First Competition on Presentation Attack Detection on ID Card

Juan E. Tapia[†1], Naser Damer[†2,3], Christoph Busch[†1], Juan M. Espin[†4], Javier Barrachina[†4], Alvaro S. Rocamora [†4], Krištof Ocvirk [5,*], Leon Alessio [5,*], Borut Batagelj[5,*], Sushrut Patwardhan[6,*], Raghavendra Ramachandra[*,6], Raghavendra Mudgalgundurao [*,6], Kiran Raja[*,6], Daniel Schulz [*,7], Carlos Aravena [*,7]

[1]Hochschule Darmstadt (h_da), da/sec-Biometrics and Internet Security Research, Germany.
[2]Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany,
[3]Department of Computer Science, TU Darmstadt, Darmstadt, Germany,
[4]Facephi company, Spain,
[5] Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia,
[6] Norwegian University of Science and Technology (NTNU), Gjøvik, Norway,
[7] ID VisionCenter (IDVC), Santiago, Chile
[†]Organizer [*]Competitors,

Abstract

This paper summarises the Competition on Presentation Attack Detection on ID Cards (PAD-IDCard) held at the 2024 International Joint Conference on Biometrics (IJCB 2024). The competition attracted a total of ten registered teams, both from academia and industry. In the end, the participating teams submitted five valid submissions, with eight models to be evaluated by the organisers. The competition presented an independent assessment of current state-of-the-art algorithms. Today, no independent evaluation on cross-dataset is available; therefore, this work determined the state-of-the-art on ID cards. To reach this goal, a sequestered test set and baseline algorithms were used to evaluate and compare all the proposals. The sequestered test dataset contains ID cards from four different countries. In summary, a team that chose to be "Anonymous" reached the best average ranking results of 74.80%, followed very closely by the "IDVC" team with 77.65%.

1. Introduction

The accelerated evolution in consumer smartphone cameras and the COVID-19 pandemic has increased the interest in remote biometric verification systems. The capacity to reach the customer remotely for services such as e-commerce, digital banking, and general fintech requires robust systems for remote identity verification. One approach for this verification is using a picture of an official identity document, such as a national ID card, and comparing the data with a frontal face photograph (selfie) of the person in question, both captured remotely by the user "in the wild" condition.

Remote identity verification processes can encounter attacks in which a user's identity is impersonated. Today, the simplest and most common attacks are presentation attacks. In a presentation attack, the attacker impersonates some of the captured samples, official documents, or selfies. With the intention to mitigate said problem, face presentation attack detection (face PAD) is a widely studied field. However, document attack detection is a new field and difficult to access due to the privacy of identity documents.

Nowadays, presentation attacks on ID cards, such as printing a photo or displaying it on a screen, are widespread. Also, the number of images available for training and testing is limited due to privacy concerns. As a result, many solutions trained on small datasets overfit intra-dataset conditions because the train and test set stem from the same source, limiting the generalisation capabilities.

PAD-IDCard 2024 is the first competition in the ID card series. It offers **(a)** an independent assessment of current state-of-the-art ID Card Presentation Attack Detection algorithms and **(b)** an evaluation protocol, including real printed-out and screen replay attacks and bona fide ID card images. Researchers can follow the evaluation protocol after the competition is closed to benchmark their solutions with PAD-IDCard winners and baselines. Today, there exists no independent evaluation where the approaches are evaluated in a cross-dataset scenario; hence, it is still being determined which methods, if any, perform well under such more realistic test conditions.

The rest of the article is organised as follows: Section 2 summarises the related works of PAD in ID cards. Section 3 describes the datasets and depicts examples of images. Sections 4 and 5 describe the submission and evaluation process. The metrics used for the evaluation and the proposed methods are described in Sections 5 and 6, respectively. The experiments and results are presented in Section 7. Lastly, Section 9 summarises the findings of the competition and discusses potential future work.

2. Related work

In recent years, several works have proposed PAD methods to detect presentation attacks where ID cards are used to circumvent the security of remote verification systems [22, 24].

Berenguel et al. [3] developed an application to classify ID documents forged by a scan-printing operation. Their application allows the capture of Spanish ID documents using a mobile device and the assessment of their validity. The counterfeit detection module apply texture descriptors, principal component analysis, and feature pooling to classify regions of interest using linear Support Vector Machines (SVM). The final decision of labelling a document as genuine or counterfeit is performed by a Naïve Bayes classifier.

Gonzalez et al. [10] presented a two-stage method for detecting tampered ID cards, which was trained and evaluated on a dataset with real Chilean national ID cards. The proposed method uses a pre-trained MobileNet model[12] to detect borderlines in the photo ID zone caused by composite tampering, while a second lightweight CNN, termed "BasicNet", was trained from scratch to detect the physical source of the document.

Mudgalgundurao et al. [18] proposed to adapt a pixel-wise supervision model in [8] that is used, along with a binary classification objective, to train presentation attack detectors on an in-house dataset of German ID cards and residence permits. The proposed system uses a simplified DenseNet [13] architecture, which the authors compare against baseline face PAD approaches.

Chen et al. [6] employed a scheme based on Siamese networks for document recapture detection. The network is trained on triplets of patches extracted from bona fide, recaptured, reference documents. A custom "forensics loss" aimed at attracting genuine and reference representations while repelling recaptured and reference representations. The authenticity of a questioned document is evaluated using the distance metrics from three triplets. The authors created a synthetic university student ID card dataset to test their system.

Benalcazar and Tapia et al. [2] explored the effectiveness of computer vision algorithms and generative models for data augmentation while training fraud detection networks. The authors propose populating templates with synthetic data to create additional bona fide presentations and training a StyleGAN-ADA network to generate synthetic bona fide samples from scratch.

Magge et al. [15] explored the application of the Meijering filter [17] for detecting recaptured identity documents. The authors created a dataset of recaptured images based on the publicly available BID [23] dataset and used it to train an SVM classifier on the raw histogram data obtained using the filter. Although their system does not compare well with approaches that utilise neural networks, it remains an attractive alternative due to being transparent and explainable.

Most of the aforementioned studies trained and tested their proposed systems on private datasets using bona fide presentations of ID cards obtained from Government entities, company services, and banks. As such, it is difficult to scrutinise and improve upon these systems since the data can not be distributed publicly due to privacy concerns.

Currently, open-set datasets like MIDV 500 [1], MIDV 2019 [4], MIDV 2020 [5] and DLC 2021 [20], despite offering a rich amount of country representations and document types, fall short due to their limited number of unique user identities and few examples of bona fide and screen displays attacks on ID cards. Conversely, private datasets are not available to compare the results. These fundamental limitations undermine the potential of PAD ID card models to accurately learn and generalise across the wide variability inherent in ID cards.

As a starting point, the quantity of unique user data is crucial for teaching models to discern between bona fide and fake documents. It is a complicated task because of the minor subject base of these datasets. Table 1 summarises the most relevant datasets in this field and can be used as a starting point to train a PAD system.

Table 1. Summary ID card datasets available in the State of the art.

Author	Datasets	Images	User	Comments
Soares et al. [23]	BID-Data	28,800	8	Synthetic data No Genuine data
Mudgalgundurao et al. [18].	Private	104,882	86	Genuine ID card
González et al.[10]	Private	54,980	5,000	Genuine ID card
González et al.[9]	Private	190,000	16,000	Genuine ID card
Benalcazar et al.[2]	Private	38,477	9,286	Genuine ID card
Markham et al. [16]	Open-set	500	50	Generated from templates - Transfer style
Arzalov et al. [1] Bulatov et al.	Open-set MIDV-500 [1] MIDV-2020 [5]	500 Videos	50	No Genuine ID card
Polevoy et al. [20]	Open-set DLC2021	1000	1000	No Genuine ID card Generated from templates
Koliaskina et al. [14]	Open-set MID Holo	700 Video		No Genuine ID card - Utopia ID card 300 holographic - 400 videos
Park et al. [19]	Open-Set KID-2K	34,662	82	No Genuine ID card For 46 people who do not exist **

Given the significant limitations of available public datasets for studying PAD card applications, a private dataset for training and testing was created for this competition from digital users' documents to develop a baseline evaluation. The dataset was generated in-house by h_da and the Spanish company Facephi.

Figure 1. Example of images used to validate the ID card PAD model. Left to right: Bona fide, Composite, Printed and Screen.

3. Datasets

No training datasets were provided to the participants for the competition. Each team used any available training dataset, such as open-set, private, or synthetic datasets, for research and commercial purposes.

For the development of the baseline method, we created a private training dataset consisting of ID cards from four countries: Spain (ESP), Chile (CHL), Argentina (ARG), and Costa Rica (CRI). This set contains bona fide images, which represent pictures captured directly from a genuine ID card. The composite attack represents an ID card image modified by swapping the face or the text area between two ID cards. The composite attacks were created manually and automatically based on data augmentation techniques. The print attacks were created using an ID card printed out on glossy paper and PVC cards with different resolutions. The screen attacks were made by capturing ID card displays on various screens such as tablets, smartphones, and laptops.

The test set partition was sequestered for all the participants, and it consisted of ID cards from four different countries: Chile (CHL), Guatemala (GUA), Panama (PAN), and Mexico (MEX). For each country bona fide and attack images are contained. All the attacks were created manually and automatically, following the same conditions as those previously defined for the private training dataset used for the baseline methods. These images also present different qualities, which means visual artefacts in the area of face photos and high-quality images without any visual artefacts. The printed attacks were created from glossy paper and PVC cards. The screen images were captured from several sources and resolutions, such as tablets, smartphones, and laptop screens. Some of the ID card images are ICAO compliant, and others are not. This competition did not test injection attacks.

One small set of 4 images was provided to all the teams that submitted a model in order to validate the results. This set contains one ID card image of each type: bona fide, composite, print, and screen ID card image, as shown in Figure 1.

Tables 2, 3, and 4 describe the details of each group of images used for the experiments 1, 2, and 3 explained in Section 6.

4. Submission process

All the participants submitted a link (e.g., a link to Google Drive or Dropbox) with a folder with the compiled model in a "conda environment" with the file "enviroment.yml" for Python 3.8, where it indicates all the libraries necessary to run their model. This model accepted as input a "CSV file" with the path input images as a "filename". The output from the model was another CSV file with three columns, "filename", "score", and "class", as Presentation Attack Detection (PAD) for all the "test" samples. The value 0 means "bona fide", and the value 1 means "attacks". The evaluation Python file provided should accept two parameters: The path to the evaluation_list.csv (–evaluation) and the output path of the scores (–output).

5. Evaluation Criteria

The detection performance of biometric PAD algorithms is standardised by ISO/IEC 30107-3[1]. The most relevant metrics for this study are Attack Presentation Classification Error Rate (APCER), Bona fide Presentation Classification Error Rate (BPCER), and BPCER_{AP}. Those metrics determine the error rates when classifying an instance between bona fide and the different Presentation Attack Instrument Species (PAIS).

The APCER metric measures the percentage of attack presentations incorrectly classified as bona fide for each different PAIS. The worst-case scenario is considered when evaluating an entire system. The computation method is detailed in Equation 1, where the value of N_{PAIS} corresponds to the number of attack presentation images, RES_i is 1 if the ith image is classified as an attack, or 0 if it was classified as a bona fide presentation according to a predefined threshold.

$$APCER_{PAIS} = 1 - \frac{1}{N_{PAIS}} \sum_{i=1}^{N_{PAIS}} RES_i \qquad (1)$$

On the other hand, the BPCER metric measures the proportion of bona fide presentations wrongly classified as attacks. The BPCER can be computed using Equation 2, where N_{BF} is the amount of bona fide presentation images, and RES_i takes the same values described earlier for the

[1] https://www.iso.org/standard/79520.html

APCER metric. The two metrics determine the system's performance and are subject to a specific operation point.

$$BPCER = \frac{\sum_{i=1}^{N_{BF}} RES_i}{N_{BF}} \quad (2)$$

Finally, $BPCER_{AP}$ and the Equal Error Rate (EER) are used to analyse the PAD system's performance for a specific operating point. The latter is the operating point where APCER and BPCER are equal. This operating point corresponds to the intersection with the diagonal line in a Detection Error Trade-off (DET) curve, which is also reported for all the experiments. On the other hand, the $BPCER_{AP}$ is the BPCER value when the APCER is $100/AP$. In this work, we use: $BPCER_{10}$, $BPCER_{20}$ and $BPCER_{100}$, which correspond to APCER values of 10%, 5% and 1%, respectively.

An average ranking determines the winning team. A weighting factor was selected to increase the metric's contribution in the most challenging operational points, such as $BPCER_{100}$. The team with the lowest AV_{Rank} won the competition. This metric weighted the $BPCER_{10,20,100}$ as follows:

$$AV_{rank} = BPCER_{10} \times 0.2 + BPCER_{20} \times 0.3 + BPCER_{100} \times 0.5 \quad (3)$$

6. Methods

6.1. Baseline description

Three baselines were defined, according to the available datasets, to explore the real conditions and possible scenarios of today's state of the art:

Baseline 1: Training with a private dataset (bona fide and attack) and evaluating with the sequestered test dataset.

Baseline 2: Training with only bona fide from private datasets plus only attacks from open-set datasets (MIDV-500 and MIDV-Holo) and evaluating with the sequestered test dataset.

Baseline 3: Training with a mixture of private and open-set datasets such as MIDV-500 and MID-Holo, evaluating with the sequestered test dataset.

The preprocessing pipeline for the images input into our networks begins with the segmentation and alignment of the ID card present in each input image, following the method proposed in [16]. This initial step ensures that the ID card is correctly isolated and positioned, providing a consistent starting point for further processing. The images are reshaped to 384×384 for both experiments. Once the ID card is properly segmented, aligned, and reshaped, a series of augmentations are applied to the dataset to improve the robustness of the models. These augmentations include horizontal flipping to simulate variations in orientation and adjustments in brightness and contrast to account for different lighting conditions. Additionally, random JPEG compression is applied to introduce variations in image quality, mimicking real-world scenarios where image compression artefacts may be present. Using a random Gaussian filter helps simulating different levels of blur, while random hue adjustments and grey-scale conversions further diversify the dataset by altering the colour properties of the images. These preprocessing steps collectively ensure that the models are trained on a wide variety of image conditions, enhancing their ability to generalise and perform accurately across different scenarios.

For this competition, three different architectures were evaluated for each baseline experiment:

MobileVITv2[2], Efficientv2-S[3], and MobileNetv3-Large[4]. In all of them, the ImageNet weights were used to initialise and train all the models. All the baseline models were trained with three classes: bona fide, composite, print plus PVC, and screen. The bona fide images (64,933) are common in the three baselines. The MobileVITv2 reached the best performance and is used to compare with the competitors in the rest of the work.

A sequestered dataset with a total of 21,000 images was created for evaluation with ID cards from four countries and four PAIs. Table 2 provides an overview of the dataset used for Baseline 1, which considers only private datasets for training.

Table 2. Baseline 1 dataset - Training/Validation using a private dataset

	Train	Val	Test	Total
Bona fide	64,933	11,458	5,000	81,391
Composite	90,101	15.900	5,000	111,001
Print	30,398	5,364	6,000	41,762
Screen	82,232	14,511	5,000	101,743
Total	267,664	47,233	21,000	335,897

Table 3. Baseline 2 dataset - Training/Validation on bona fide presentations from a private dataset and attacks from public datasets

	Train	Val	Test	Total
Bona fide	64,933	11,458	5,000	81,391
Composite	9,283	1,638	5,000	15,921
Print	26,623	4,698	6,000	37,321
Screen	19,608	3,460	5,000	28,068
Total	120,447	21,254	21,000	162,701

Table 3 provides an overview of the dataset used for

[2] https : / / github . com / leondgarse / keras _ cv _ attention_models/tree/main

[3] https://github.com/google/automl/tree/master/efficientnetv2

[4] https : / / github . com / kuan - wang / pytorch - mobilenet-v3

Baseline 2, which considers only open-set datasets for training plus private bona fide presentations. This experiment considers MIDV-500 and MIDV-Holo.

Table 4 provides an overview of the dataset used for Baseline 3, which considers private datasets (267,664 images) plus open-set datasets for training (55,514 images). This experiment considers MIDV-500 and MIDV-Holo.

Table 4. Baseline 3 dataset - Training/Validation on a mixture of private and public datasets

	Train	Val	Test	Total
Bona fide	64,933	11,458	5,000	81,391
Composite	99,384	17,538	5,000	121,922
Print	57,021	10,062	6,000	73,083
Screen	101,840	17,971	5,000	124,811
Total	323,178	57,029	21,000	401,207

6.2. Submission and Team proposals

Ten teams were registered for the competition. However, five different teams have submitted their models for evaluation. In total, 11 models were evaluated, 8 submitted by the teams, plus three baselines. Each team described its own proposed method.

Team Asmodeus This team from NTNU proposed two models, AsmodeusV1 and AsmodeusV2, based on Dynamic Snake convolution (DSC) [21] for detecting high-frequency artefacts in added-in print photos, screen photos, and during GAN synthesis. DSC uses a deformed kernel to learn high-frequency noise. The model consists of sequential DSC modules with a fully connected layer at the end for classification.

Team "Anonymous" This team that chose to be Anonymous, proposed two models for the Document Presentation Attack Detection (DocPAD) based on the MobilenetV3-large Convolutional Neural Network (CNN) [11]. This CNN was trained to detect artefacts commonly found in display and print attacks. The model was trained using academic datasets such as MIDV, DLC, and internal data developed by their QA team. The Anonymous team was leveraging its capabilities. The system aims to accurately identify and prevent presentation attacks, ensuring the integrity and reliability of identity verification processes.

Team FRIFE This team from the University of Ljubljana developed a model for detecting composite attacks. The team decided to develop a line detection algorithm based on traditional computer vision techniques with the goal of detecting visible lines in unusual places on the ID cards. For recapture attacks, they trained a single Xception-based model [7] to detect both screen and print-out recaptures.

Team IDVC-PAD-IDCARD For this challenge, the IDVisionCenter (IDVC) company team proposed models called IDVC_V1 and IDVC_V2. The two models are based on a pipeline composed of an ID-Card detection followed by a PAD algorithm. For the ID-Card detector, they trained a network based on the YOLO algorithm, using international open-set datasets and also their private dataset, using a resolution of 416×416 pixels. Then, for the PAD algorithm, they trained a network based on MobileNet[11], using four distinct attack classes along with the bona fide class. The resolution for the PAD detection algorithm is 224×224 pixels. They manually created three presentation attack instruments: Printed, Replay (Screen), and Composite. They also created a presentation attack instrument automatically, consisting of swapping the face image for different ID cards. The difference between IDVC_V1 and IDVC_V2 is that V2 includes open-set datasets in the training stage.

Team Secure-ID This second team from NTNU as well proposed a method based on the MIDV-500 dataset that served as the foundation for the competition. To generate synthetic ID cards, the GitHub repository [5] was utilised. The Canon TS-5000 printer was employed for print attacks, utilising papers of various qualities along with colour and grayscale images for detection purposes. Two Android smartphones and two monitors under different backgrounds and lighting settings were used to create replay attacks. A pixel-wise classification method is proposed to detect presentation attacks of the printed and digitally replayed attacks. The approach to using pixel-wise supervision is to leverage minute cues on various artefacts, such as Moiré patterns and artefacts left by the printers [25].

7. Experiment and Results

This challenge involved two kinds of evaluations. The first evaluation addresses the three baseline methods, which were fine tuned on the private dataset after being pre-trained network as described in Section 6.1. The second evaluation compares all submissions from all teams.

The sequestered test set is common for all the submissions. Thus, each team trained its own model, but all were evaluated on the same test set. The test dataset is composed of ID cards from 4 different countries.

All the submissions were evaluated as a binary model to determine the winning team, which means bona fide versus attacks. As complementary information, the sequestered test dataset was evaluated separately by countries, i.e. Chile, Guatemala, Panama, and Mexico.

Figure 2 shows the DET curves for the baselines 1, 2, and 3. The black line considers bona fide presentations versus all the attacks (i.e., composite, print, and screen). Each plot also includes the analysis of each PAIS isolated for research purposes. The green curve shows the composite attack, the red represents the printed attack, and the blue represents a

[5] https://github.com/Oriolrt/SIDTD_Dataset

screen attack. A single analysis by baselines 1, 2, and 3 are depicted in Figures 3, 4 and 5.

Further, the green curve shows that the composite attack reached a lower error rate for baseline 1, which means training in private datasets and testing in sequestered datasets. For baselines 2 and 3, the PAIS shows similar results for each one.

Figure 6, 7, 8, and 9 show the DET curve for the Anonymous, FRIFE, IDVC_V2, and Secure-ID teams, respectively, for all countries evaluated together, followed by the single evaluation for Chile, Guatemala, Panama, and Mexico subsets.

In summary, the best results were reached by the Anonymous_V1 team submission with an AV_{Rank} of 74.30% for the bona fide versus attack of all the countries together. The Asmodeus team reached the lowest results because both submitted models always reported the same score, i.e. 0.86 and 0.99 for Asmodeus_V1 and Asmodeus_V2, respectively. Both models presented by this team can not generalise well to different attacks reaching a higher AV_{Rank}.

Table 5 shows all the submitted models' summary results evaluated based on Average Rank. The last column show the overall rank.

Table 5. Summary submission results.

Team Name	EER (%)	BPCER10 (%)	BPCER20 (%)	BPCER100 (%)	Average Rank (eq. 3) (%)	Rank
PAD-IDCard 2024 Competing Algorithms						
Anonymous_V1	21.87	46.06	65.82	90.70	**74.30**	1
Anonymous_V2	29.01	63.36	76.82	92.22	81.82	4
Asmodeus_V1	N/A[6]	N/A	N/A	N/A	N/A	7
Asmodeus_V2	N/A[7]	N/A	N/A	N/A	N/A	7
FRIFE	44.09	87.96	93.06	99.92	95.47	5
IDVC_V1	22.96	65.40	74.60	84.38	77.65	2
IDVC_V2	25.91	66.10	74.42	86.16	78.62	3
SecureID	50.63	90.94	95.42	99.42	96.52	6
PAD-IDCard 2024 Baseline Algorithms						
Baseline1	4.58	1.84	4.20	14.96	9.10	-
Baseline2	7.17	5.26	9.78	24.40	16.18	-
Baseline3	9.02	8.14	13.28	28.58	19.90	-

As we mentioned before, the sequestered dataset was also analysed by country, separated by Chile, Guatemala, Panama, and Mexico.

- For the Chilean ID card, the IDVC_V2 team achieved the best results by far, with a lower Average rank of 4.34%. The screen attack was identified as the most challenging PAI.
- For the Guatemala ID card, the IDVC_V1 team reached the best results with an average rank of 41.25%.
- For the Panama ID card, the Anonymous team reached the best results, with an average rank of 37.58%.
- For Mexico, the Anonymous team reached the best results with an average rank of 76.94%.

[6]The Asmodeus_V1 model always delivers the score 0.86

[7]The Asmodeus_V2 model always delivers the score 0.99

Table 6. Summary results on **Chile**

Team Name	EER (%)	BPCER10 (%)	BPCER20 (%)	BPCER100 (%)	Average Rank (%)
Anonymous_V1	16.00	29.80	47.00	79.30	59.71
Anonymous_V2	16.31	28.10	44.50	67.60	52.77
Asmodeus_V1	N/A	N/A	N/A	N/A	N/A
Asmodeus_V2	N/A	N/A	N/A	N/A	N/A
FRIFE	38.13	83.80	93.60	100	94.84
IDVC_V1	4.90	1.30	4.70	14.80	9.07
IDVC_V2	3.00	0.60	1.90	7.30	**4.34**
Secure-ID	42.46	83.10	90.70	98.80	93.23
Baseline1	0.70	0.01	0.01	0.50	0.25
Baseline2	3.10	0.70	1.90	6.00	3.71
Baseline3	3.98	1.50	3.20	8.70	5.61

Table 7. Summary results on **Guatemala**

Team Name	EER (%)	BPCER10 (%)	BPCER20 (%)	BPCER100 (%)	Average Rank (%)
Anonymous_V1	15.30	22.10	33.90	62.80	45.99
Anonymous_V2	24.88	54.50	69.01	88.90	73.05
Asmodeus_V1	N/A	N/A	N/A	N/A	N/A
Asmodeus_V2	N/A	N/A	N/A	N/A	N/A
FRIFE	47.88	93.30	97.30	100	97.85
IDVC_V1	6.38	2.10	9.60	75.90	**41.25**
IDVC_V2	15.88	22.30	34.00	77.30	53.31
SecureID	50.75	92.00	95.40	99.20	96.62
Baseline1	2.50	0.30	1.00	5.10	2.91
Baseline2	7.90	6.80	10.10	20.20	14.49
Baseline3	13.69	15.40	21.50	32.90	25.98

Table 8. Summary results on **Panama**

Team Name	EER (%)	BPCER10 (%)	BPCER20 (%)	BPCER100 (%)	Average Rank (%)
Anonymous_V1	13.06	16.20	27.80	52.00	**37.58**
Anonymous_V2	15.90	27.20	47.80	85.50	62.53
Asmodeus_V1	N/A	N/A	N/A	N/A	N/A
Asmodeus_V2	N/A	N/A	N/A	N/A	N/A
FRIFE	41.68	92.00	94.60	99.99	96.77
IDVC_V1	10.78	13.40	44.50	86.20	59.13
IDVC_V2	18.78	30.60	40.50	69.00	52.77
SecureID	53.65	93.90	96.90	99.60	97.47
Baseline1	7.20	6.40	10.40	20.10	14.45
Baseline2	12.98	15.80	22.70	32.80	26.37
Baseline3	11.78	13.10	18.00	27.50	21.77

It is essential to highlight that all the submission models reached the highest error (average rank) compared with baseline 1 trained with the MobileVIT model.

Tables 6, 7, 8, and 9 show all the submissions evaluated in a single country as complementary information for Chile, Guatemala, Panama, and Mexico respectively. For each table, the best results are shown in bold.

For the Asmodeus team, estimating the EER, BPCER_{10}, BPCER_{20}, and BPCER_{100} values was not possible because the models always delivered the same score values (0.86 and 0.99). These scores were also checked using validation scores directly with the team.

Table 9. Summary results on **Mexico**

Team Name	EER (%)	BPCER10 (%)	BPCER20 (%)	BPCER100 (%)	Average Rank (%)
Anonymous_V1	22.48	51.80	71.10	90.50	**76.94**
Anonymous_V2	29.88	64.30	72.90	88.20	78.83
Asmodeus_V1	N/A	N/A	N/A	N/A	N/A
Asmodeus_V2	N/A	N/A	N/A	N/A	N/A
FRIFE	46.86	86.40	94.40	99.40	95.3
IDVC_V1	24.70	65.90	83.30	97.80	87.07
IDVC_V2	30.28	73.40	87.10	96.50	89.06
SecureID	57.08	96.10	98.40	99.90	98.69
Baseline1	5.80	2.20	6.70	27.30	16.10
Baseline2	2.38	0.50	0.70	5.70	3.16
Baseline3	4.10	0.80	3.00	20.50	11.31

8. Analysis

The competition results show that the generalisation capabilities to predict PAD between different countries and attacks are still challenging. According to our analysis, the number of images available for training in open datasets limits the performance of the approaches. The open-set datasets present fewer bona fide images per subject or use a printed PVC ID card to simulate a genuine image. Further on, the same ID card is used to create many attacks. As a result, imbalanced datasets are obtained. This factor confuses the classifier with the print and screen attack, obtaining a high EER. Conversely, the teams that used private datasets based on ID cards with many different subjects and reduced the number of images per subject in the datasets obtained the best results. The screen attack (blue curve) is identified as the most challenging in the baseline and for the team which achieved the best result. The high resolution of different screens available in the market makes this attack very hard to detect.

The different security factors present on ID cards, such as holograms, watermarks, and others, were created based on a physical inspection using active factors and lights. Thus, these factors are not a challenge that is easily reproducible in a PAD system based on one image in remote systems.

We can also identify that ID cards in countries based on ICAO compliance standards are more accessible to detect and classify, such as Chile and Panama. The standardised position of the face photo, letter sizes, and other factors support the learning process. Conversely, Guatemala and Mexico ID cards do not follow the ICAO standards and present a lot of variability in photos, illumination, and where the different information is positioned. A general and agnostic system is still a challenge, as demonstrated in the test evaluation results for specific countries.

9. Conclusion

The results from this competition indicate that ID card PAD is still far away from fully solving this research challenge. Significant differences in accuracy among baseline algorithms, which were trained with different data considering private and open-set datasets, stress the importance of access to extensive and diversified training datasets encompassing a large number of PAIs.

As a future work, we will propose a new version of the competition based on synthetic ID cards to reduce the lack of bona fide images and attacks. In order to measure the generalisation capabilities, new approaches based on meta-learning approaches, such as zero-shot and few-shot learning, are suggested to improve the generalisation capabilities of countries not included in the training set.

This competition and the benchmark will contribute to our efforts as a biometric community to win the PAD arms race.

Acknowledgements

This competition was sponsored by Facephi, R&D area. Further, this work was supported by the European Union's Horizon 2020 research and innovation program under grant agreements 883356 (iMARS) and 101121280 (EINSTEIN), and the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[1] V. V. Arlazarov, K. B. Bulatov, and T. S. Chernov. MIDV-500: A dataset for identity documents analysis and recognition on mobile devices in video stream. *CoRR*, abs/1807.05786, 2018.

[2] D. Benalcazar, J. E. Tapia, S. Gonzalez, and C. Busch. Synthetic id card image generation for improving presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 18:1814–1824, 2023.

[3] A. Berenguel, O. R. Terrades, J. Lladós, and C. Canero. e-counterfeit: a mobile-server platform for document counterfeit detection. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 9, pages 15–20. IEEE, 2017.

[4] K. Bulatov, D. Matalov, and V. V. Arlazarov. Midv-2019: challenges of the modern mobile-based document ocr. In W. Osten and D. P. Nikolaev, editors, *Twelfth International Conference on Machine Vision (ICMV 2019)*. SPIE, 2020.

[5] K. B. Bulatov, E. Emelianova, D. V. Tropin, N. S. Skoryukina, Y. S. Chernyshova, A. V. Sheshkus, S. A. Usilin, Z. Ming, J.-C. Burie, M. M. Luqman, and V. V. Arlazarov. Midv-2020: A comprehensive benchmark dataset for identity document analysis. *ArXiv*, abs/2107.00396, 2021.

[6] C. Chen, S. Zhang, F. Lan, and J. Huang. Domain-agnostic document authentication against practical recapturing attacks. *IEEE Transactions on Information Forensics and Security*, 17:2890–2905, 2022.

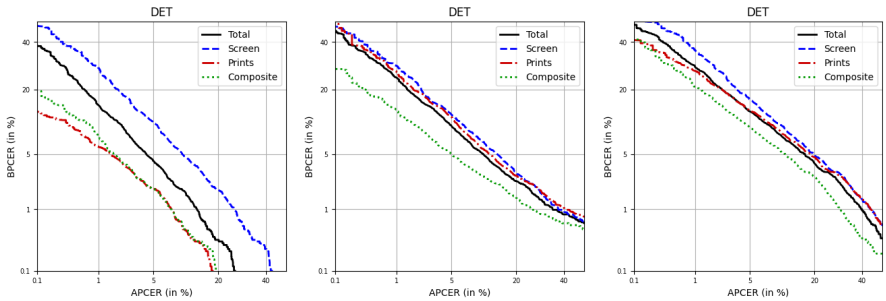

Figure 2. DET Baseline results for MobileVIT models. The left to right plots show Baselines 1, 2, and 3 results, respectively. The black line represents the binary results of bona fide versus attacks.

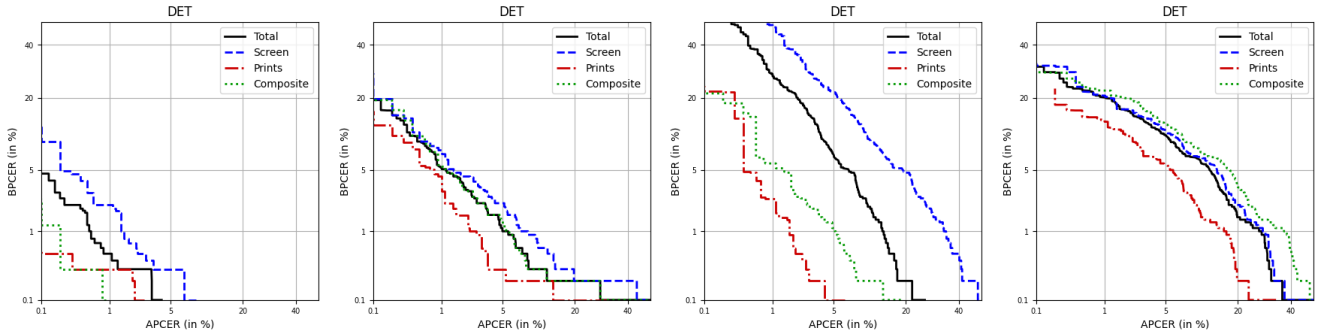

Figure 3. DET Baseline results for MobileVIT models. The left to right plots show **Baseline 1** for 4 different ID card countries, Chile, Guatemala, Panama, and Mexico subsets. The black line represents the binary results of bona fide versus attacks.

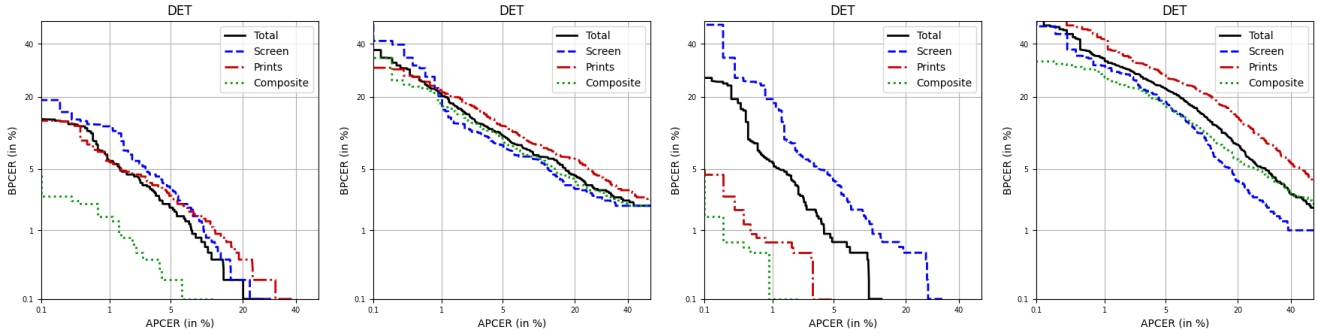

Figure 4. DET Baseline results for MobileVIT models. The left to right plots show **Baseline 2** for 4 different ID card countries, Chile, Guatemala, Panama, and Mexico subsets. The black line represents the binary results of bona fide versus attacks.

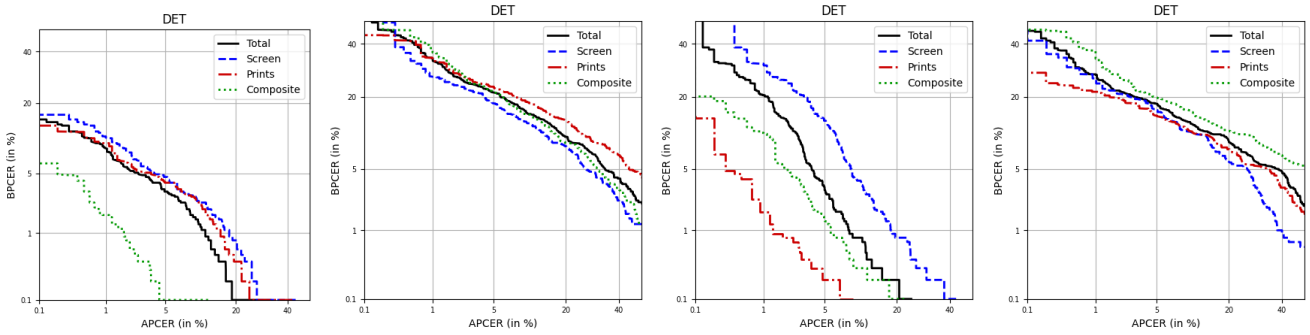

Figure 5. DET Baseline results for MobileVIT models. The left to right plots show **Baseline 3** for 4 different ID card countries, Guatemala, Panama, and Mexico subsets. The black line represents the binary results of bona fide versus attacks.

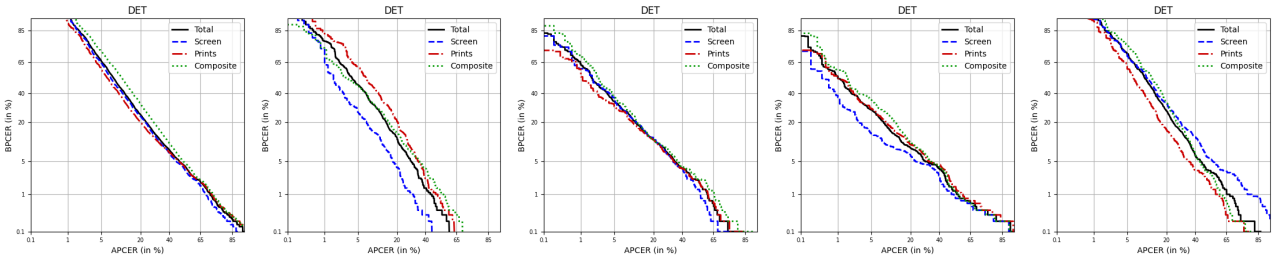

Figure 6. DET result for **Anonymous_V1** model for test set. The left-to-right plots show all countries together, followed by Chile, Guatemala, Panama, and Mexico subsets. The black line represents the binary results of bona fide versus attacks.

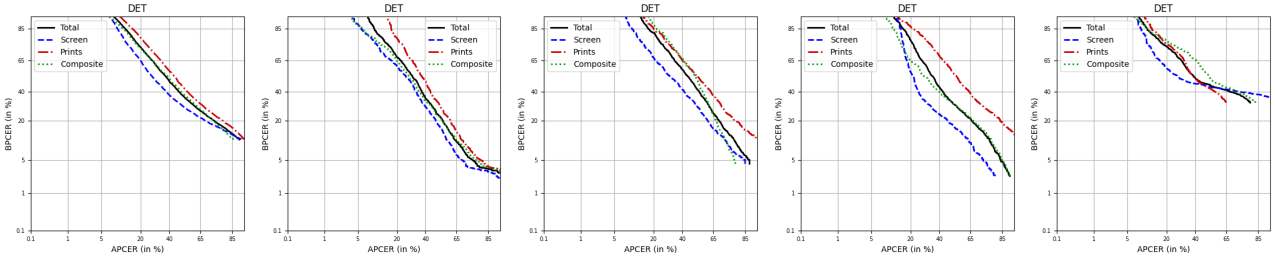

Figure 7. DET result for **FRIFE** model for the test set. The left-to-right plots show all countries together, followed by Chile, Guatemala, Panama, and Mexico subsets. The black line represents the binary results of bona fide versus attacks.

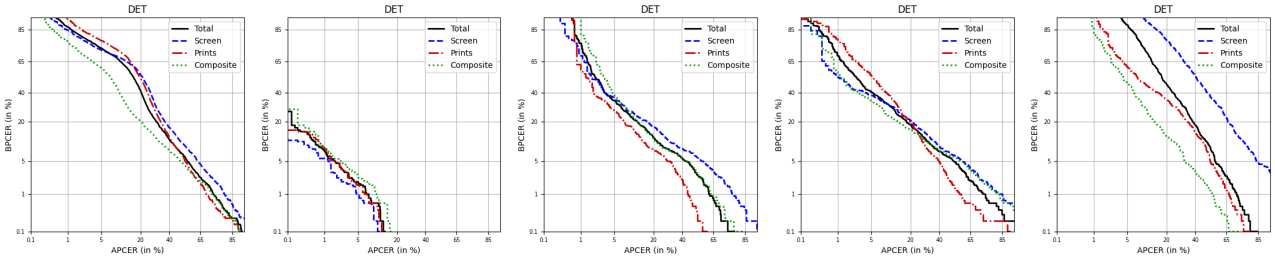

Figure 8. DET result for **IDVC_V2** model for the test set. The left-to-right plots show all countries together, followed by Chile, Guatemala, Panama, and Mexico subsets. The black line represents the binary results of bona fide versus attacks.

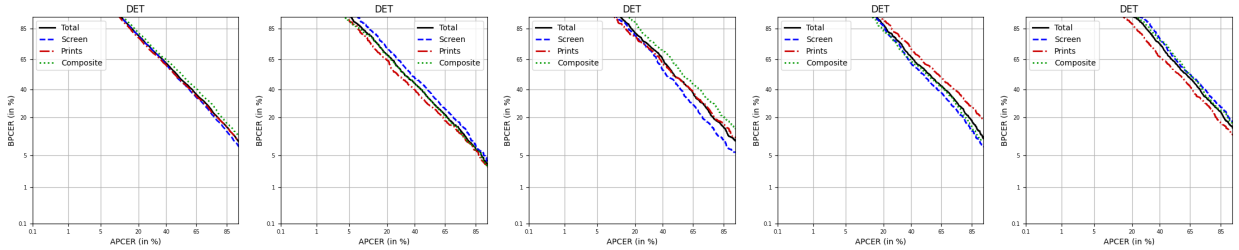

Figure 9. DET result for **SecureID** model for the test set. The left-to-right plots show all countries together, followed by Chile, Guatemala, Panama, and Mexico subsets. The black line represents the binary results of bona fide versus attacks.

[7] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807, 2017.

[8] N. Damer, N. Spiller, M. Fang, F. Boutros, F. Kirchbuchner, and A. Kuijper. PW-MAD: pixel-wise supervision for generalized face morphing attack detection. In *Advances in Visual Computing - 16th International Symposium, ISVC 2021, Virtual Event, October 4-6, 2021, Proceedings, Part I*, volume 13017 of *Lecture Notes in Computer Science*, pages 291–304. Springer, 2021.

[9] S. Gonzalez and J. E. Tapia. Improving presentation attack detection for id cards on remote verification systems, 2023.

[10] S. González, A. Valenzuela, and J. Tapia. Hybrid two-stage architecture for tampering detection of chipless id cards. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(1):89–100, 2021.

[11] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen,

M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le. Searching for mobilenetv3. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1314–1324, 2019.

[12] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017.

[13] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[14] L. Koliaskina, E. Emelianova, D. Tropin, V. Popov, K. Bulatov, D. Nikolaev, and V. Arlazarov. Midv-holo: A dataset for id document hologram detection in a video stream. In *International Conference on Document Analysis and Recognition*, pages 486–503. Springer, 2023.

[15] J. Magee, S. Sheridan, and C. Thorpe. An Investigation into the Application of the Meijering Filter for Document Recapture Detection. 2023. Publisher: Technological University Dublin.

[16] R. P. Markham, J. M. E. López, M. Nieto-Hidalgo, and J. E. Tapia. Open-set: Id card presentation attack detection using neural style transfer. *IEEE Access*, 12:68573–68585, 2024.

[17] E. Meijering, M. Jacob, J. Sarria, P. Steiner, H. Hirling, and M. Unser. Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry Part A*, 58A(2):167–176, Apr. 2004.

[18] R. Mudgalgundurao, P. Schuch, K. Raja, R. Ramachandra, and N. Damer. Pixel-wise supervision for presentation attack detection on identity document cards. *IET Biometrics*, 11(5):383–395, Sept. 2022.

[19] E.-J. Park, S.-Y. Back, J. Kim, and S. S. Woo. Kid34k: A dataset for online identity card fraud detection. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, page 5381–5385, New York, NY, USA, 2023. Association for Computing Machinery.

[20] D. V. Polevoy, I. V. Sigareva, D. M. Ershova, V. V. Arlazarov, D. P. Nikolaev, Z. Ming, M. M. Luqman, and J.-C. Burie. Document Liveness Challenge Dataset (DLC-2021). *Journal of Imaging*, 8(7):181, June 2022.

[21] Y. Qi, Y. He, X. Qi, Y. Zhang, and G. Yang. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6070–6079, October 2023.

[22] Y. Shi and A. K. Jain. Docface+: ID document to selfie matching. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 1(1):56–67, 2019.

[23] D. J. R. Soares Alysson De Sa and B. L. D. Bezerra. BID Dataset: a challenge dataset for document processing tasks. In *Anais Estendidos da Conference on Graphics, Patterns and Images (SIBRAPI Estendido 2020)*, pages 143–146, Brasil, Nov. 2020. Sociedade Brasileira de Computação.

[24] J. Tapia, C. Busch, H. Zhang, R. Ramachandra, and K. Raja. Simulating print/scan textures for morphing attack detection. In *31st European Signal Processing Conference (EUSIPCO)*, pages 610–614, 2023.

[25] C. Yang, Z. Yang, Y. Ke, T. Chen, M. Grzegorzek, and J. See. Doing more with moiré pattern detection in digital photos. *IEEE Transactions on Image Processing*, 32:694–708, 2023.