Unified Physical-Digital Attack Detection Challenge

Haocheng Yuan[1]*, Ajian Liu[2]*, Junze Zheng[1], Jun Wan[1,2,6†]
Jiankang Deng[3], Sergio Escalera[4], Hugo Jair Escalante[7,8], Isabelle Guyon[8], Zhen Lei[2,5,6]
[1]M.U.S.T, Macau; [2]MAIS, CASIA, China; [3]ICL, UK; [4]CVC, Spain; [5]CAIR, HKISI, CAS;
[6]SAI, UCAS, China; [7]INAOE, CINVESTAV, Mexico; [8]ChaLearn, USA
[1]Allen.hcyuan@outlook.com, [2]{ajian.liu,jun.wan}@ia.ac.cn, [2]zlei@nlpr.ia.ac.cn

Abstract

Face Anti-Spoofing (FAS) is crucial to safeguard Face Recognition (FR) Systems. In real-world scenarios, FRs are confronted with both physical and digital attacks. However, existing algorithms often address only one type of attack at a time, which poses significant limitations in real-world scenarios where FR systems face hybrid physical-digital threats. To facilitate the research of Unified Attack Detection (UAD) algorithms, a large-scale UniAttackData dataset has been collected. UniAttackData is the largest public dataset for Unified Attack Detection, with a total of 28,706 videos, where each unique identity encompasses all advanced attack types. Based on this dataset, we organized a Unified Physical-Digital Face Attack Detection Challenge to boost the research in Unified Attack Detections. It attracted 136 teams for the development phase, with 13 qualifying for the final round. The results re-verified by the organizing team were used for the final ranking. This paper comprehensively reviews the challenge, detailing the dataset introduction, protocol definition, evaluation criteria, and a summary of published results. Finally, we focus on the detailed analysis of the highest-performing algorithms and offer potential directions for unified physical-digital attack detection inspired by this competition. Challenge Website: `https://sites.google.com/view/face-anti-spoofing-challenge/welcome/challengecvpr2024`

1. Introduction

Face Anti-Spoofing (FAS) is essential to ensure the security of Face Recognition (FR) systems by identifying whether the image is live or fake. Attacks and corresponding detection methods can be classified into physical and digital categories. Physical Attacks (PA) involve the presentation of face replicas, like prints, masks, and screen replies. With the release of several high-quality 2D datasets [33, 54, 56], and 3D-Mask datasets [10, 16, 30, 38], existing works [17, 18, 23, 25, 31, 32, 53] demonstrated a satisfying result, while other efforts [19, 28, 51] have concentrated on leveraging multi-modal information to uncover spoofing clues. Wang et al. [46] introduce a Multi-Domain Incremental Learning (MDIL) approach for Presentation Attack Detection (PAD), effectively acquiring new domain knowledge while preserving performance across previously learned domains. Digital Attacks (DA) aim to manipulate faces before they are presented for verification, involving Deepfakes [11, 12, 41, 47, 48] and adversarial attacks [37, 58]. With the release of the dataset [14, 44, 45], promising results have been obtained by current works [5, 9, 13, 27, 34, 42, 57, 59] through distinguishing digitally ma-

Ranking	Team Name	Leader Name, Affiliation
1	MTFace	Xianhua He, Meituan
2	SeaRecluse	Minzhe Huang, Akuvox
3	duileduile	Jiaruo Yu, INTSIG Information Co. Ltd
4	BSP-Idiap	Anjith George, Idiap Research Institute
5	VAI-Face	Vu Minh Quan, Viettel AI
6	L&L&W	Tongming Wan, Central South University
7	SARM	Jun Lan, SARM
8	M2-Purdue	Shu Hu, Purdue University
9	Cloud Recesses	Peipeng Yu, Nanyang Technological Univeristy
10	ImageLab	Sabari Nathan, Couger Inc., Sethu Institute of Technology, Thiagarajar college of engineering
11	BOVIFOCR-UFPR	Bernardo Biesseck, Federal University of Paraná
12	Inria-CENATAV-Tec	Luis Santiago Luévano García, Inria, CENATAV, Tecnologico de Monterrey
13	Vicognit	Manoj Sharma, Bennett Univeristy

Table 1. List of the team and affiliation names in the final ranking of this challenge.

*Equal contribution.
†Corresponding author.

nipulated facial artifacts. Recent work [29] utilizes large-scale VLMs and text features to dynamically adjust classifier weights, enhancing the exploration of generalizable visual features.

However, related studies continue to investigate PA Detection (PAD) and DA Detection (DAD) as distinct tasks, leading to extensive computing resources consumed during deployment. We identify two main reasons for the scarcities of Unified Attack Detection (UAD): (1) Lack of a large-scale dataset that unifies both physical and digital attacks. While two notable datasets, GrandFake [4] and JFSFDB [52] have been introduced to address the UAD challenge, their approach primarily combines PA and DA datasets. (2) Lack of a public baseline benchmark measuring the UAD algorithms. Although physical and digital attacks are categorized as fake in the final classification procedure, the significant differences between these attack types increase the intra-class distances. Current methods are tailored to address either physical or digital threats specifically. UAD algorithms and baselines for evaluating such algorithms are urgently needed.

Striving to propel advancements in the research community regarding UAD, we address the issues analyzed above through the following two aspects: (1) We collected and published a large-scale Unified Physical-Digital Attack dataset named UniAttackData [8]. Compared to current unified datasets, it has several advantages, such as the complete attack types of each ID, the most advanced forgery methods, and the amount of data. (2) We establish a broader and more valuable testing protocol, which emphasizes evaluating the generalization ability of UAD algorithms. (3) Based on this dataset, we successfully held the ***Unified Physical-Digital Attack Detection Challenge at CVPR2024***, which attracted 136 teams worldwide. The top three teams achieved results significantly surpassing our baseline. A summary containing the names of the team and affiliations who reached the final phase is shown in Tab. 1.

To sum up, the contributions of this paper are summarized as follows:

- We describe the design of the Unified Physical-Digital Attack Detection Challenge at CVPR2024.
- We organized this challenge around the UniAttackData, proving the suitability of such a resource for boosting research on the topic.
- We report and analyze the solutions developed by participants.
- We highlight critical factors in detecting both physical and digital by examining the top-ranked algorithms and suggest future research directions through this competition.

2. Related Work

2.1. Face Anti-spoofing datasets for Challenges

Print-Attack [1] addresses the prevalent issue of bypassing 2D FR systems using spoofed photographs. It comprises 400 samples: 200 genuine accesses and 200 videos using printed photos across 50 identities. **Replay-Attack** [3] is a dedicated resource focusing on two-dimensional PAs, which includes 1,300 video clips featuring various attack types with 50 subjects, recorded under controlled and adverse lighting conditions. **OULU-NPU** [2] is a public PAD database designed to test the generalization of FR systems across different environments like lighting and background, various smartphones, and presentation attack instruments. It includes 5,940 high-resolution videos of 55 subjects, recorded in three environments with six smartphone models.

CASIA-SURF [55] constitutes a large-scale resource that includes data from $1,000$ distinct subjects, featuring $21,000$ videos per subject, with each encompassing three modalities: RGB, Depth, and Infrared images. **CASIA-SURF CeFA** [22] is a pioneering resource that explores ethnic bias in FAS by encompassing data from three ethnicities, three modalities, and 1,607 individuals, complete with 2D and 3D attack types.

CASIA-SURF HiFiMask [26] is a large-scale, high-fidelity mask dataset introduced to address the limitations of current 3D mask attack detection benchmarks. It contains over $54,600$ videos from 75 individuals wearing 225 realistically crafted masks using seven different sensor types. **CASIA-SURF SuHiFiMask** [7] is a large-scale dataset designed for advancing FAS research. It specializes in protecting FR's security in distant surveillance scenarios and contains attack data from 101 individuals across various age ranges in 40 different surveillance settings.

UniAttackData [8] is the largest known dataset that unifies physical-digital attacks, involving each of 1,800 subjects with 2 physical and 12 digital attacks, which applied by SOTA attack methods over the last three years.

2.2. Face Anti-spoof Challenges

Since 2019, the Institute of Automation of the Chinese Academy of Sciences (CASIA) has held a series of FAS competitions based on the International Conference on Computer Vision.

The first competition **Multi-modal Face Anti-spoofing Attack Detection Challenge at CVPR-2019** [20] aimed to advance research in PAD by utilizing the large-scale dataset, CASIA-SURF. The challenge, held during CVPR2019, drew participation from over 300 global entities, with a majority from industry in the final stage, highlighting the practical significance of FAS in real-life applications.

The second competition **Cross-ethnicity Face Anti-**

spoofing Recognition Challenge at CVPR-2020 [21] addresses racial bias in FR systems, leveraging CASIA-SURF CeFA. The challenge featured separate tracks for single-modal (RGB) and multi-modal (combining with depth and infrared) FAS techniques. 340 teams participated in the development phase, with 11 and 8 teams advancing to the final evaluation for the single-modal and multi-modal tracks, respectively. The competition served as a catalyst for research focused on mitigating racial bias in facial spoofing detection.

The third competition **3D High-Fidelity Mask Face Presentation Attack Detection Challenge at ICCV-2021** [24] addressed the security risks associated with advanced 3D masks against FR systems. The challenge utilized the HiFiMask dataset featuring diverse and established a demanding testing protocol to measure algorithm performance, particularly distinguishing real vs. fake faces across varying conditions. The competition attracted 195 teams. Notably, many finalists hailed from industry backgrounds, underscoring the practical importance of this research for real-world applications.

The fourth competition **Surveillance Face Presentation Attack Detection Challenge at CVPR-2023** [6] was conducted to enhance the security of FR systems in surveillance settings, particularly in long-distance scenarios by leveraging a large-scale SuHiFiMask dataset. A total of 180 teams participated, with 37 advancing to the finals.

3. Challenge Overview

3.1. UniAttackData Dataset

As far as we know, the UniAttackData [8] is the largest unified physical-digital attack dataset, with a total of $28,706$ videos of $1,800$ subjects. It encompasses $1,800$ live face videos, $5,400$ videos showcasing PAs, and $21,506$ videos with DAs. Complete attack types for each ID are constructed to maintain consistency across physical and digital dimensions. This procedure avoids leading models inadvertently focusing on features unrelated to FAS tasks.

To facilitate the use of the dataset by the participating teams, the following pre-processing steps were carried out: (1) We clip and crop the facial part of the image from the original videos. (2) We choose one frame of each video and reconstruct the file's names, like *train/000001.png*, to hide the attack clues, like ethics or types.

3.2. Challenge Protocol and Data Statistics

To thoroughly evaluate the performance of UAD frameworks, we established two distinct protocols within the UniAttackData framework. According to Tab. 2, Protocol 1 is designed to scrutinize performance across unified attack tasks. Protocol 2, on the other hand, is tailored to assess algorithmic generalization across "unseen" attack types. Employing a "leave-one-type-out testing" strategy, we further divide Protocol 2 into two sub-protocols, which enables a thorough investigation into the generalization capacity of FAS mechanisms against a broad and evolving threat landscape.

Protocal	Class	Types				Total
		Live	Phys	Adv	Digital	
P1	train	3000	1800	1800	1800	8400
	eval	1500	900	1800	1800	6000
	test	4500	2700	7106	7200	21506
P2.1	train	3000	0	9000	9000	21000
	eval	1500	0	1706	1800	5006
	test	4500	5400	0	0	9900
P2.2	train	3000	2700	0	0	5700
	eval	1500	2700	0	0	4200
	test	4500	0	10706	10800	26006

Table 2. Two protocols used in this challenge.

3.3. Challenge Process and Timeline

The challenge was held on the CodaLab platform, including two phases as follows:

Development Phase: (Start: Feb. 1st, 2024 – Ended: Feb. 22nd). During this phase, participants can train their models with labeled training data and predict scores on unlabeled development data. Two data subsets include the same type of attacks (unified attacks for phase 1, digital for phase 2.1, and physical for phase 2.2). Participants can submit their predictions on the development set and get timely feedback via leader board.

Final Phase: (Start: Feb. 23rd – Ended: Mar.3rd). In the final phase, we released the development set labels for participants to refine their models for the test set, while also making the unlabeled test data available. Teams are required to predict outcomes for the test data and submit these to CodaLab. It is crucial that models are trained solely on the training set, without access to the development or test set data. Note that on the CodaLab platform, the last submission is considered the final entry.

The final ranking of participants was obtained from the submissions' performance in the testing sets. To be eligible for prizes, winners had to publicly release their code under a license and provide a fact sheet describing their solution.

3.4. Evaluation Metrics

In evaluating the performance for this challenge, we adopted the ISO/IEC 30107-3 metrics, which include the Attack Presentation Classification Error Rate (APCER), the Normal/Bona Presentation Classification Error Rate (NPCER/BPCER), and the Average Classification Error Rate (ACER). These metrics quantify the detection system's accuracy in identifying live and fake face presentations. The determination of ACER for the test set relies on the Equal Error Rate (EER) established during the development phase. Additionally, we used the Area Under Curve (AUC) metric as an additional performance measure, assessing model discrimination between fake and real samples across thresholds. Rankings were primarily based on the Average Classification Error Rate (ACER), with AUC as a secondary criterion, to thoroughly evaluate each algorithm's efficacy against spoofing attacks.

4. Description of solutions

4.1. MTFace

Due to the significant disparity between samples applying physical attacks and those applying digital attacks, team MTFace proposed a framework named Optimized Data Augmentation for Comprehensive Face Attack Detection Across Physical and Digital Domains. This framework focuses prominently on data augmentations and balanced training loss. The overall pipeline, including data augmentation and loss-balancing training, is shown in Fig. 1.

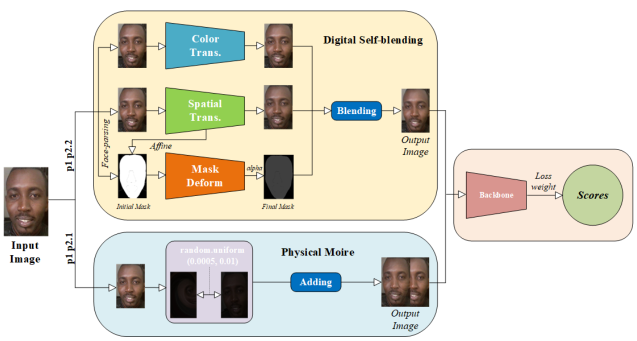

Figure 1. The overall pipeline of MTFace's method. The moiré augmentation improves the model's PAD ability, while self-blending augmentation improves the model's DAD ability.

During pre-processing, all uncropped images are face-detected and cropped 20 pixels outward by the bound box. Finally, all live data in the training set is parsed to obtain the face mask area for further use in subsequent self-blending methods. As shown in Fig. 1, a moiré data augmentation and a self-blending augmentation are applied to the different subsets based on the diverse structure of protocols. MTFace analyzes that screen replay will cause moiré patterns on the surface texture of the image. Thus, the moiré data augmentation method is designed and applied to protocol p1 and p2.1 so that the characteristics of physical attacks are introduced, improving the model's ability to detect physical attacks across domains. Second, inspired by the work [43], a self-blending augmentation method is defined. For protocols p1 and p2.2, extra live-based data has been added with digital attack features by color, spatial transformation, and mask deformation. To achieve a more balanced training approach in response to the varying ratios of bona and spoof samples across three protocols, MTFace adjusted the cross-entropy loss weights accordingly. For protocol p1, they maintained an equal loss weight ratio of live to fake samples at 1:1, aiming for an even-handed learning process. In protocol p2.1, they adopted a 5:1 loss weight ratio, favoring live data to encourage the model to learn live features more effectively. For protocol p2.2, they applied a 2:1 loss weight ratio to better equilibrate the impact of different sample types on the model's learning. This led to a marked enhancement in their experimental outcomes. Finally, ResNet-50 is set as the backbone, and the ImageNet pre-train weights are loaded. Their promising experiment results have beaten all other competitors in this challenge.

4.2. SeaRecluse

The team SeaRecluse presented a solution titled "Cross-domain Face Anti-spoofing in Unified Physical-Digital Attack Dataset." The team's approach included several key steps and techniques: SCRFD is first applied to uncropped images in the training set for face cropping. For three protocols, the dataset partitioning methods differ: 80% of the p1, 60% of p2.1, and 40% of p2.2's training data are used along with the validation data for training, while the remaining data serves as the validation set. Regarding data augmentation and preprocessing, cropped images undergo no additional processing, whereas uncropped images are supplemented with data through loose and tight facial cropping. For each of the three tasks, distinct data augmentation measures are taken based on their respective focuses: Task P1 does not perform any enhancement operations; for P2.1, to balance the ratio of real face samples, real face data is downsampled and edge blank pixels are filled to reach a fixed size, effectively tripling the amount of real face data; in contrast, for P2.2, similar augmentation strategies are adopted for fake face data, introducing 4x and 8x downsampling, thereby increasing the fake face data volume by sevenfold. During the data review stage, some images' incorrect aspect ratios were rectified and restored to their original proportions. All tasks implement normal data augmentation operations, with Task P2.1 also incorporating Gaussian blur. SeaRecluse chose ConvNeXt V2 backbone considering the competition requirements and local training resources for training. During the training phase, Image CutMix and label smoothing [36] techniques were leveraged to enhance the model's generalization capability.

4.3. duileduile

The team duileduile has devised a two-stage method integrating pre-training and fine-tuning techniques, employing the Swin Transformer model [35] as its backbone. This approach showcases a robust framework for generalizing the detection of unified face representations, and the solution pipeline is shown in Fig. 2.

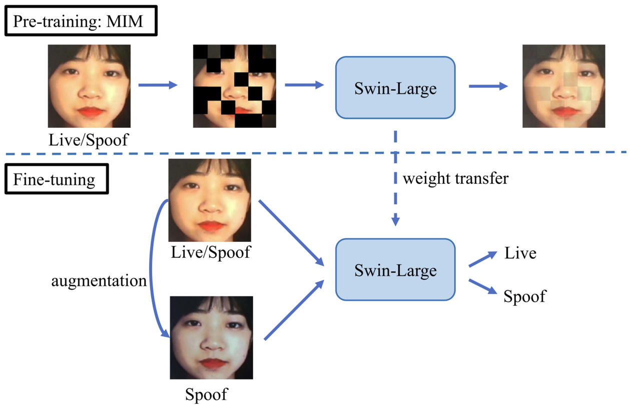

Figure 2. The pipeline of Team duileduile's framework with pre-training stage and fine-tuning stage.

The backbone model, the Swin-Large model [35], extracts features resulting in 1536-dimensional vectors, which, coupled with the image's dimensional attributes, aid in differentiating between live and spoof images. Both pre-training and fine-tuning stages follow identical configurations across all protocols. This self-contained strategy yields a competitive edge in cross-attack-type scenarios and simplifies the transplanting of the method across various baselines. The simMIM strategy [50] is utilized with unlabeled training and development sets at the pre-training stage, involving segmenting an image into non-overlapping patches, concealing portions and prompting the model to infer the complete image. This masked image modeling (MIM) methodology fosters the model's proficiency in feature extraction from incomplete visual information, which is helpful for protocol 2 in the testing phase. During the fine-tuning stage, they introduce a tailored data augmentation sequence. The Gaussian Noise augmentation represents digital threats, while ColorJitter and screen display simulations, including moiré pattern overlays and gamma correction, are used to mimic physical attack types. These sequential augmentations are applied exclusively to training samples with designated probabilities.

4.4. BSP-Idiap

The Dual Branch Pixel-wise Binary Supervision (DBPixBiS) algorithm, presented by the BSP-Idiap team, is a novel FAS solution that extends their previous work DeepPixBiS. As shown in Fig. 3, their DBPixBiS utilizes a dual-branch network architecture that takes Fourier transform representations as input to capture both spatial and frequency domain properties of images. In the RGB branch of the model, central difference convolutional blocks are employed to process the visual information differently from standard convolutional operations. A vital aspect of the method lies in its pixel-wise binary supervision, where each pixel in the feature maps is guided by an attentional angular margin loss function during the training phase. This supervision strategy regularizes the learning process, reducing the over-fitting problem and promoting better generalization across diverse spoofing scenarios. The architecture includes an additional branch tailored for Fourier-transformed inputs to detect spoofing artifacts in the frequency domain. During inference, the mean value of the resulting feature map is used as the final spoofing detection score.

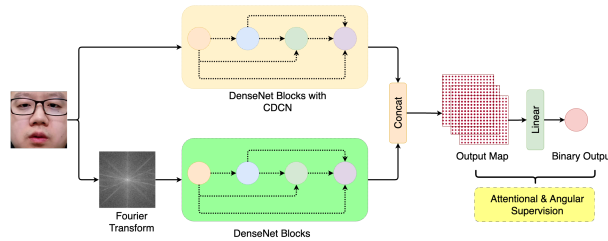

Figure 3. This figure shows the DBPixBiS's architecture, the feature map, and the final outputs are supervised by angular loss.

4.5. VAI-Face

The team VAI-Fac's approach is distilled into Fig. 4, briefly outlining their methodology. For this task, they utilized the Dinov2 Vision Transformer (ViT) large model [39]. A key element of their strategy is the application of disparate augmentation techniques to live and fake images to enhance model performance. Live images are treated with RandomResizedCrop and HorizontalFlip, while fake images are subjected to a more extensive augmentation suite that includes distortions, blurs, and custom cutouts to simulate the anomalies often found in spoofed images. Their learning strategies include utilizing OneCycleLR with finely-tuned hyperparameters and label smoothing, emphasizing precision in enhancing the model's learning efficiency. The VAI-Face team further bolstered their model's robustness with mixup augmentation and opted for the ADAN optimizer. In summary, the VAI-Face team's methodology showcases a considered balance between sophisticated data augmentation strategies, a thorough hyperparameter tuning process, and leveraging the capabilities of a powerful ViT model, making it a standout contribution in the arena of UAD.

4.6. L&L&W

The team L&L&W has introduced a method centered on patch-based feature learning and incorporating novel attention and sampling techniques for FAS. L&L&W's methodology starts with extracting small patches from images, which are then processed to learn discriminative features. They employ Centre Difference Attention to capture fine-grained intrinsic features, highlighting the subtle cues es-

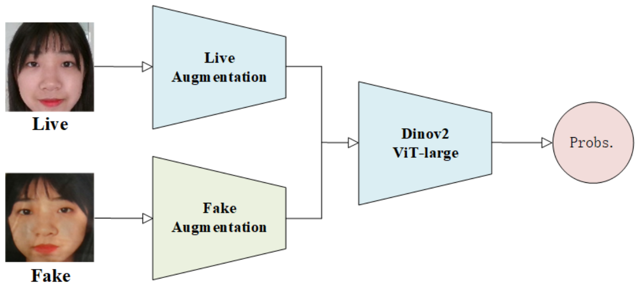

Figure 4. The overall structure of team VAI-Face's solution.

sential for spoof detection. Furthermore, they have developed a High-Frequency Wavelet Sampler, focusing on high-frequency components to robustly identify forgery traces in the spatial domain. This indicates a dual focus on spatial and frequency domains for a comprehensive feature extraction strategy. In validation and testing, they create 36 crops of an image and use the average of the predictions as the final score. This approach suggests an emphasis on capturing varied aspects of the images to enhance the prediction reliability.

4.7. SARM

Through analysis, the SARM team employed a multi-attention (MAT) network structure to train detection models at each stage, noticing the variation of protocols. Their approach consists of two stages: initially, training pre-trained detection models for each phase using supervised contrastive learning; subsequently, fine-tuning these pre-trained models in the second stage to generate FAS models. During training, they combine cross-entropy loss and supervised contrastive learning loss. While employing some general training strategies and the Adam optimizer for the relatively straightforward protocol P1, they adopted a label-flip augmentation strategy to generate more synthetic training data for P2.1 and P2.2. Specifically, they utilize image style transformation techniques based on the OpenCV library to convert real face images into synthetic face images, labeling these generated synthetic images as fake. This strategy significantly improves performance and reduces the domain gap between the training and testing sets.

4.8. M2-Purdue

To enhance the study in UAD, the M2-Purdue team proposed a method that blends feature extraction using CLIP with a Multi-Layer Perceptron (MLP) based classification system, namely Robust Face Attack Detection. The framework of their approach is depicted in Fig. 5.

Their methodology commences with pre-processing, which resizes all images to 224x224 pixels to standardize the input data. Following this, the team utilizes the CLIP [40] model's image encoder to extract detailed feature

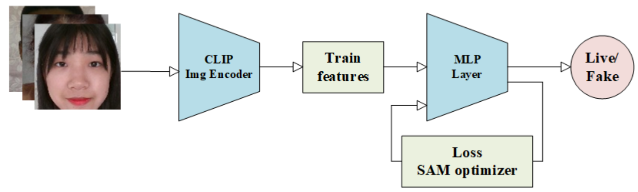

Figure 5. The overall framework of team M2-Purdue's algorithm.

representations of the visual content. Next, the MLP, composed of three layers, serves as the pivot of their classification mechanism. This MLP is fine-tuned using a composite loss function that combines the Conditional Value at Risk (CVAR) loss and the Area Under the ROC Curve (AUC) loss, weighted by a parameter λ. This fusion is described as Eq. 1:

$$\mathcal{L} = \lambda\mathcal{L}_{\text{CVAR}} + (1-\lambda)\mathcal{L}_{\text{AUC}} \tag{1}$$

where λ balances the focus between the two losses, harnessing their strengths. Also, the Adaptive Moment Estimation (Adam) optimizer handles the model's optimization, further refining the training process.

4.9. Cloud Recesses

The Cloud Recesses team has advanced a technique titled "Random Masking for Face Anti-Spoofing Detection." They proposed an innovative scheme, depicted in Fig. 6. The central premise of their method pivots on data augmentation through the deliberate obscuration of critical facial features. Initially, face landmarks are detected to inform the random masking process, intentionally veiling key areas such as the eyes, nose, and mouth. Subsequently, these modified images are channeled into the EfficientNet architecture for model training, targeting the enhancement of the model's predictive acumen on validation sets. RetinaFace is employed for face detection in data pre-processing, yielding a 256x256 cropped face image. After that, dlib is utilized to detect 68 facial landmarks, outlining essential face components. During training, random occlusion of 3-5 regions per image is executed, attenuating the impact of identifiable facial features on the training outcome.

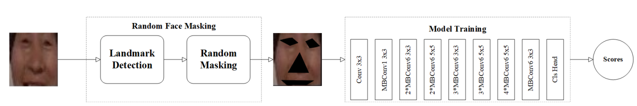

Figure 6. The overview of team Cloud Recesses's scheme. Face landmarks are first detected and used to guide random face masking. Then, the EfficientNet model is trained for classification.

4.10. ImageLab

The Image Lab team unveiled a sophisticated framework named Multiattention-Net for face anti-spoofing. Their pro-

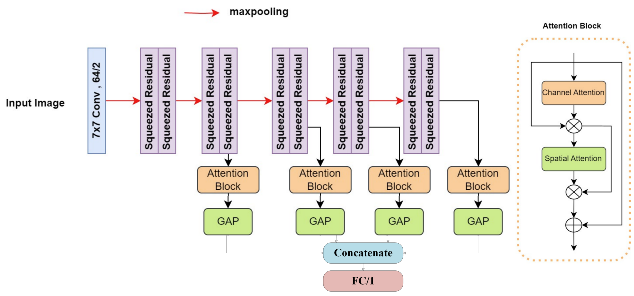

Figure 7. The overall architecture of Multiattention-Net proposed by team ImageLab.

posed solution is elucidated in Fig. 7. The Multiattention-Net initiates with a 7x7 convolutional layer to grasp complex local patterns within the input image. The network is then deepened with ten modified squeezed residual blocks that systematically downsample the information through max pooling to extract more abstract and global features. Spatial information extraction is performed at various levels of these blocks, followed by applying a dual attention block [49] that accentuates critical features within the spatial domain. A global average pooling (GAP) strategy is employed to diminish feature dimensions, and the outputs from different spatial levels are concatenated before being passed on to a fully connected layer. The training regimen is bolstered by a Binary Focal Cross entropy loss function, delicately weighted with a class balancing factor α and a focusing parameter λ, designed to tackle the challenges posed by imbalanced datasets by imposing more significant penalties on incorrect predictions, particularly for the minority class. The equation can be represented as Eq. 2:

$$\mathcal{L}(y, \hat{y}) = -\alpha \cdot (1-\hat{y})^{\gamma} \cdot \log(\hat{y}) - (1-\alpha) \cdot \hat{y}^{\gamma} \cdot \log(1-\hat{y}) \quad (2)$$

where $\mathcal{L}$ is the binary focal cross-entropy loss, y is the ground truth label (0 or 1), $\hat{y}$ is the predicted probability, α is the class balancing factor (0.25 if apply class balancing=True), γ is the focusing parameter (3 in this case)

4.11. BOVIFOCR-UFPR

The BOVIFOCR-UFPR team crafted a sophisticated approach to UAD by leveraging 3D face reconstruction and angular margin loss, as portrayed in their method overview in Fig. 8. The proposed architecture draws inspiration from 3DPC-Net [17] and pivots on an encoder-decoder setup. First, faces are detected, aligned, and cropped uniformly in the preparatory stages. The ground truth 3D point clouds corresponding to these images are initially reconstructed using a high-quality reconstruction method [15]. Then, the encoder, built upon a ResNet-50 backbone, extracts high-level features from the input RGB images. The decoder transforms these features into a 3D point cloud representing the facial structure, subsequently used to distinguish between live and spoof faces. During the training phase, the Arcface loss function is used to refine its discriminative power, and the fine-tuned Stochastic Gradient Descent optimizer collectively orchestrates the learning process. Furthermore, the Chamfer loss function, known for its $O(N^2)$ complexity, is employed alongside the Arcface loss function, which bears a complexity of $O(CD)$, with C denoting the number of classes and D the dimensionality of the encoder's feature output.

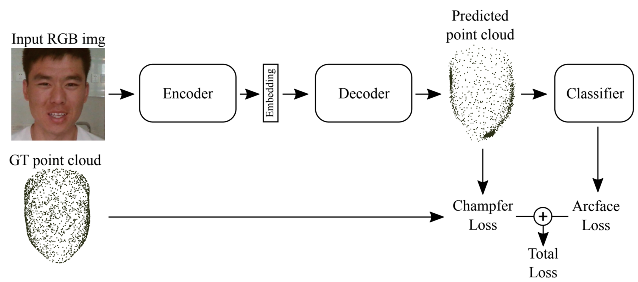

Figure 8. General architecture of BOVIFOCR-UFPR's method. The model receives pairs of RGB face images and its corresponding ground truth 3D point cloud. The predicted point cloud is also fed to a final classifier.

4.12. Inria-CENATAV-Tec

The team Inria-CENATAV-Tec refined UAD with their contribution, "MobileNetV3-spoof with hyperparameter tuning." Their methodology is encapsulated in Fig. 9, which details a flowchart of their pre-processing and UAD approach. In the pre-processing phase, the team employs ResNet-50 for landmark detection to ascertain face positioning. If the face is detected, it is aligned using the insightface template; otherwise, the image is resized and saved in its original state. The team utilizes the MobileNetV3-large-1.25 backbone for feature extraction and spoof detection, aiming to balance model complexity with performance. The model training leverages a Stochastic Gradient Descent optimizer with a multi-step learning rate strategy. Additionally, various augmentation techniques are applied to the data before normalization, aligning with each protocol's mean and standard deviation. Their work aligns with the contemporary need for lightweight models that do not compromise the precision of FAS measures.

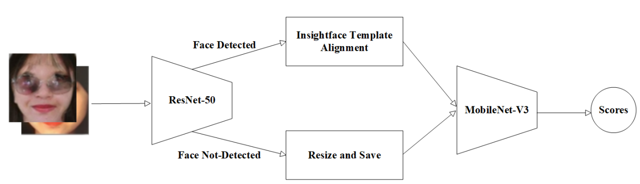

Figure 9. The approach of team Inira-CENATAV-Tec's Preprocessing and UAD stages.

4.13. Vicognit

The team Vicognit presents "FASTormer: Leveraging Vision Transformers for Face Anti-Spoofing," a novel approach utilizing the Vision Transformer architecture for FAS. Their strategy is visualized in Fig. 10, which outlines the architecture of the proposed method. During the training phase, the team meticulously tunes the hyperparameters, such as learning rate and weight decay, to ensure that the model's parameters are optimized effectively, leading to robust convergence and generalization. The Vicognit team's solution hinges on a transformer-based model, indicating their focus on harnessing transformers' capabilities to capture the intricate patterns necessary for distinguishing between bona and spoof face representations without reducing the dimensionality of the input data.

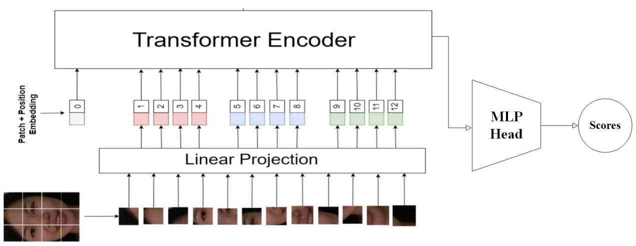

Figure 10. The approach of team Inira-CENATAV-Tec's Preprocessing and UAD stages.

5. Challenge Results

5.1. Challenge Results Report

We adopted four metrics to evaluate the performance of the solutions: APCER, NPCER, ACER, and AUC. Please note that although we report performance for various evaluation measures, the leading metric is ACER. See Tab. 3, which lists the results and ranking of the top 13 teams; we can draw three conclusions: (1) The ACER performance of the top 3 teams was significantly higher than the other teams. (2) The first-rank team achieved the best results in ACER, AUC, and BPCER, while the fifth-rank team got the best APCER result. (3) The top 5 teams are from the industry, which indicates that developing UAD algorithms in real-world applications is crucial. (4) There is a considerable variation among teams concerning the ACER scores, inferring the desperation of the in-depth research on UAD in the Face Anti-Spoofing area.

5.2. Competition Summary and Future Work

Through the challenge, we summarize the practical ideas for UAD: (1) For backbone networks, larger models with more parameters often perform better in complex real-world applications. (2) At the data level, pre-processing methods are crucial to enriching data variety, which helps avert model overfitting, while category balance enhances algorithmic stability. (3) Learning discriminative features from incomplete faces would be an effective path to UAD. In the following work, we will further improve the performance in the following aspects: (1) We will explore more efficient approaches by applying recent VLMs like CLIP [40] to guide the training process of UAD. (2) We will continue to create a more complete UAD dataset with more complete attack types and high-quality images. (3) We will try to develop a more advanced and authorized protocol for UAD tasks.

Rank	Team	ACER(%)	ACPER(%)	BPCER(%)	AUC(%)
1	MTFace	**2.3396**	0.9259	**3.7533**	**99.6923**
2	SeaRecluse	3.4369	0.3999	6.4737	96.5631
3	duileduile	5.5111	5.5185	5.5037	98.6830
4	BSP-Idiap	16.2263	9.3630	23.0698	96.3351
5	VAI-Face	17.1324	**0.2593**	34.0055	87.7566
6	L&L&W	23.6949	11.8889	35.5009	81.0384
7	SARM	27.1958	0.5037	53.8879	98.3233
8	M2-Purdue	33.4651	0.8593	66.0709	86.4003
9	Cloud Reccesses	34.5701	1.8741	67.2661	71.6810
10	ImageLab	34.9434	5.6148	64.2717	76.6426
11	BOVIFOCR-UFPR	35.3620	6.8444	63.8795	73.7304
12	Inria-CENATAV-Tec	37.3436	7.5259	67.1612	74.2011
13	Vicognit	52.1031	80.0000	24.2062	46.7025

Table 3. Team and results are listed in the final ranking of this challenge.

6. Conclusion

We organized the ***Unified Physical-Digital Face Attack Detection Challenge at CVPR2024*** based on the UniAttackData dataset and running on the CodaLab platform. 133 teams registered for the competition, and 13 made it to the final stage and submitted their codes. In the final stage of the competition, the codes are verified and re-produced by the organizers, and the reproduced results are used for the final rankings. We first present the associated dataset, protocols, and evaluation metrics. Then, we review the solutions of the participating ranked teams and report the results of the final phase. Finally, we summarize the conclusions related to the challenges and point out effective methods for Unified Attack Detection by this challenge.

Acknowledgments

This work was supported by the National Key Research and Development Program of China under Grant 2021YFE0205700, Beijing Natural Science Foundation JQ23016, the Science and Technology Development Fund of Macau Project 0123/2022/A3, and 0070/2020/AMJ, the Chinese National Natural Science Foundation Project 62276254, U23B2054, and the InnoHK program.

References

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

[2] 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. 2

[3] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG)*, pages 1–7. IEEE, 2012. 2

[4] Debayan Deb, Xiaoming Liu, and Anil K Jain. Unified detection of digital and physical face attacks. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–8. IEEE, 2023. 2

[5] Xiaoyi Dong, Jianmin Bao, Dongdong Chen, Ting Zhang, Weiming Zhang, Nenghai Yu, Dong Chen, Fang Wen, and Baining Guo. Protecting celebrities from deepfake with identity consistency transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9468–9478, 2022. 1

[6] Hao Fang, Ajian Liu, Jun Wan, Sergio Escalera, Hugo Jair Escalante, and Zhen Lei. Surveillance face presentation attack detection challenge. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 6361–6371. IEEE, 2023. 3

[7] Hao Fang, Ajian Liu, Jun Wan, Sergio Escalera, Chenxu Zhao, Xu Zhang, Stan Z Li, and Zhen Lei. Surveillance face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 2023. 2

[8] Hao Fang, Ajian Liu, Haocheng Yuan, Junze Zheng, Dingheng Zeng, Yanhong Liu, Jiankang Deng, Sergio Escalera, Xiaoming Liu, Jun Wan, et al. Unified physical-digital face attack detection. *arXiv preprint arXiv:2401.17699*, 2024. 2, 3

[9] Joel Frank, Thorsten Eisenhofer, Lea Schönherr, Asja Fischer, Dorothea Kolossa, and Thorsten Holz. Leveraging frequency analysis for deep fake image recognition. In *International conference on machine learning*, pages 3247–3258. PMLR, 2020. 1

[10] Anjith George, Zohreh Mostaani, David Geissenbuhler, Olegs Nikisins, André Anjos, and Sébastien Marcel. Biometric face presentation attack detection with multi-channel convolutional neural network. *IEEE transactions on information forensics and security*, 15:42–55, 2019. 1

[11] Guillaume Heusch, Anjith George, David Geissbühler, Zohreh Mostaani, and Sébastien Marcel. Deep models and shortwave infrared information to detect face presentation attacks. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(4):399–409, 2020. 1

[12] Fa-Ting Hong, Longhao Zhang, Li Shen, and Dan Xu. Depth-aware generative adversarial network for talking head video generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3397–3406, 2022. 1

[13] Hsin-Ping Huang, Deqing Sun, Yaojie Liu, Wen-Sheng Chu, Taihong Xiao, Jinwei Yuan, Hartwig Adam, and Ming-Hsuan Yang. Adaptive transformers for robust few-shot cross-domain face anti-spoofing. In *European Conference on Computer Vision*, pages 37–54. Springer, 2022. 1

[14] Marek Kowalski. Faceswap github. *Faceswap github",[online] Available: https://github.com/MarekKowalski/FaceSwap*, 2018. 1

[15] Biwen Lei, Jianqiang Ren, Mengyang Feng, Miaomiao Cui, and Xuansong Xie. A hierarchical representation network for accurate and detailed face reconstruction from in-the-wild images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 394–403, 2023. 7

[16] Haoliang Li, Wen Li, Hong Cao, Shiqi Wang, Feiyue Huang, and Alex C Kot. Unsupervised domain adaptation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 13(7):1794–1809, 2018. 1

[17] Xuan Li, Jun Wan, Yi Jin, Ajian Liu, Guodong Guo, and Stan Z Li. 3dpc-net: 3d point cloud network for face anti-spoofing. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2020. 1, 7

[18] Ajian Liu and Yanyan Liang. Ma-vit: Modality-agnostic vision transformers for face anti-spoofing. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 1180–1186, 2022. 1

[19] Ajian Liu, Zichang Tan, Xuan Li, Jun Wan, Sergio Escalera, Guodong Guo, and Stan Z Li. Static and dynamic fusion for multi-modal cross-ethnicity face anti-spoofing. *arXiv preprint arXiv:1912.02340*, 2019. 1

[20] Ajian Liu, Jun Wan, Sergio Escalera, Hugo Jair Escalante, Zichang Tan, Qi Yuan, Kai Wang, Chi Lin, Guodong Guo, Isabelle Guyon, et al. Multi-modal face anti-spoofing attack detection challenge at cvpr2019. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 0–0, 2019. 2

[21] Ajian Liu, Xuan Li, Jun Wan, Yanyan Liang, Sergio Escalera, Hugo Jair Escalante, Meysam Madadi, Yi Jin, Zhuoyuan Wu, Xiaogang Yu, et al. Cross-ethnicity face anti-spoofing recognition challenge: A review. *IET Biometrics*, 10(1):24–43, 2021. 3

[22] Ajian Liu, Zichang Tan, Jun Wan, Sergio Escalera, Guodong Guo, and Stan Z Li. Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1179–1187, 2021. 2

[23] Ajian Liu, Zichang Tan, Jun Wan, Yanyan Liang, Zhen Lei, Guodong Guo, and Stan Z Li. Face anti-spoofing via adversarial cross-modality translation. *IEEE Transactions on Information Forensics and Security*, 16:2759–2772, 2021. 1

[24] Ajian Liu, Chenxu Zhao, Zitong Yu, Anyang Su, Xing Liu, Zijian Kong, Jun Wan, Sergio Escalera, Hugo Jair Escalante, Zhen Lei, et al. 3d high-fidelity mask face presentation attack detection challenge. In *Proceedings of the IEEE/CVF Inter-*

national Conference on Computer Vision Workshops, pages 814–823, 2021. 3

[25] Ajian Liu, Jun Wan, Ning Jiang, Hongbin Wang, and Yanyan Liang. Disentangling facial pose and appearance information for face anti-spoofing. In *2022 26th international conference on pattern recognition (ICPR)*, pages 4537–4543. IEEE, 2022. 1

[26] Ajian Liu, Chenxu Zhao, Zitong Yu, Jun Wan, Anyang Su, Xing Liu, Zichang Tan, Sergio Escalera, Junliang Xing, Yanyan Liang, et al. Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 17:2497–2507, 2022. 2

[27] Ajian Liu, Zichang Tan, Yanyan Liang, and Jun Wan. Attack-agnostic deep face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 6335–6344, 2023. 1

[28] Ajian Liu, Zichang Tan, Zitong Yu, Chenxu Zhao, Jun Wan, Yanyan Liang Zhen Lei, Du Zhang, Stan Z Li, and Guodong Guo. Fm-vit: Flexible modal vision transformers for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 2023. 1

[29] Ajian Liu, Shuai Xue, Jianwen Gan, Jun Wan, Yanyan Liang, Jiankang Deng, Sergio Escalera, and Zhen Lei. Cfpl-fas: Class free prompt learning for generalizable face anti-spoofing. *arXiv preprint arXiv:2403.14333*, 2024. 2

[30] Siqi Liu, Pong C Yuen, Shengping Zhang, and Guoying Zhao. 3d mask face anti-spoofing with remote photoplethysmography. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14*, pages 85–100. Springer, 2016. 1

[31] Shubao Liu, Ke-Yue Zhang, Taiping Yao, Mingwei Bi, Shouhong Ding, Jilin Li, Feiyue Huang, and Lizhuang Ma. Adaptive normalized representation learning for generalizable face anti-spoofing. In *Proceedings of the 29th ACM international conference on multimedia*, pages 1469–1477, 2021. 1

[32] Yaojie Liu and Xiaoming Liu. Spoof trace disentanglement for generic face anti-spoofing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3813–3830, 2022. 1

[33] 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. 1

[34] 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. 1

[35] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 4, 5

[36] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986, 2022. 4

[37] Cheng Luo, Qinliang Lin, Weicheng Xie, Bizhu Wu, Jinheng Xie, and Linlin Shen. Frequency-driven imperceptible adversarial attack on semantic similarity. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15315–15324, 2022. 1

[38] Ishan Manjani, Snigdha Tariyal, Mayank Vatsa, Richa Singh, and Angshul Majumdar. Detecting silicone mask-based presentation attack via deep dictionary learning. *IEEE Transactions on Information Forensics and Security*, 12(7):1713–1723, 2017. 1

[39] Maxime Oquab, Timothée Darcet, Theo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Russell Howes, Po-Yao Huang, Hu Xu, Vasu Sharma, Shang-Wen Li, Wojciech Galuba, Mike Rabbat, Mido Assran, Nicolas Ballas, Gabriel Synnaeve, Ishan Misra, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision, 2023. 5

[40] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 6, 8

[41] Felix Rosberg, Eren Erdal Aksoy, Fernando Alonso-Fernandez, and Cristofer Englund. Facedancer: Pose-and occlusion-aware high fidelity face swapping. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3454–3463, 2023. 1

[42] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1–11, 2019. 1

[43] Kaede Shiohara and Toshihiko Yamasaki. Detecting deepfakes with self-blended images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18720–18729, 2022. 4

[44] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Face2face: Real-time face capture and reenactment of rgb videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2387–2395, 2016. 1

[45] Justus Thies, Michael Zollhöfer, and Matthias Nießner. Deferred neural rendering: Image synthesis using neural textures. *Acm Transactions on Graphics (TOG)*, 38(4):1–12, 2019. 1

[46] Keyao Wang, Guosheng Zhang, Haixiao Yue, Ajian Liu, Gang Zhang, Haocheng Feng, Junyu Han, Errui Ding, and Jingdong Wang. Multi-domain incremental learning for face presentation attack detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5499–5507, 2024. 1

[47] Qiulin Wang, Lu Zhang, and Bo Li. Safa: Structure aware face animation. In *2021 International Conference on 3D Vision (3DV)*, pages 679–688. IEEE, 2021. 1

[48] Ting-Chun Wang, Arun Mallya, and Ming-Yu Liu. One-shot free-view neural talking-head synthesis for video conferencing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10039–10049, 2021. 1

[49] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. *arXiv e-prints*, pages arXiv–2003, 2020. 7

[50] Zhenda Xie, Zheng Zhang, Yue Cao, Yutong Lin, Jianmin Bao, Zhuliang Yao, Qi Dai, and Han Hu. Simmim: A simple framework for masked image modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9653–9663, 2022. 5

[51] Zitong Yu, Yunxiao Qin, Xiaobai Li, Zezheng Wang, Chenxu Zhao, Zhen Lei, and Guoying Zhao. Multi-modal face anti-spoofing based on central difference networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 650–651, 2020. 1

[52] Zitong Yu, Rizhao Cai, Zhi Li, Wenhan Yang, Jingang Shi, and Alex C Kot. Benchmarking joint face spoofing and forgery detection with visual and physiological cues. *IEEE Transactions on Dependable and Secure Computing*, 2024. 2

[53] Ke-Yue Zhang, Taiping Yao, Jian Zhang, Ying Tai, Shouhong Ding, Jilin Li, Feiyue Huang, Haichuan Song, and Lizhuang Ma. Face anti-spoofing via disentangled representation learning. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX 16*, pages 641–657. Springer, 2020. 1

[54] 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. 1

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

[56] 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. 1

[57] Hanqing Zhao, Wenbo Zhou, Dongdong Chen, Tianyi Wei, Weiming Zhang, and Nenghai Yu. Multi-attentional deepfake detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2185–2194, 2021. 1

[58] Yiqi Zhong, Xianming Liu, Deming Zhai, Junjun Jiang, and Xiangyang Ji. Shadows can be dangerous: Stealthy and effective physical-world adversarial attack by natural phenomenon. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15345–15354, 2022. 1

[59] Bojia Zi, Minghao Chang, Jingjing Chen, Xingjun Ma, and Yu-Gang Jiang. Wilddeepfake: A challenging real-world dataset for deepfake detection. In *Proceedings of the 28th ACM international conference on multimedia*, pages 2382–2390, 2020. 1