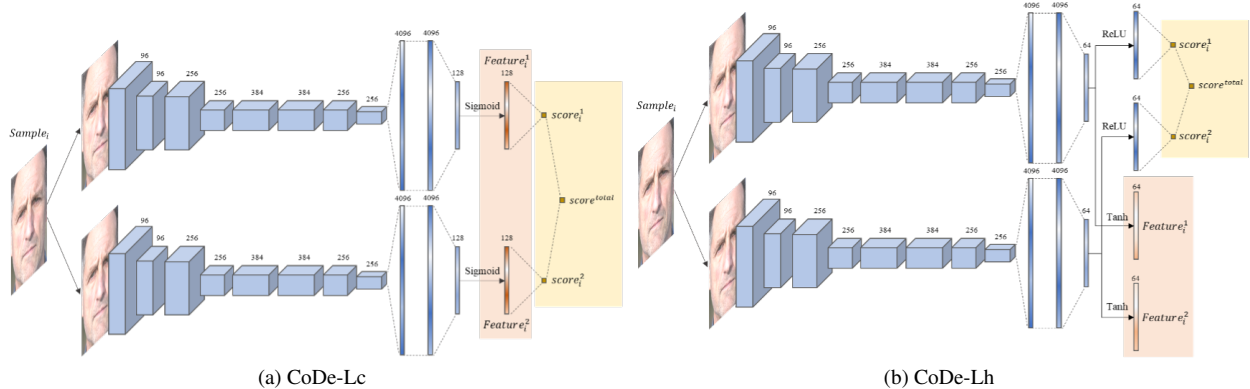


Figure 2. The pipeline of (a) CoDe-Lc and (b) CoDe-Lh (rank-3)

each branch’s prediction. The total loss was calculated as the cumulative sum of all the aforementioned losses. For CoDe-Lh, the cosine similarity was replaced with the hypersphere loss [43]. The input images were resized to dimensions of 224×224 , and data augmentation techniques were applied, including random horizontal flipping, scaling and rotating, gamma adjustment, RGB shifting, and color jittering. Moreover, for CoDe-Lh, additional augmentation was introduced by applying random Gaussian blur. The Adam optimizer with a learning rate of $1e-4$ and weight decay of $5e-4$ was utilized, along with an exponential learning scheduler with a gamma value of 0.998. The batch size during training was set to 128, and the number of training epoch was defined as 200.

OrthPADNet: The proposed method, OrthPADNet (as shown in Figure 3), is a fusion of two networks: the PDA net and the ID net. A two-stream backbone is applied to both networks. The two-stream backbone applies two ResNet18 [29] to extract features from both the original input and its CLAHE-augmented [45] version to fuse a final image feature. The PAD net has a structure similar to [31], which extracts two perpendicular features from the raw image feature extracted by the backbone, but only one of them is used for the PAD task. The ID net extracts two perpendicular features from the raw features, which are used for PAD and ID classification tasks respectively, and also only the PAD-related part is used for the final decision. Cross entropy loss with orthogonal projection loss [48] is used for classification, and cosine similarity loss is used for guaranteeing the orthogonality of the features. Apart from the face detection and cropping process provided by the organizers, they applied JPEG compression, rotation, flipping, RGB shift, and eye dropout [14] for data augmentation.

hdafvPAD: The Fisher Vectors (FV) approach for face PAD [26] derives a kernel from the parameters of a generative model such as Gaussian Mixture Models (GMM) on K -components. In essence, this representation characterises how the distribution of a set of local descriptors, extracted from unknown PAI species, differs from the distribution of

known attacks and bona fides, which is previously learned by the generative model. Thus, the most significant properties of the sub-population are summarised. Since image convolution with a suitable filter can effectively quantify frequency variations, local dense-Binarized Statistical Image Features (BSIF) features are used in the approach for image description [25]. Given the high correlation among the three RGB color components, these local descriptors are first decorrelated by Principal Component Analysis, thus reducing their size to $d = 64$ components while retaining 95% of the system variance. Then, the FV representation captures the average first- and second-order statistic differences between the local features and each semantic sub-group previously learned by the GMM. The final FV components are assumed to be independent of each other, allowing the correct use of statistical techniques (e.g., Support Vector Machines) which rely upon assumptions of independence. Overall, the final transformed features are more robust to new samples, which may stem from unknown scenarios and thus differ from the samples used for training.

idvcVT: The submitted idvcVT is based on the Swin-Transformer architecture [44] with a multi-class linear classifier as the final stage. Swin-Transformer builds hierarchical feature maps by merging image patches in deeper layers and has linear computation complexity to input image size due to computation of self-attention only within each local window. Two attack classes were taken into consideration in our case: bona fide and attacks. Input images were transformed according to ImageNet’s transformations of RGB images and resized to $256 \times 256 \times 3$ pixels. The model was fine-tuned from ImageNet 1K weights for 50 epochs and results were computed using epoch 10 (best validation performance). Softmax was used on the linear classifier’s output to get the score of the bona fide class.

Anonymous-1: A ResNet-50 [30] architecture enhanced with MixStyle for domain generalization [57] and Descriptive Convolutions [33] with four additional fully connected linear layers for classification. A saliency channel [37] was added to the RGB inputs to enhance the detection of cer-

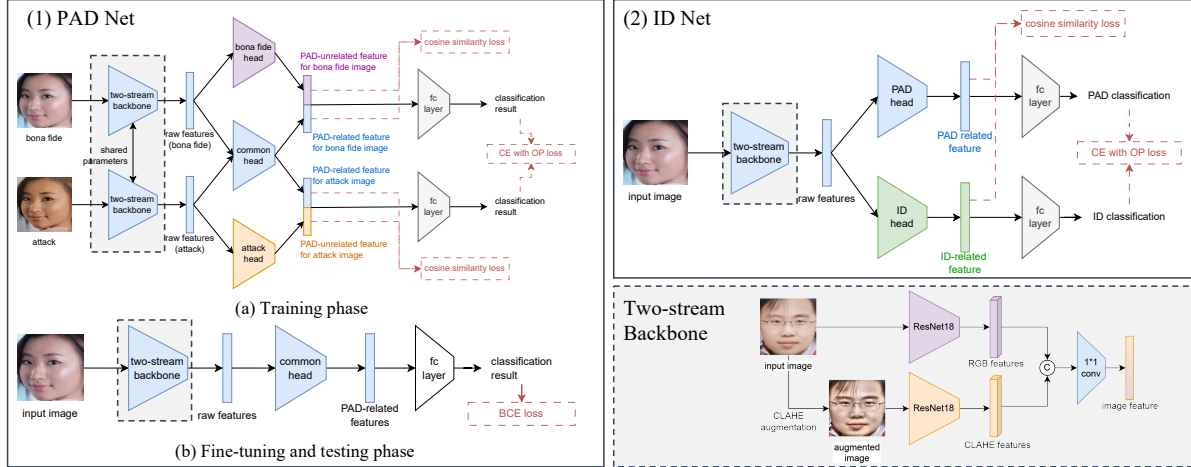


Figure 3. The pipeline of OrthPADNet (rank-4) consisting of (1) PAD Net learning PAD task-related features for classification by learning perpendicular feature pairs, and (2) ID Net learning PAD-related but ID-unrelated information. Both PAD Net and ID Net contain a dual-stream backbone to extract features from both the RGB and its CLAHE-augmented version as CLAHE augmentation magnifies the high-frequency details in the images.

tain spoof attacks. The model was supervised using Binary Focal Loss with class weights adjusted to tackle class imbalance. The ADAM optimizer with a betas of [0.9, 0.999] and a weight decay of $5e-4$ was used following a cosine annealing scheduler with a period of 20 epochs. The model was trained for 50 epochs with a batch size of 64. Training data is augmented with color jitter, random erasing, and channel normalization.

4. Results and Analysis

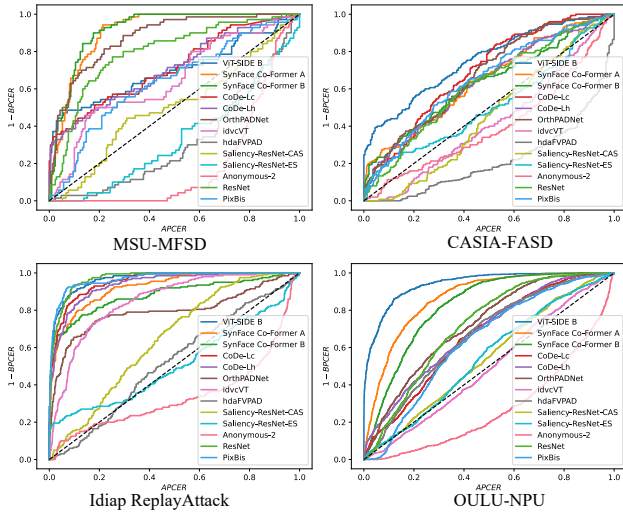


Figure 4. ROC curves of 11 submitted solutions and two baseline methods tested on four authentic face PAD benchmarks.

This section presents the evaluation results of the submitted solutions and the baselines on four authentic face PAD datasets in terms of the metrics introduced in Section 2. Note that APCER at BPCER of 20% is used for ranking solutions and the final ranking is based on the average ranking on all benchmarks. As shown in Table 8, Team ID R&D

Inc with solution ViT-SIDE B achieved the top-1 rank.

4.1. Analysis

MSU-MFSD: Table 4 presents the results of all submitted solutions and the two baseline methods on MSU-MFSD benchmarks. We observe that (1) three solutions outperformed baseline ResNet and six solutions outperformed PixBis in terms of $APCER@BPCER=20\%$. (2) SynFace Co-Former B and A achieved the lowest two error rates. (3) Multiple branches help in capturing fine-grained and subtle PAD patterns, as evidenced by the top-3 ranked solutions.

CASIA-FASD: Table 5 presents the results of all submitted solutions and two baseline methods on CASIA-FASD benchmarks. We observe that (1) five solutions outperformed baseline methods. (2) ViT-SIDE B achieved the best performance, followed by CoDe-Lc and OrthPADNet. (3) all solutions achieved APCER higher than 48% at BPCER of 20%, indicating that CASIA-FASD is more challenging than the other three benchmarks. This might be attributed to that CASIA-FASD includes a photo attack with eye region cut out, which is not seen in the training data.

Idiap ReplayAttack: Table 6 presents the results of all submitted solutions and two baseline methods on the Idiap ReplayAttack benchmark. We observe that (1) Most submitted solutions achieved their best performance on this benchmark, maybe due to a diverse range of replay attacks (more than print attacks) in the training data of SynthASpoof [21]. (2) The best performance in terms of $APCER@BPCER=20\%$ is achieved by the ViT-SIDE B solution, where the achieved APCER is 4.40%.

OULU-NPU: Table 7 presents the results of all submitted solutions and two baseline methods on OULU-NPU benchmark. We observe that (1) ViT-SIDE B achieved superior performance compared to all solutions, with an

Solutions	MSU-MFDS [52]						Rank
	HTER(%)↓	AUC(%)↑	BPCER(%)↓ @		APCER(%)↓ @		
			APCER 10%	APCER 20%	BPCER 10%	BPCER 20%	
ResNet [21]	25.48	79.54	57.14	32.86	62.38	33.33	
PixBis [21]	38.33	63.87	84.29	61.43	80.00	68.57	
SynFace Co-Former B	16.67	91.61	32.86	10.00	19.52	12.86	1
SynFace Co-Former A	18.57	90.76	34.29	12.86	20.48	17.62	2
OrthPADNet	20.95	87.59	35.71	24.29	33.81	23.81	3
CoDe-Lc	37.14	71.45	58.57	48.57	68.57	57.14	4
CoDe-Lh	39.05	70.58	61.43	51.43	74.76	60.00	5
idvcVT	45.71	64.58	84.29	51.43	84.76	61.43	6
ViT-SIDE B	36.67	69.78	51.43	47.14	89.52	72.38	7
Saliency-ResNet-CAS	47.62	50.14	94.29	81.42	90.48	82.86	8
Anonymous-2	72.86	19.98	100.00	100.00	68.57	90.00	9
hdaFVPAD	65.71	31.95	98.57	97.14	96.2	92.38	10
Saliency-ResNet-ES	58.09	33.49	100	98.71	97.62	92.38	10

Table 4. The comparative evaluation results of the submitted solutions on MSU-MFSD [52] benchmark. The ranking is based on APCER@BPCER=20%. The top-3 ranked solutions, using multiple branches to capture more subtle PAD patterns, demonstrated an enhanced generalizability.

Solutions	CASIA-FASD[56]						Rank
	HTER(%)↓	AUC(%)↑	BPCER(%)↓ @		APCER(%)↓ @		
			APCER 10%	APCER 20%	BPCER 10%	BPCER 20%	
ResNet [21]	39.22	62.00	84.67	66.67	79.78	68.44	
PixBis [21]	38.44	64.79	79.33	62.67	78.00	59.33	
ViT-SIDE B	33.33	75.21	57.33	45.33	65.56	49.11	1
CoDe-Lc	37.11	69.08	78.67	66.67	68.44	50.44	2
OrthPADNet	39.78	67.32	74.67	62.00	66.67	56.00	3
CoDe-Lh	39.33	63.70	76.67	64.00	79.11	63.33	4
SynFace Co-Former A	41.11	64.49	72.00	64.00	78.89	64.67	5
SynFace Co-Former B	40.00	63.05	74.67	61.33	81.56	71.78	6
Saliency-ResNet-CAS	51.55	45.31	99.33	90.67	86.89	79.78	7
Saliency-ResNet-ES	51.11	51.09	80.00	74.00	87.78	83.11	8
idvcVT	56.44	41.82	98.67	89.33	91.56	85.33	9
Anonymous-2	60.22	40.04	90.67	84.00	94.22	89.33	10
hdaFVPAD	71.33	21.95	100.00	97.56	99.33	97.33	11

Table 5. The comparative evaluation results of the submitted solutions on CASIA-FASD [56] benchmark. The submitted solutions generalized not well on CAISA-FASD, in comparison to on the other three benchmarks. The possible reason is that CASIA-FASD contains a printed photo attack sample with eye region cut out, which is not present in the synthetic training data.

APCER of 9.14% at 20% BPCER and the rank 2 solution SynFace Co-Former A achieved an APCER of 22.88% at 20% BPCER. (2) The top 3 ranked solutions employed transformer as the base network, suggesting the relatively higher generalizability of self-attention based transformer.

Figure 4 presents the achieved performance in terms of ROC curves on each benchmark by 11 submitted solutions and two baseline methods. A consistent observation can be made: 1) Most of the submitted solutions achieved better performance on MSU-MFSD, Idiap ReplayAttack, and OULU-NPU benchmarks, in comparison to CASIA-FASD. 2) Most of the presented solutions are very competitive and achieve better performance compared to the two baseline models. The results of Anonymous-2 solutions indicate a strong over-fitting of the model. However, it is hard to make a specific analysis without information on the submitted model. Notely, all models were limited to training on the synthetic data without any access to authentic faces and the best checkpoint is mostly determined by the fixed epochs or the training loss (details in Table 3 and descriptions in Section 3). This further proved the feasibility of

using synthetic data for developing face PADs. In addition, the possible domain gaps between authentic and synthetic data can be targeted in the future by including more attack types and formulate such problem in cross-domain.

4.2. Comparison and Final Ranking

Table 8 presents the final ranking based on the average rank achieved on the four authentic face PAD benchmarks. From the ranking outcome, we made the following general observations: (1) Solutions using transformer-based architecture as the base network generally exhibited higher PA detectability compared to CNNs. For example, ViT-SIDE B based on ViT-Tiny and SynFace Co-Former based on Swin-Transformer ranked first and second, respectively. (2) The incorporation of diverse data augmentation techniques helped in enhancing the generalizability of PADs. This can be observed when comparing SynFace Co-Former and idvcVT, both of which employ Swin-Transformer as the base architecture. (3) Using multiple branches or models contributed to an accurate and generalized PAD decision. (4) Deep-learning based solutions obtained outperformed hand-crafted feature-based methods in most cases.

Solutions	Idaip ReplayAttack[10]						Rank
	HTER(%)↓	AUC(%)↑	BPCER (%)↓@		APCER (%)↓@		
			APCER 10%	APCER 20%	BPCER 10%	BPCER 20%	
ResNet [21]	8.90	96.96	7.00	2.50	8.50	5.10	
PixBis [21]	7.50	96.88	7.50	5.50	6.50	3.00	
ViT-SIDE B	9.80	96.67	10.50	3.50	10.50	4.40	1
CoDe-Lc	12.10	95.31	14.50	7.00	15.70	6.60	2
CoDe-Lh	13.90	93.84	18.50	9.00	16.90	9.40	3
SynFace Co-Former A	16.30	92.31	23.00	14.50	24.20	13.20	4
SynFace Co-Former B	18.80	88.20	27.50	19.00	41.90	18.70	5
idvcVT	23.10	85.08	42.00	25.50	40.40	25.40	6
OrthPADNet	23.70	79.55	35.50	25.50	78.00	57.30	7
Saliency-ResNet-CAS	40.80	64.48	85.00	73.00	67.10	57.90	8
hdaFVPAD	47.80	52.10	94.00	83.00	89.70	74.70	9
Saliency-ResNet-ES	50.70	51.25	75.50	72.00	94.10	86.20	10
Anonymous-2	64.00	34.05	87.50	81.00	96.50	94.50	11

Table 6. The comparative evaluation results of the submitted solutions on Idiap ReplayAttack [10] benchmark. Most submitted solutions obtained the best performance on this benchmarks which may be benefit from a diverse replay attack in training set.

Solutions	OULU-NPU[1]						Rank
	HTER(%)↓	AUC(%)↑	BPCER (%)↓@		APCER (%)↓@		
			APCER 10%	APCER 20%	BPCER 10%	BPCER 20%	
ResNet [21]	31.48	71.48	79.80	59.70	57.07	45.56	
PixBis [21]	35.77	67.71	86.57	64.75	65.56	49.60	
ViT-SIDE B	13.26	94.04	18.79	8.69	17.68	9.14	1
SynFace Co-Former A	21.67	86.44	43.63	23.74	33.08	22.88	2
SynFace Co-Former B	25.35	82.02	61.11	34.24	40.23	29.34	3
OrthPADNet	34.92	71.69	73.13	54.55	60.76	49.14	4
hdaFVPAD	37.89	66.49	81.82	62.53	70.10	54.89	5
CoDe-Lh	38.11	68.33	75.66	58.59	69.09	55.73	6
CoDe-Lc	37.58	66.30	81.21	66.46	68.46	56.94	7
Saliency-ResNet-ES	46.03	54.7	89.29	81.11	83.61	70.71	8
Saliency-ResNet-CAS	47.55	54.4	89.89	77.68	83.31	71.67	9
idvcVT	51.09	49.68	93.03	83.13	86.21	76.59	10
Anonymous-2	66.39	27.86	99.09	95.35	98.06	95.61	11

Table 7. The comparative evaluation results of the submitted solutions on OULU-NPU [1] benchmark. The top-3 ranked solutions obtained a superior performance. In addition to their innovative designs, this may be related to two shared factors 1) the base networks of these solutions are self-attention based transformers, 2) extensive augmentation techniques.

Teams	Solutions	Ranks					
		M	C	I	O	Avg	Final rank
ID R&D Inc	ViT-SIDE B	7	1	1	1	2.50	1
SCU-DIG	SynFace Co-Former A	2	5	4	2	3.25	2
	SynFace Co-Former B	1	6	5	3	3.75	
HIT	CoDe-Lc	4	2	2	7	3.75	3
	CoDe-Lh	5	4	3	6	4.50	
BUCEA	OrthPADNet	3	3	7	4	4.25	4
idvc	idvcVT	6	9	6	10	7.75	5
Anonymous-1	Saliency-ResNet-CAS	8	7	8	9	8.00	6
	Saliency-ResNet-ES	10	8	10	8	9.00	
hda	hdaFVPAD	10	11	9	5	8.75	7
Anonymous-2	Anonymous-2	9	10	11	11	10.25	8

Table 8. The final ranking of the participating teams based on the average ranking of their best performance of the submitted solutions on four benchmarks. The bold number is the best performance of each team.

In the final ranking, Team ID R&D Inc with solution ViT-SIDE B won the competition with an average rank of 2.50, the second place was achieved by the SCU-DIG team with SynFace Co-Former A model (average rank 3.25), and the third place is obtained by Team HIT with CoDe-Lc model (average rank 3.75), as detailed in Table 8.

5. Conclusion

In this paper, we summarized the results and observations of the SynFacePAD 2023: Competition on Face Pre-

sentation Attack Detection Based on Privacy-aware Synthetic Training Data. In total, 14 teams registered for participation, and eight of them submitted 11 valid submissions to address the challenges of face PAD while considering privacy and legal concerns associated with authentic development data. The evaluation of the submitted solutions was conducted on four publicly available authentic face PAD benchmarks. The competition showcased various innovative approaches, resulting in improved performance compared to the considered two baseline methods. The enhanced PAD performance demonstrated the feasibility of using synthetic data and highlighted the potential of synthetic data in the development of face PAD systems.

Acknowledgment: This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE. J.F. is supported by project BBforTAI (PID2021-127641OB-I00MICINN/FEDER).

References

- [1] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid. OULU-NPU: A mobile face presentation attack database with real-world variations. In *FG*, pages 612–618. IEEE Computer Society, 2017.
- [2] F. Boutros, N. Damer, M. Fang, F. Kirchbuchner, and A. Kuijper. Mixfacenets: Extremely efficient face recognition networks. In *International IEEE Joint Conference on Biometrics, IJCB 2021, Shenzhen, China, August 4-7, 2021*, pages 1–8. IEEE, 2021.
- [3] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2022, New Orleans, LA, USA, 2022*, pages 1577–1586. IEEE, 2022.
- [4] F. Boutros, N. Damer, and A. Kuijper. Quantface: Towards lightweight face recognition by synthetic data low-bit quantization. In *26th International Conference on Pattern Recognition, ICPR 2022, Montreal, QC, Canada, August 21-25, 2022*, pages 855–862. IEEE, 2022.
- [5] F. Boutros, M. Huber, P. Siebke, T. Rieber, and N. Damer. Sface: Privacy-friendly and accurate face recognition using synthetic data. In *IJCB*, pages 1–11. IEEE, 2022.
- [6] F. Boutros, M. Klemm, M. Fang, A. Kuijper, and N. Damer. Unsupervised face recognition using unlabeled synthetic data. In *17th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2023, Waikoloa Beach, HI, USA, January 5-8, 2023*, pages 1–8. IEEE, 2023.
- [7] F. Boutros, V. Struc, J. Fierrez, and N. Damer. Synthetic data for face recognition: Current state and future prospects. *Image and Vision Computing*, 135:104688, 2023.
- [8] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin. Alumentations: Fast and flexible image augmentations. *Information*, 11(2), 2020.
- [9] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *FG*, pages 67–74. IEEE Computer Society, 2018.
- [10] I. Chingovska, A. Anjos, and S. Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *BIO SIG*, volume P-196 of *LNI*, pages 1–7. GI, 2012.
- [11] César Augusto Fontanillo López and Abdullah Elbi. On synthetic data: a brief introduction for data protection law dummies, 2022.
- [12] N. Damer, M. Fang, P. Siebke, J. N. Kolf, M. Huber, and F. Boutros. Mordiff: Recognition vulnerability and attack detectability of face morphing attacks created by diffusion autoencoders. In *IWBF*, pages 1–6. IEEE, 2023.
- [13] N. Damer, C. A. F. López, M. Fang, N. Spiller, M. V. Pham, and F. Boutros. Privacy-friendly synthetic data for the development of face morphing attack detectors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2022, New Orleans, LA, USA, June 19-20, 2022*, pages 1605–1616. IEEE, 2022.
- [14] S. Das, S. Seferbekov, A. Datta, M. S. Islam, and M. R. Amin. Towards solving the deepfake problem: An analysis on improving deepfake detection using dynamic face augmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3776–3785, 2021.
- [15] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 4690–4699. Computer Vision Foundation / IEEE, 2019.
- [16] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [17] M. Fang, H. Ali, A. Kuijper, and N. Damer. Patchswap: Boosting the generalizability of face presentation attack detection by identity-aware patch swapping. In *IEEE International Joint Conference on Biometrics, IJCB 2022, Abu Dhabi, United Arab Emirates, October 10-13, 2022*, pages 1–10. IEEE, 2022.
- [18] M. Fang, F. Boutros, and N. Damer. Unsupervised face morphing attack detection via self-paced anomaly detection. In *IEEE International Joint Conference on Biometrics, IJCB 2022, Abu Dhabi, United Arab Emirates, October 10-13, 2022*, pages 1–11. IEEE, 2022.
- [19] M. Fang, N. Damer, F. Kirchbuchner, and A. Kuijper. Learnable multi-level frequency decomposition and hierarchical attention mechanism for generalized face presentation attack detection. In *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2022, Waikoloa, HI, USA, January 3-8, 2022*, pages 1131–1140. IEEE, 2022.
- [20] M. Fang, N. Damer, F. Kirchbuchner, and A. Kuijper. Real masks and spoof faces: On the masked face presentation attack detection. *Pattern Recognit.*, 123:108398, 2022.
- [21] M. Fang, M. Huber, and N. Damer. Synthaspoof: Developing face presentation attack detection based on privacy-friendly synthetic data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1061–1070, June 2023.
- [22] J. Galbally, S. Marcel, and J. Fierrez. Image quality assessment for fake biometric detection: Application to iris, fingerprint and face recognition. *IEEE Trans. on Image Processing*, 23(2):710–724, February 2014.
- [23] A. George and S. Marcel. Deep pixel-wise binary supervision for face presentation attack detection. In *ICB*, pages 1–8. IEEE, 2019.
- [24] L. F. Gomez, J. Fierrez, et al. PAD-Phys: Exploiting physiology for presentation attack detection in face biometrics. In *IEEE Conf. on Computers, Software, and Applications (COMPSAC)*, June 2023.
- [25] L. J. Gonzalez-Soler, M. Gomez-Barrero, and C. Busch. Fisher vector encoding of dense-bsif features for unknown face presentation attack detection. In *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIO SIG)*, pages 1–6. IEEE, 2020.
- [26] L. J. Gonzalez-Soler, M. Gomez-Barrero, and C. Busch. On the generalisation capabilities of fisher vector-based face presentation attack detection. *IET Biometrics*, 10(5):480–496, 2021.

- [27] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *ECCV (3)*, volume 9907 of *Lecture Notes in Computer Science*, pages 87–102. Springer, 2016.
- [28] A. Hadid, N. Evans, S. Marcel, and J. Fierrez. Biometrics systems under spoofing attack: an evaluation methodology and lessons learned. *IEEE Signal Processing Magazine*, 32(5):20–30, September 2015.
- [29] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [30] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [31] R. He, X. Wu, Z. Sun, and T. Tan. Learning invariant deep representation for nir-vis face recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.
- [32] J. Hernandez-Ortega, J. Fierrez, A. Morales, and J. Galbally. *Introduction to Presentation Attack Detection in Face Biometrics and Recent Advances*. Springer, 2023. 3rd Ed.
- [33] P.-K. Huang, H.-Y. Ni, Y.-Q. Ni, and C.-T. Hsu. Learnable descriptive convolutional network for face anti-spoofing. In *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*. BMVA Press, 2022.
- [34] M. Huber, F. Boutros, A. T. Luu, K. B. Raja, R. Ramachandra, N. Damer, P. C. Neto, T. Gonçalves, A. F. Sequeira, J. S. Cardoso, J. Tremoço, M. Lourenço, S. Serra, E. Cermeño, M. Ivanovska, B. Batagelj, A. Kronovsek, P. Peer, and V. Struc. SYN-MAD 2022: Competition on face morphing attack detection based on privacy-aware synthetic training data. In *IJCB*, pages 1–10. IEEE, 2022.
- [35] International Organization for Standardization. ISO/IEC DIS 30107-3:2016: Information Technology – Biometric presentation attack detection – P. 3: Testing and reporting, 2017.
- [36] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila. Training generative adversarial networks with limited data. In *NeurIPS*, 2020.
- [37] I. Katramados and T. P. Breckon. Real-time visual saliency by division of gaussians. In *2011 18th IEEE International Conference on Image Processing*, pages 1701–1704, 2011.
- [38] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard. The megaface benchmark: 1 million faces for recognition at scale. In *CVPR*, pages 4873–4882. IEEE Computer Society, 2016.
- [39] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [40] J. N. Kolf, J. Elliesen, F. Boutros, H. Proença, and N. Damer. Syper: Synthetic periocular data for quantized light-weight recognition in the NIR and visible domains. *Image Vis. Comput.*, 135:104692, 2023.
- [41] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, may 2017.
- [42] H. Li, S. J. Pan, S. Wang, and A. C. Kot. Domain generalization with adversarial feature learning. In *CVPR*, pages 5400–5409. IEEE Computer Society, 2018.
- [43] Z. Li, H. Li, K.-Y. Lam, and A. C. Kot. Unseen face presentation attack detection with hypersphere loss. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2852–2856, 2020.
- [44] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.
- [45] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.
- [46] S. Purnapatra, N. Smalt, K. Bahmani, P. Das, D. Yambay, A. Mohammadi, A. George, T. Bourlai, S. Marcel, S. Schuckers, M. Fang, N. Damer, F. Boutros, A. Kuijper, A. Kantarci, B. Demir, Z. Yildiz, Z. Ghafoory, H. Dertli, H. K. Ekenel, S. Vu, V. Christophides, D. Liang, G. Zhang, Z. Hao, J. Liu, Y. Jin, S. Liu, S. Huang, S. Kuei, J. M. Singh, and R. Ramachandra. Face liveness detection competition (livdet-face) - 2021. In *International IEEE Joint Conference on Biometrics, IJCB 2021, Shenzhen, China, August 4-7, 2021*, pages 1–10. IEEE, 2021.
- [47] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, and D. Tao. Synface: Face recognition with synthetic data. In *ICCV*, pages 10860–10870. IEEE, 2021.
- [48] K. Ranasinghe, M. Naseer, M. Hayat, S. Khan, and F. S. Khan. Orthogonal projection loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12333–12343, 2021.
- [49] R. Shao, X. Lan, J. Li, and P. C. Yuen. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *CVPR*, pages 10023–10031. Computer Vision Foundation / IEEE, 2019.
- [50] R. Shao, X. Lan, and P. C. Yuen. Regularized fine-grained meta face anti-spoofing. In *AAAI*, pages 11974–11981. AAAI Press, 2020.
- [51] R. Tolosana, C. Rathgeb, et al. *Future Trends in Digital Face Manipulation and Detection*, chapter Future Trends in Digital Face Manipulation and Detection, pages 463–482. January 2022.
- [52] D. Wen, H. Han, and A. K. Jain. Face spoof detection with image distortion analysis. *IEEE Trans. Inf. Forensics Secur.*, 10(4):746–761, 2015.
- [53] R. Wightman. Pytorch image models. <https://github.com/rwightman/pytorch-image-models>, 2019.
- [54] Z. Yu, X. Li, J. Shi, Z. Xia, and G. Zhao. Revisiting pixel-wise supervision for face anti-spoofing. *IEEE Trans. Biom. Behav. Identity Sci.*, 3(3):285–295, 2021.
- [55] Y. Zhang, Z. Yin, Y. Li, G. Yin, J. Yan, J. Shao, and Z. Liu. Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations. In *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020*,

Proceedings, Part XII, volume 12357 of *Lecture Notes in Computer Science*, pages 70–85. Springer, 2020.

- [56] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. A face antispoofing database with diverse attacks. In *ICB*, pages 26–31. IEEE, 2012.
- [57] K. Zhou, Y. Yang, Y. Qiao, and T. Xiang. Mixstyle neural networks for domain generalization and adaptation, 2021.