Face Liveness Detection Competition (LivDet-Face) - 2024

Lambert Igene[1] †, Afzal Hossain[1], Mohammad Zahir Uddin Chowdhury[1], Humaira Rezaie[1], Ayden Rollins[1], Jesse Dykes[1], Rahul Vijaykumar[1], Alain Komaty[2], Sébastien Marcel[2], Stephanie Schuckers[1], [1]Clarkson University, USA, [2]Idiap Research Institute, Switzerland,
†igenela@clarkson.edu

Juan E. Tapia[*3] Carlos Aravena[*3] Daniel Schulz[*3] Banafsheh Adami[*4] Nima Karimian[*4] Diogo Nunes[*5] João Marcos[*5] Nuno Gonçalves[*5] Lovro Sikošek[*6] Borut Batagelj[*6] Aleksandr Alenin[*7] Alhasan Alkhaddour[*7] Anton Pimenov[*7] Artem Tregubov[*7] Igor Avdonin[*7] Maxim Kazantsev[*7] Mikhail Pozigun[*7] Vasiliy Pryadchenko[*7] Nima Schei[*8] David Pabon[*8] Manuela Tiedemann[*9]

[3]Hochschule Darmstadt (HDA), Darmstadt, Germany, [3]IDVisionCenter (IDVC), Santiago, Chile,

[4]West Virginia University, Morgantown, USA, [5]Institute of Systems and Robotics, University of Coimbra, Portugal,

[6]Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, [7]ID R&D Inc, Barcelona, Spain, [8]Hummingbirds AI, Florida, USA, [9]DERMALOG Identification Systems GmbH, Hamburg, Germany
*Competitors

Abstract

Imagine a world where a copy of your face could trick the most advanced security systems. This isn't science fiction; it's a real challenge today. LivDet-Face is a competition that aims to advance the detection of attacks at the biometric sensor, known as Presentation Attack Detection (PAD). This international contest is a key benchmark in biometric security, offering an unbiased look at the latest innovations in face PAD and demonstrating progress over time in detecting and preventing sophisticated attacks. Through the International Joint Conference on Biometrics (IJCB) platform, LivDet-Face 2024 provides a standardized evaluation process, access to advanced Presentation Attack Instruments (PAI), and a comprehensive dataset of bona fide face images. The competition had two main categories: algorithms and systems. A total of sixteen algorithms and one system were submitted for this year's competition. Anonymous submissions topped both image and video subcategories with an ACER of 4.93% and 4.13%, respectively. In the systems category, Team Dermalog, despite being the sole submission, achieved an impressive ACER of 3.12%.

1. Introduction

Face biometric recognition systems are integral to electronic identity management and authentication, widely adopted in government and industry for their user-friendliness, convenience, and high accuracy. Despite their benefits, these systems are vulnerable to presentation attacks, such as printed photos, video replays, and face masks, which pose significant security risks [2, 7, 16, 21]. These attacks aim to conceal identities, impersonate others, or enroll fake identities into recognition systems [10].

To counter these threats, robust Presentation Attack Detection (PAD) systems are essential, employing both hardware and software solutions to distinguish between genuine and fraudulent attempts, thereby ensuring security and integrity. Presentation Attack Detection (PAD) has emerged as a crucial technique in this domain, drawing significant attention [1, 9, 18]. While numerous PAD methods have achieved impressive results on benchmark datasets [4, 5], many struggle in real-world scenarios, leaving face PAD an unresolved challenge under unconstrained conditions [3, 21].

In response to these challenges, we introduce LivDet-Face, a pioneering international competition within the

LivDet series, aimed at advancing face liveness detection technologies. This competition provides a platform for evaluating state-of-the-art PAD algorithms against unseen face presentation attacks.

Key achievements of the LivDet-Face 2024 competition include:

- State-of-the-Art Report: An assessment of current PAD technologies based on independent testing of nine algorithms for image, seven algorithms for video, and one hardware for system category.
- Using levels defined by FIDO: A dataset prepared according to Fast ID Online (FIDO) Biometric Requirements [17] and evaluated using standard PAD metrics defined by the International Organization for Standardization (ISO) [19].
- Diverse PAI Species: The competition utilized the broadest spectrum of Presentation Attack Instruments (PAIs) to date, grouped into nine different PAIs captured with nine sensors.
- Introduction of Novel PAIs: Three novel PAIs were introduced: projection on 2D, projection on 3D, and bobblehead models of bona fide subjects.

1.1. Performance Evaluation Metrics

For LivDet-Face 2024, we maintain the evaluation metrics used in LivDet-Face 2021, following ISO/IEC 30107-3 guidelines [19]:

Attack Presentation Classification Error Rate (APCER): Measures the proportion of attack presentations (PAIs) incorrectly classified as genuine, indicating the system's susceptibility to accepting fraudulent inputs. **Bonafide Presentation Classification Error Rate (BPCER):** Quantifies the proportion of genuine presentations incorrectly classified as attacks, highlighting the system's ability to reject legitimate inputs.

The overall performance is determined using the Weighted Average of APCER ($\text{APCER}_{\text{average}}$), which considers the average APCER across all PAI categories, weighted by sample count. This metric provides a balanced assessment of algorithm performance. For competitive ranking, the Average Classification Error Rate (ACER) is calculated as the average of $\text{APCER}_{\text{average}}$ and BPCER. Although ACER has been deprecated in recent ISO/IEC guidelines for industry-related PAD evaluations, it remains relevant in LivDet-Face 2024 for comparative assessment of participant algorithms.

2. Comparison of Face Spoofing Challenges in the Past 5 Years

In recent years, significant advancements have been made in PAD algorithm performance through the development of various PAIs for training and testing. Below are notable challenges that have contributed to this progress and how LivDet-Face 2024 distinguishes itself.

2.1. CelebA-Spoof Challenge (2020)

The CelebA-Spoof Challenge utilized a large-scale dataset of over 600,000 images featuring print, replay, and mask attacks to advance face anti-spoofing technology [23]. While effective for image-based spoofing detection, it primarily focused on controlled environments.

2.2. CVPR Face Anti-Spoofing Challenge (2021-2023)

The CVPR Face Anti-Spoofing Challenge provided comprehensive datasets covering various attack types annually, pushing the boundaries of spoof detection in terms of accuracy and generalizability using standard metrics like APCER and BPCER. These datasets included attacks such as print, replay, and 3D mask attacks, offering a broad spectrum for evaluating the robustness of face anti-spoofing algorithms [22].

Comparison with LivDet-Face 2024: LivDet-Face 2024 offers a unique dataset with videos captured using two cameras from five different smartphones under varied lighting conditions, providing a more rigorous evaluation compared to the relatively controlled environments of the CVPR challenges.

2.3. Wild Face Anti-Spoofing Challenge (2023)

This challenge focused on evaluating algorithms in uncontrolled environments with diverse spoofing attacks, aiming to benchmark performance under real-world conditions [20].

Comparison with LivDet-Face 2024: While the Wild Face Anti-Spoofing Challenge emphasized real-world variability, LivDet-Face 2024 innovates further by introducing novel attack types like projection on 2D surfaces, projection on 3D masks, and bobblehead attacks, along with realistic testing conditions using multiple smartphones.

LivDet-Face 2024 stands out with its comprehensive dataset, novel attack types, and realistic testing conditions that address current challenges and pave the way for future innovations in face PAD.

3. Competition Design

3.1. LivDet-Face 2024

LivDet-Face 2024, co-organized by Clarkson University (USA) and the Idiap Research Institute (Switzerland) [6, 11], is the second edition of the LivDet competition focused on face PAD [15]. It assesses state-of-the-art facial PAD algorithms and systems against both traditional and novel PAIs. The competition features two main categories:

Algorithm and System, with Algorithm further divided into Image and Video subcategories.

The competition includes nine PAI types for both subcategories: bobblehead, half cloth mask, high-quality 3D mask, low-quality 3D mask, print attack, projection attack 2D, projection attack 3D, silicon mask, and replay attack of bona fide subjects. The overall test samples and two specific PAI types (high-quality 3D masks and video displays of bona fide subjects) were not disclosed to the competitors.

Algorithm performance was measured using an output score ranging from 0 to 100, with a threshold set at 50. Scores below 50 were classified as PAIs, while scores of 50 and above were classified as bona fide presentations. A score of -1000, however, was used to indicate undetected samples or other types of failures, as specified in the competition's instructions [14]. During PAI testing, a score of -1000 was considered a correct rejection of PAIs and did not contribute to the attack presentation classification error. Conversely, a score of -1000 for bona fide sample testing was considered incorrect and was included in the BPCER calculation.

All evaluations were conducted by the competition organizers, ensuring unbiased results.

4. Experimental Protocol

LivDet-Face 2024 competition welcomed participation from international academic and industrial organizations, both anonymously and non-anonymously. Non-anonymous competitors were included as co-authors in the publication. Ten teams registered globally, submitting nine entries for the image category, seven for the video category, and one system submission. All submissions were successfully tested.

4.1. Training Dataset

No official training dataset was provided for LivDet-Face 2024. Participants were encouraged to use any available data, whether public or proprietary, to train their algorithms. To aid familiarization, the organizers shared a few examples of known PAIs. However, the remaining samples of the disclosed PAI types were kept unknown to competitors.

4.2. Test Dataset

The LivDet-Face 2024 competition utilized a diverse dataset from Clarkson University and Idiap Research Institute, comprising 17,512 images (2,235 bona fide and 15,277 PAI samples) and 16,291 videos (2,225 bona fide and 14,066 PAI samples). The data was collected using nine sensors (DSLR, iPhone 6s/12/X, Samsung Galaxy S9, Google Pixel, Redmi 6pro/9A, Basler aA1920-150uc) from 121 subjects. Video lengths for testing were up to 5 seconds. PAIs were grouped into eight categories for both images and videos with a summary in Table 1.

4.2.1 Test Dataset Details

Paper Display: The dataset includes 100 low-quality and 100 high-quality paper images and videos from 25 subjects, captured using four sensors.

Laptop Display: 100 samples for both images and videos from 25 subjects using four sensors.

2D Photo Mask: 100 samples each for both images and videos, involving eye cut-outs from photo paper face images. These samples were collected from 25 subjects using four sensors.

3D Mask: low, medium, and high-quality masks created using 3D printed models from front and side photos of subjects [8, 12, 15]. It features 24 low-quality, 12 medium-quality, and 12 high-quality samples for both images and videos, all captured using four sensors.

Silicon Mask: 141 image and video samples of silicon masks, were collected using five sensors.

Bobblehead: Includes 3D models created from single images, processed into bobbleheads, with 90 samples collected from 5 subjects using three sensors.

White Resin and Filament 3D Masks: 15 image and 15 video samples were collected from white resin and filament 3D masks. These masks were generated using at least 40 images and were printed with resin and filament materials, produced at the Clarkson University Makerspace workshop using a Prusa MK3 3D printer for filament and a FormLabs printer for resin.

Projection Attack 3D: 150 image and 78 video samples were collected by projecting single-face images onto white resin and filament 3D masks using a projector under various lighting conditions.

Print Attacks: 4,665 video recordings of face photos printed on matte and glossy paper using laser and ink-jet printers.

Replay Mobile Attacks: 11,200 bona fide face videos replayed to smartphone cameras, with one phone used for replaying and another for recording.

Replay TV Attacks: 5,600 bona fide face videos replayed on a TV screen under different lighting conditions.

Projector Attacks: 2,800 video recordings of bona fide face videos projected under various conditions using white and green screens.

Bona Fide Face Videos: 8,400 videos of real faces, without masks, from 70 subjects recorded over two sessions. Each video is 10 seconds long and was captured using the selfie camera of five smartphones, simulating scenarios similar to those during the COVID-19 pandemic.

This comprehensive dataset with sample images shown in Figure 1, provides a realistic and varied testing environment to ensure robust algorithm performance, making it an essential tool for evaluating face PAD systems.

Figure 1. Example images of presentation attack types present in the LivDet-Face 2024 test dataset. Top (left to right): paper, bobblehead, projection on a 3D mask, 3D high-quality mask, 3D low-quality mask, projection on 2D, silicon mask, and half-cloth mask.

Table 1. Test Dataset Summary

Class	Types of PAIs	Total Images	Total Videos	Sensors
Bona fide	-	2235	2225	DSLR, iPhone 6s:12:X, Redmi 9A:6 Pro, Pixel, S9
PAI	BOBBLEHEAD (BH)	90	15	DSLR, iPhone X, S9
PAI	HALF CLOTH (HC)	759	714	DSLR, iPhone 6s:12:X, Redmi 9A:6 Pro, Pixel, S9
PAI	HQ 3D MASK (HQ)	111	33	DSLR, iPhone X, Pixel, S9
PAI	LQ 3D MASK (LQ)	73	72	DSLR, iPhone X, Pixel, S9
PAI	PRINT ATTACK (PP)	3472	3091	DSLR, iPhone 6s:12:X, Redmi 9A:6 Pro, Pixel, S9
PAI	PROJECTION ATTACK 2D (2D-PA)	1400	1400	iPhone 6s:12:X, Redmi 9A:6 Pro, Pixel, S9
PAI	PROJECTION ATTACK 3D (3D-PA)	721	90	DSLR, iPhone X, S9
PAI	REPLAY ATTACK (RA)	8510	8510	DSLR, iPhone 6s:12:X, Redmi 9A:6 Pro, Pixel, S9
PAI	SILICON MASK (SM)	141	141	DSLR, iPhone X, S9, Pixel and Basler aA1920-150uc

5. LivDet-Face 2024 Competition Algorithms

A total of eight teams from five countries participated, submitting nine algorithms for image, seven algorithms for video, and one in the system category. The competition saw teams HDA-IDVC, WVU, Aeminium, UNLJ-FRI-FE, IDLiveFace, Hummingbirds.ai, DERMALOG, and one anonymous submission. All competitors were given the option to present their results anonymously. Each team was also invited to submit a description of their algorithms, which are detailed below.

5.1. IDLiveFace

Team IDLiveFace submitted their algorithm exclusively for the image category of the competition. The algorithm adopts a passive, single-image liveness detection system utilizing an ensemble of EfficientNet and transformer models. This solution is designed to accurately differentiate between genuine and spoofed face images with high throughput and low latency. The transformers (Vision Transformers) focus more attention on specific patches during classification, effectively concentrating on crucial areas that might indicate spoofing. In contrast, EfficientNet models use broader features across the images, allowing for a comprehensive yet detailed analysis of facial features. This varied focus is essential for addressing sophisticated spoofing attempts, such as high-fidelity masks and 4K screen replays, which pose significant challenges in liveness detection. Each model within the ensemble employs distinct image preprocessing methods to ensure focus on different aspects of the images. This diversity in preprocessing enhances the overall detection capability by covering a broader range of potential spoofing indicators. The models were trained on a diverse dataset, encompassing various ethnic backgrounds and attack vectors, to prepare the system for real-world applications. The ensemble is carefully configured to maintain a small size and reduce computational demands, enabling high throughput without compromising detection accuracy.

5.2. Hummingbirds.ai

Team Hummingbirds.ai submitted solutions for both image and video categories using their DepthFusion Specular-Diffuse Attention Network (DFSD-AttentionNet). This model leverages both RGB and depth information to enhance image processing from multiple perspectives. The integration of depth information improves the separation of flat surface features, aiding in scene geometry understanding.

DFSD-AttentionNet employs a specular-diffuse separa-

tion mechanism, distinguishing between direct specular reflections and scattered diffuse reflections. This separation aids in capturing surface property nuances. Convolutional layers in the separation module learn distinct filters for each reflection type, while attention modules dynamically weigh feature importance, thereby enhancing the model's capability.

A pre-trained ConvNext model, adapted for depth images, serves as the depth extractor. This transfer learning approach utilizes pre-learned features from large datasets, thereby enhancing performance and generalization. The final layers integrate features from both RGB and depth inputs to produce the output.

DFSD-AttentionNet combines advanced deep learning techniques with multi-modal inputs, demonstrating potential in visual data understanding. As a preliminary experiment, it offers insights for further research and development, paving the way for sophisticated systems to tackle complex image processing challenges.

5.3. HDA-IDVC

For this challenge, the HDA/IDVC Team submitted solutions for both image and video categories. In their solution, they proposed an algorithm leveraging the SwinTransformer (ST) architecture with a multi-class linear classifier as the final stage for attack detection [13]. The algorithm addresses five distinct attack classes along with the bona fide class. Input RGB images are preprocessed by normalizing using ImageNet transforms and resized to 256x256 pixels. The model is fine-tuned from ImageNet 1K weights over 200 epochs, optimizing for the best Equal Error Rate (EER) metric on a validation set. Softmax activation is employed on the linear classifier's output to obtain the bona fide class score. The training data encompass a blend of proprietary datasets and publicly available datasets, including OULU-NPU, CASIA, and Replay-Mobile.

5.4. UNLJ-FRI-FE

Team UNLJ-FRI-FE submitted their algorithm for the image category, utilizing a prediction algorithm based on a SE-ResNeXt model variant, pre-trained on ImageNet-12k. The model was selected due to its performance on ImageNet. For the training data, the team chose the 3DMAD dataset, which contains mask attack and bona fide videos from 17 subjects, with 15 videos per subject, captured using Microsoft Kinect, each containing around 300 frames. This dataset offers 76,500 images, split into non-overlapping training and testing sets (54,000 and 22,500 images, respectively). Although the dataset includes texture and depth data, only RGB data was utilized. As such, the frames from the videos were used as training data.

First, an off-the-shelf MTCNN model for face detection was used on each image to crop out the face, and the images were then resized. The ADAM optimizer was employed along with Binary Cross Entropy loss. The initial learning rate was set relatively low (1e-5) and was halved every five epochs to prevent overfitting. The model was then trained for 25 epochs, with performance monitored on a set of 37 example images provided by the organizers.

5.5. Aeminium

The developed method aims to be a security-focused PAD system, comprising several subsystems, each tasked with a specific PAD function. The PAD system includes the following subsystems: (1) Suspect Context Detector; (2) Print Attack Detector; (3) Replay Attack Detector; (4) Moiré Pattern Detector; and (5) Mask Attack Detector.

Although each subsystem is trained for a particular task, it demonstrates some level of discrimination against other non-trained PAD tasks. Consequently, the final solution is designed to leverage the discrimination capabilities of each trained PAD subsystem for both expected and unseen scenarios. As a result, the final system can be characterized as an ensemble model with a high detection capability for both seen and unseen attack scenarios. To achieve this, a maximum score fusion strategy was employed. Additionally, as a frame-based solution, the final video score is obtained by averaging the scores of all frames.

One of the main disadvantages of this system lies in its sensitivity to attack detection, primarily due to the chosen fusion strategy, which results in a high BPCER. For future work, novel fusion strategies will be explored to maintain the system's ability to detect unseen attacks while reducing the overall BPCER.

5.6. WVU

Video Liveness Detection: The proposed method utilizes a novel architecture that combines a 3D Convolutional Neural Network (3D CNN) with the Swin Transformer to efficiently capture spatial and temporal features. Frames extracted from videos are processed by the 3D CNN to create a spatiotemporal rPPG (ST-rPPG) block, which captures physiological signals. This block is then fed into the Swin Transformer, which partitions the input into windows and performs shifted window operations to enhance feature extraction across both spatial and temporal dimensions. During training, rPPG signals are sampled from different locations within the same video. For liveness detection, the Heart Rate (HR) is calculated from the Power Spectral Density (PSD) of the signal. An HR between 40-100 bpm indicates a bona fide video, while values outside this range suggest spoofing. The trained model processes new videos, analyzing cues to output a liveness score, which is thresholded for a binary decision.

Image Liveness Detection: Team WVU developed a deep learning approach using autoencoders combined with

the VGG19 architecture. The pre-trained VGG19 network serves as the encoder, extracting high-level features from input images, which are then compressed into a lower-dimensional latent space. The decoder reconstructs the original image from this compressed representation using transposed convolutional layers. Training involves minimizing the reconstruction loss (mean squared error) between input images and their reconstructions, using only bona fide face images. Post-training, the autoencoder accurately reconstructs genuine faces and produces higher reconstruction errors for spoofed images. Reconstruction errors are computed for each input image, with a threshold set to classify images as bona fide (below threshold) or spoofed (above threshold). This method leverages VGG19's feature extraction and focuses on reconstruction errors to distinguish bona fide from spoofed faces without requiring extensive labeled datasets.

5.7. DERMALOG (System):

Team Dermalog submitted the only hardware system for the system category. The FLC1 camera integrates an Intel RealSense (aligned RGB and depth image) with a thermal sensor to continuously capture images, track faces, and evaluate their Presentation Attack Detection (PAD) score.

The core PAD algorithm processes a 96x96 normalized thermal image through a MobileNet-V2 model with a width multiplier of 0.35 to generate a single PAD score. The final PAD score is calculated as the average of eight single-shot PAD scores, improving overall performance and handling outliers. Additionally, any face height below 10 centimeters is classified as an attack.

The thermal image is normalized using a similarity transformation based on facial landmarks, followed by value range adjustment, where the image's mean value is shifted to zero and divided by half the min-max distance.

A PAD score is calculated only if the face is fully in view of all sensors, looking into the camera, within 1.5 meters, and not occluded.

6. Results and Discussion

This section discusses the performance of the competing algorithms and systems, evaluated using APCER for each PAI category and BPCER for bona fide samples at a threshold of 50, as announced before the competition. A summary of error rates for both image and video subcategories is provided in Table 2 and Table 3, and performance comparisons are shown in Figure 2a,b.

LivDet-Face 2024 Image Category: Team Anonymous emerged as the winner with the lowest ACER of 4.93%, followed by Team IDLIVEFACE with 5.53% and Team WVU at 9.90%. The winning team's algorithm achieved the lowest BPCER of 1.34%. Team Anonymous performed well with low-quality 3D masks, bobbleheads, and silicon masks but had an average APCER of 8.52%, indicating lower performance against sophisticated spoofs like replay attacks and 2D projection attacks. Team IDLIVEFACE detected all bobbleheads, half cloth, silicon masks, and both high and low-quality 3D masks with nearly 0% APCER, except for print (2.42%) and replay attacks (0.01%). However, their BPCER was higher at 10.60%.

High APCER values for "Projection Attack 2D" and "Print Attack" suggest these attack types are particularly challenging and require focused research to improve detection capabilities. High BPCER values indicate many algorithms struggle with accurately identifying bona fide samples, this could also be due to a face processing problem such as face detection failure, quality checks, etc., that results in a -1000 score result for some samples. This leads to false rejections, so training with more diverse and representative bona fide data is needed to improve genuine user classification. Advanced machine learning techniques and diverse training datasets can enhance robustness and reliability across different attack scenarios.

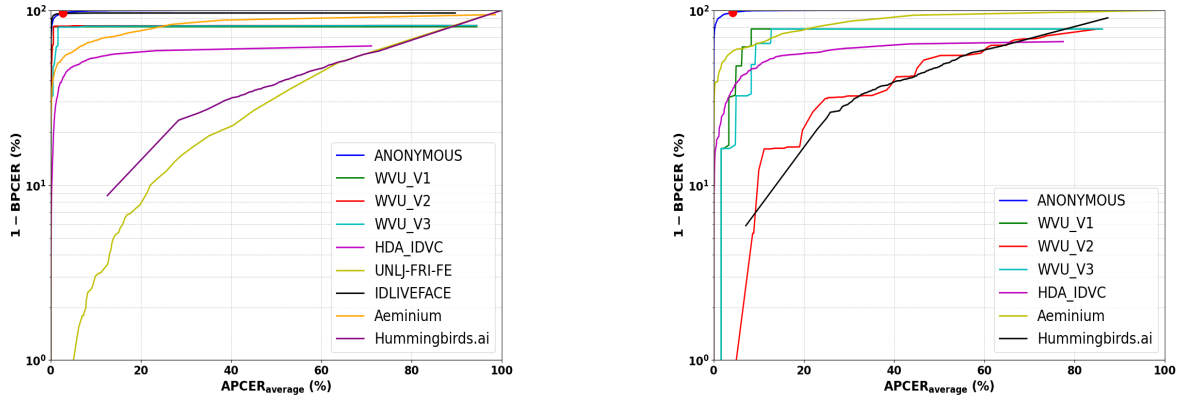
LivDet-Face 2024 Video Category: Team Anonymous also won the video category with an ACER of 4.13%, followed by Team WVU algorithms 1 and 3, both with 11.92%. Team Anonymous had the lowest BPCER of 2.16%. Team Aeminium achieved third place with an ACER of 25.89% and the second-best avgAPCER of 2.42%. Teams like HDA-IDVC and WVU showed avgAPCERs of 2.24% and 2.01%, respectively, but struggled with bona fide samples, with BPCERs as high as 61.53% and 49.35%. The top competitors performed better against low-quality PAIs than high-quality ones. Team WVU excelled with half-cloth, high-quality, and low-quality 3D masks (APCER = 0%) but struggled with projection attack 3D (APCER = 35.90%).

The challenges with "Projection Attack 2D" and "Print Attack" remain consistent across both image and video categories, highlighting the need for focused research to develop more effective detection mechanisms. High BPCER values across multiple datasets suggest that video algorithms, like image algorithms, need improvement in accurately identifying bona fide samples. Enhancing algorithm robustness and consistency across different attack scenarios is crucial. Advanced machine learning techniques, including deep learning models, can help improve generalization capabilities, making them more reliable and effective in real-world applications. While algorithms like IDLIVEFACE show promising results in both image and video categories, significant work remains to address high BPCER values and improve detection capabilities for challenging attack types. Focusing on these areas can substantially enhance the reliability and effectiveness of these algorithms in practical applications.

LivDet-Face 2024 System Category: Team Dermalog's

Table 2. Face PAD Competition Summary: Image Category

Competitor Name	Presentation Attack Instruments Level Types									Overall Performance		
	Level A		Level B		Level C					APCER$_{avg}$%	BPCER%	ACER%
	Print Attack		Display Attacks		3D Face Masks							
	HC	PP	RA	2D-PA	BH	3D-PA	LQ	HQ	Silicon			
Anonymous	2.47	12.72	3.61	35.07	0	5.02	0	3.60	0	8.52	1.34	4.93
HDA-IDVC	12.65	25.25	1.50	1.43	0	0.28	0	0	0	1.97	62.98	32.47
IDLIVEFACE	0	2.42	0.01	0	0	0	0	0	1.42	0.46	10.60	5.53
Aeminium	8.56	1.74	2.69	4.00	0	0	0	3.57	21.28	2.78	48.23	25.50
WVU-V1	0.40	2.15	0.16	0.21	14.44	4.72	0	0.99	2.13	0.91	19.62	10.26
WVU-V2	0.25	2.29	0.25	0.36	14.44	4.85	0	0.99	2.13	0.88	18.93	9.90
WVU-V3	4.61	3.48	0.99	1.86	14.44	4.72	8.33	7.92	2.13	2.09	19.91	11.00
UNLJ-FRI-FE	28.55	6.90	3.04	0.46	10	3.17	0	9.09	0	5.17	98.27	51.72
Hummingbirds.ai	43.08	51.05	33.77	27.57	70	59.22	-	80.46	-	41.61	59.71	50.66



(a) Image Category

(b) Video Category

Figure 2. ROC curves for all nine algorithms for the image category and seven algorithms for the video category of the competition, presenting the overall performance on samples representing all nine PAIs. The overall APCER is evaluated based on (APCER$_{average}$).

Table 3. Face PAD Competition Summary: Video Category

Competitor Name	Presentation Attack Instruments Level Types									Overall Performance%		
	Level A		Level B		Level C					APCER$_{avg}$%	BPCER	ACER
	Print Attack		Display Attacks		3D Face Masks							
	HC	PP	RA	2D-PA	BH	3D-PA	LQ	HQ	Silicon			
Anonymous	1.4	10.09	1.96	26.86	0	0	0	3.03	0	6.11	2.16	4.13
HDA-IDVC	12.61	1.55	1.96	1.67	33.33	2.56	9.72	9.09	7.02	2.42	61.53	31.97
Aeminium	5.60	1.86	2.22	3.43	0	0	0	4.76	0	2.42	49.35	25.89
WVU-V1	0	4.84	1.05	0	40.00	35.90	0	0	58.62	2.01	21.83	11.92
WVU-V2	38.27	44.75	45.11	31.64	40.00	35.90	0	47.62	58.62	42.63	53.36	48.00
WVU-V3	0	4.84	1.05	0	40.00	35.90	0	0	58.62	2.09	19.91	11.00
Hummingbirds.ai	43.93	43.93	32.07	28.00	33.33	60.26	47.22	63.64	34.32	37.96	59.5	47.85

Table 4. Face PAD Competition Summary: System Category

DERMALOG (system)	APCER%	Spoof Samples	Errors	BPCER%	Bonafide Samples	Error
Bobblehead	0	5	0			
Bonafide				0	33	0
HQ Mask	30.00	10	3			
LQ Mask	18.18	11	2			
Print Paper	0	23	0			
Silicone Mask	0	5	0			
White Filament Mask	0	8	0			
White Resin Mask	0	8	0			
Full Cloth Mask	0	10	0			
Weights		80	5	BPCER%	ACER%	
APCER$_{avg}$%	6.25			0	3.12	

system submission achieved an average APCER of 6.25% and an ACER of 3.12%, as shown in Table 4. The system demonstrated excellent performance, achieving a perfect 0% APCER and BPCER for most Presentation Attack Instruments (PAIs). However, it faced difficulties with high-quality and low-quality 3D masks, which resulted in significantly higher APCERs of 30% and 18.18%, respectively. This increase in APCER is primarily attributed to the system's tendency to misclassify these PAIs as bona fide presentations when the 3D masks are warmed to human body temperature, leading to false recognition of them as genuine. This suggests that while the Dermalog system is highly effective in detecting most PAIs, it struggles with the more lifelike and detailed facial features presented by 3D masks.

7. Conclusion

The LivDet-Face 2024 competition introduced five novel PAIs (Projection Attack 2D, Projection Attack 3D, flexible 3D silicon masks, video display samples of bona fide subjects, and Print Attack) and compared state-of-the-art algorithms in image and video categories. The winning image algorithm achieved an ACER of 4.93% (APCER = 8.52%, BPCER = 1.34%), while the winning video algorithm achieved an ACER of 4.13% (APCER = 6.11%, BPCER = 2.16%).

Nine image algorithms, seven video algorithms, and one system were tested in real-world scenarios with various PAIs, attack types, environments, and sensors.

Key trends revealed that while algorithms generally detected attacks well (low APCER), they struggled with accurately identifying bona fide samples (high BPCER). Projection Attack 2D and Print Attack were particularly challenging, leading to higher error rates. Robust performance from some algorithms, like IDLIVEFACE, contrasted with considerable variability in others, highlighting the need for enhanced algorithm robustness and generalization.

Factors contributing to performance degradation included:

a) Increased complexity in test datasets due to unknown PAI types.

b) Limited access to large public datasets for new attack types.

c) Lack of standardized training datasets, leaving training choices to competitors.

d) Variability between training and test datasets in environmental factors, sensor types, and PAI quality.

e) Face processing challenges, including face detection failures and quality check issues.

These findings underscore that face PAD remains a challenging and evolving field. Significant differences in algorithm accuracy emphasize the need for large, diverse training datasets encompassing a wide range of PAIs. This competition and the benchmark dataset will help enhance the security and reliability of biometric systems. Continued research and innovation are crucial to advancing face presentation attack detection.

8. Acknowledgement

This material is based upon the work supported in part by the National Science Foundation under Grant No. 1650503, the Center for Identification Technology and Research (CITeR), and the European Union's Horizon 2020- research and innovation program, SOTERIA, under grant agreement 101018342.

References

[1] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu. Face anti-spoofing using patch and depth-based cnns. In *2017 IEEE international joint conference on biometrics (IJCB)*, pages 319–328. IEEE, 2017. 1

[2] J. Bigun, H. Fronthaler, and K. Kollreider. Assuring liveness in biometric identity authentication by real-time face tracking. In *Proceedings of the 2004 IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, 2004. CIHSPS 2004.*, pages 104–111. IEEE, 2004. 1

[3] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face anti-spoofing based on color texture analysis. In *2015 IEEE international conference on image processing (ICIP)*, pages 2636–2640. IEEE, 2015. 1

[4] M. M. Chakka, A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, et al. Competition on counter measures to 2-d facial spoofing attacks. In *2011 international joint conference on biometrics (IJCB)*, pages 1–6. IEEE, 2011. 1

[5] I. Chingovska, J. Yang, Z. Lei, D. Yi, S. Z. Li, O. Kahm, C. Glaser, N. Damer, A. Kuijper, A. Nouak, et al. The 2nd competition on counter measures to 2d face spoofing attacks. In *2013 international conference on biometrics (ICB)*, pages 1–6. IEEE, 2013. 1

[6] Clarkson University (CITeR). Co-organizing team. http://livdet.org/. 2

[7] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel. Can face anti-spoofing countermeasures work in a real world scenario? In *2013 international conference on biometrics (ICB)*, pages 1–8. IEEE, 2013. 1

[8] FaceGen Modeller. available at: http://facegen.com/modeller.html, 2021. 3

[9] J. Galbally, S. Marcel, and J. Fierrez. Biometric antispoofing methods: A survey in face recognition. *Ieee Access*, 2:1530–1552, 2014. 1

[10] A. Husseis, J. Liu-Jimenez, I. Goicoechea-Telleria, and R. Sanchez-Reillo. A survey in presentation attack and presentation attack detection. In *2019 International Carnahan Conference on Security Technology (ICCST)*, pages 1–13. IEEE, 2019. 1

[11] I. R. Institute. Co-organizing team. https://www.idiap.ch/en/, 2024. Accessed: 2024-08-14. 2

[12] A. S. Jackson, A. Bulat, V. Argyriou, and G. Tzimiropoulos. Large pose 3d face reconstruction from a single image via direct volumetric cnn regression. In *Proceedings of the IEEE international conference on computer vision*, pages 1031–1039, 2017. 3

[13] 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 *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9992–10002, 2021. 5

[14] LivDet-Face 2024 . competition's instructions. https://face2024.livdet.org/. 3

[15] S. Purnapatra, N. Smalt, K. Bahmani, P. Das, D. Yambay, A. Mohammadi, A. George, T. Bourlai, S. Marcel, S. Schuckers, et al. Face liveness detection competition (livdet-face)-2021. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2021. 2, 3

[16] R. Ramachandra and C. Busch. Presentation attack detection methods for face recognition systems: A comprehensive survey. *ACM Computing Surveys (CSUR)*, 50(1):1–37, 2017. 1

[17] S. Schuckers, G. Cannon, E. Tabassi, M. Karlsson, and E. Newton. Fido biometrics requirements. *Population*, 5(2-1):2–3, 2019. 2

[18] S. A. Schuckers. Spoofing and anti-spoofing measures. *Information Security technical report*, 7(4):56–62, 2002. 1

[19] I. G. Tellería. Evaluation of presentation attack detection under the context of common criteria. *Universidad Carlos III de Madrid*, 2019. 2

[20] D. Wang, J. Guo, Q. Shao, H. He, Z. Chen, C. Xiao, A. Liu, S. Escalera, H. J. Escalante, Z. Lei, et al. Wild face anti-spoofing challenge 2023: Benchmark and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6380–6391, 2023. 2

[21] D. Wen, H. Han, and A. K. Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015. 1

[22] S. Zhang, X. Wang, A. Liu, C. Zhao, J. Wan, S. Escalera, H. Shi, Z. Wang, and S. 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 (CVPR)*, June 2019. 2

[23] Y. Zhang, Z. Yin, J. Shao, Z. Liu, S. Yang, Y. Xiong, W. Xia, Y. Xu, M. Luo, J. Liu, et al. Celeba-spoof challenge 2020 on face anti-spoofing: Methods and results. *arXiv preprint arXiv:2102.12642*, 2021. 2