

Iris Liveness Detection Competition (LivDet-Iris) – The 2020 Edition

Priyanka Das^{†1}, Joseph McGrath^{†2}, Zhaoyuan Fang^{†2}, Aidan Boyd², Ganghee Jang¹,
Amir Mohammadi⁴, Sandip Purnapatra¹, David Yambay¹, Sbastien Marcel⁴, Mateusz Trokielewicz³,
Piotr Maciejewicz⁵, Kevin Bowyer², Adam Czajka², Stephanie Schuckers¹

¹Clarkson University, ²University of Notre Dame, ³Warsaw University of Technology, Poland,

⁴Idiap Research Institute, Switzerland, ⁵Medical University of Warsaw, Poland

Organizers [†]Equal Lead

Juan Tapia^{*6}, Sebastian Gonzalez^{*9}, Meiling Fang^{*7,8}, Naser Damer^{*7,8}, Fadi Boutros^{*7,8}, Arjan Kuijper^{*7,8}

⁶Universidad de Santiago - Chile, ⁷Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt,
Germany, ⁸Department of Computer Science, Technical University of Darmstadt, Darmstadt,

Germany, ⁹TOC Biometrics - Chile

*Competitors

Renu Sharma^{**}, Cunjian Chen^{**}, Arun Ross^{**}

Michigan State University

^{**}Providers of the MSU baseline algorithms

Abstract

Launched in 2013, LivDet-Iris is an international competition series open to academia and industry with the aim to assess and report advances in iris Presentation Attack Detection (PAD). This paper presents results from the fourth competition of the series: LivDet-Iris 2020. This year's competition introduced several novel elements: (a) incorporated new types of attacks (samples displayed on a screen, cadaver eyes and prosthetic eyes), (b) initiated LivDet-Iris as an on-going effort, with a testing protocol available now to everyone via the Biometrics Evaluation and Testing (BEAT)* open-source platform to facilitate reproducibility and benchmarking of new algorithms continuously, and (c) performance comparison of the submitted entries with three baseline methods (offered by the University of Notre Dame and Michigan State University), and three open-source iris PAD methods available in the public domain. The best performing entry to the competition reported a weighted average APCER of 59.10% and a BPCER of 0.46% over all five attack types. This paper serves as the latest evaluation of iris PAD on a large spectrum of presentation attack instruments.

*<https://www.idiap.ch/software/beat/>
978-1-7281-9186-7/20/\$31.00 ©2020 IEEE

1. Introduction

Iris recognition systems have been deployed in commercial and government applications across the globe for more than two decades [16]. Vulnerabilities of these systems against malicious attacks is an active area of research. One such attack that is being increasingly studied is the *presentation attack* (PA), where a sample is presented to the sensor with the goal of interfering with the correct operation of the system [4]. Presentation attacks may be carried out with different motives: (1) to impersonate an identity during verification, (2) to conceal an identity during recognition or (3) to create a virtual identity during enrollment [11]. Solutions to detect these attacks are referred to as *Presentation Attack Detection* (PAD) and include both hardware-based and software-based approaches. Software solutions are typically passive and mainly consider the static or dynamic features of the image or video presented to the system. Hardware solutions often employ active measurements of physical (color, density of tissue, optical properties) or physiological (pupil dilation) characteristics of the eye [7]. Research in presentation attack detection is an arms race: attacks on biometric systems are continually evolving, and system designers are continuously updating their security measures to efficiently detect artifacts as well as non-conformant uses of authentic biometric characteristics. LivDet-Iris is an international competition series launched in 2013 [26] to assess the current state of the art in iris PAD

by the independent evaluation of algorithms and systems on data and artifacts not seen by the competitors when designing their solutions. This paper reports on the fourth edition of this competition: *LivDet-Iris 2020*. The most significant **contributions of this paper** (and the LivDet-Iris 2020 competition itself) are:

- A report on the current state-of-the-art in iris PAD based on independent testing of **three algorithms submitted** to the competition organizers;
- Introduction of three **novel presentation attack instruments (PAI)**, when compared to previous LivDet editions: post-mortem iris images, electronic display, and fake/prosthetic/printed samples with add-ons. These attacks, combined with the printed iris images and the eyes with textured contact lenses, represent the **five different PAIs** in the test set, *i.e.* the largest spectrum of PAIs used to date in all iris PAD competitions.
- **Results from three different baseline methods** offered by the University of Notre Dame and the Michigan State University (see Sec. 4.5 for details) and **three open-source iris PAD methods** (see Sec. 4.6 for details).
- Availability of the competition through the **Biometrics Evaluation and Testing (BEAT)** [1, 2] platform (in addition to other algorithm submission options), implies some degree of privacy to the PAD algorithms as well as the test dataset.
- **Initiation of LivDet-Iris Competition as “Ongoing”**, *i.e.*, the competition benchmark will remain available to all researchers through the BEAT platform after this edition is concluded, which allows for testing all future algorithms according to the LivDet-Iris 2020 protocol, **without revealing the test data**.

2. Performance Evaluation Metrics

LivDet-Iris 2020 follows the recommendations of ISO/IEC 30107-3 [15] in employing two basic PAD metrics in its evaluations:

- **Attack Presentation Classification Error Rate (APCER)**, the proportion of attack presentations of the same PAI species incorrectly classified as bonafide presentation, *i.e.* spoof classified as live, and
- **Bonafide Presentation Classification Error Rate (BPCER)**, the proportion of bonafide presentations classified as attack presentations, *i.e.* live classified as spoof.

Both the APCER and BPCER metrics are used to evaluate the algorithms. ISO also recommends to use the maximum value of APCER when multiple PA species (or categories) are present in case of system-level evaluation, which is primarily designed for industry applications. This, however, is inconsistent with our prior competitions [26, 27, 25] and also our goal to consider the detection of all PAIs,

and not to rank the competitors by looking at their worst-performing PA. Thus, we introduced the weighted average of APCER over all PAIs:

- **Weighted Average of APCER** ($APCER_{\text{average}}$), which is the average of APCER across all PAIs, weighted by the sample counts in each PAI category, as reported in Table 2.

Only for the **purpose of competition ranking**, the Average Classification Error Rate (ACER) was computed to select the winner:

- **Average Classification Error Rate (ACER)**: the average of $APCER_{\text{average}}$ and BPCER.

Note that ACER has been deprecated in ISO/IEC 30107-3:2017 [15] in the industry-related PAD evaluations.

3. Iris PAD Evaluation Efforts To Date

Iris PAD literature offers a wide spectrum of software- and hardware-based solutions, and two recent survey papers [4, 7] provide a comprehensive overview of the current state of the art. In this section, we offer a summary of all known public iris PAD evaluation efforts to date.

3.1. MobILive

Mobile Iris Liveness Detection Competition (MobILive) was held in 2014 to assess the state of art of algorithms for iris liveness detection for **mobile applications**. The competition concentrated on the simplest PAI: printed iris images of an authorized subject presented to the sensor by an unauthorized subject. The purpose of the competition was to assess the performance of algorithms to distinguish between live iris images and paper iris printouts. The best performing algorithm achieved the mean of APCER and BPCER equal to 0.25% [18].

3.2. LivDet-Iris 2013, 2015 and 2017

LivDet-Iris 2013 [26] was the first public evaluation platform for advancements in iris PAD focused on systems and algorithms employing **ISO-compliant iris images** (in particular, acquired in near infrared light). The competition subsequently occurred in 2015 [27] and 2017 [25]. Every LivDet-Iris competition offered both software- and system-level evaluation. The software-based competition evaluates the performance of *algorithms* in the task of detection of presentation attacks. The system-based competition evaluates the performance of *complete systems* (including sensors) against physical presentation attacks. Over the past editions, ten participants submitted algorithms and none elected to submit for a system-level evaluation. Table 1 summarizes all editions, including the current (2020) installment.

Table 1: LivDet-Iris Competition Series Summary

Competition year	Presentation Attack Instruments in test data	New train / test data delivered by organizers	Number of competitors	Best performance	
				APCER	BPCER
2013	Printed Irises, Patterned Contact Lenses	Yes / Yes	3	5.7%	28.6%
2015	Printed Irises, Patterned Contact Lenses	Yes / Yes	4	5.48%	1.68%
2017	Printed Irises, Patterned Contact Lenses	Yes / Yes	3	14.71%	3.36%
2020 (reported in this paper)	Printed Irises, Patterned Contact Lenses, Fake/Prosthetic/Printed Eyes with Add-ons Eyes Displayed on Kindle, Cadaver Irises	No / Yes	3	59.10%	0.46%

3.3. LivDet-Iris 2020

The LivDet-Iris 2020 competition was launched in May 2020 and was co-organized by five organizations: Clarkson University (USA), University of Notre Dame (USA), Warsaw University of Technology (Poland), IDIAP Research Institute (Switzerland) and Medical University of Warsaw (Poland). In previous editions, this competition had two parts: *Algorithms* and *Systems*. The *Algorithms* part required participants to submit their algorithm(s) to the organizers for independent testing on the unknown test data. The *Systems* section required submission of a complete system, including hardware, designed for presentation attack detection. After submission, our team evaluated the submitted systems based on varied physical attack types. One winner was selected for each part of the competition based on the average performance of detecting spoofs and accepting live samples. Participation was encouraged from all academic and industrial institutions. In contrast to past LivDet-Iris competitions, the 2020 edition did not offer any official training data – the competitors were free to use any proprietary and/or publicly available data to design their algorithms. In this 2020 edition, for the first time, an open-source research experimentation platform, BEAT was used to host the competition. The BEAT platform facilitates further evaluations of algorithms by any researcher, using the identical test data and protocol as in this competition edition, even after the completion of the competition.

It is important to note that the entire LivDet-Iris competition series focuses on evaluation of capabilities of algorithms to **generalize to unknown circumstances**. While a very brief characterization of attack types included in the test set is provided to the competitors, the test samples are not revealed, and therefore not used for training. All evaluations are completed by the LivDet-Iris organizers, and are *not* self-reported by participants.

4. Experimental Protocol and Evaluation

4.1. Participation

Participation in LivDet-Iris 2020 was open to all academic and industrial institutions with the option to participate anonymously in both the *Algorithms* and *Systems* part.

Anonymous participation allowed the competitors to retain their identity from the co-authors’ list. Fourteen teams registered for the competition from across the globe. The organizers received three algorithm submissions from three registered teams. There were no submissions to the systems portion of the competition. All three competing teams were invited to contribute to this report by describing their PAD methods briefly.

4.2. Datasets

Training dataset LivDet-Iris 2020 was different from previous editions in that the organizers did not announce any official training set. Instead, the participants were encouraged to use all data available to them (both publicly available and proprietary) to make their solutions as effective and robust as possible. The entire past LivDet-Iris benchmarks were also made publicly available [26, 27, 25]. Additionally, the competition organizers shared 5 examples of each PAI (and these samples were not used later in evaluations) to familiarize the competitors with the test data format (pixel resolution, bits per pixel used to code the intensity, etc.).

Test dataset The testing set employed in this competition was a combination of data from all three organizers: Clarkson University (CU), University of Notre Dame (ND) and Warsaw University of Technology (WUT). The dataset consisted of 12,432 images (5,331 live and 7,101 spoof samples), as summarized in Table 2. Sample images from the dataset are shown in Figure 1. Five Presentation Attack Instruments (PAI) categories were included in the dataset:

- **Printed eyes:** 1,049 samples created using five different printers (Epson Stylus Pro 9900, HP M652, Xerox C60, OKI MB-471, Cannon Super G3) and two different print qualities (“office” and “professional”). Two different paper types were used (matte and glossy paper). Images were collected with the Iris ID iCAM7000.
- **Textured contact lenses:** 4,336 samples were acquired using LG IrisAccess 4000 and IrisGuard AD100 under different illumination setups offered by these sensors (two different illuminants in LG 4000 and six different illuminants in AD 100). This portion of the data were collected

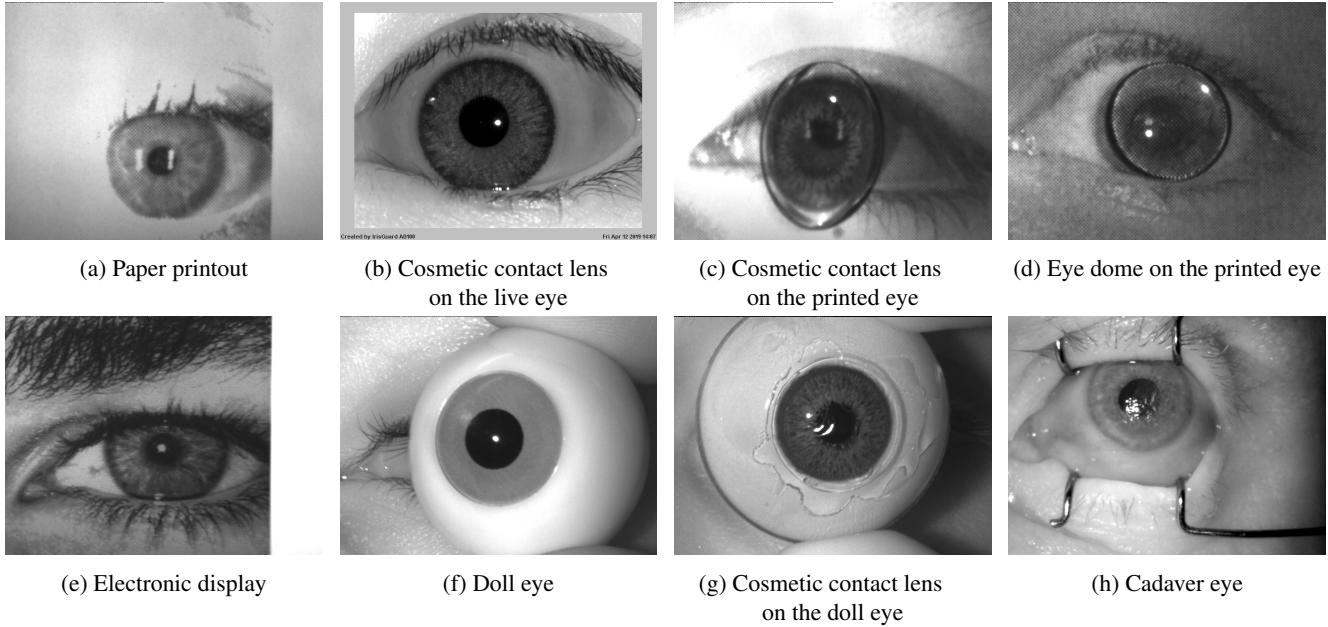


Figure 1: Example images of all presentation attack types present in the LivDet-Iris 2020 test dataset.

from 88 subjects (176 irises) wearing cosmetic contact lenses of three different brands: Johnson & Johnson, Ciba Vision, and Bausch & Lomb.

- **Eyes displayed on Kindle e-Ink:** 81 such samples were captured in NIR spectrum the Iris ID iCAM7000 sensor.
- **Fake/Prosthetic/Printed Eyes with Add-ons:** This category has five sub-categories of spoofs making up a total of 541 samples, captured in NIR spectrum by the Iris ID iCAM7000 sensor, with non-uniform distribution of samples in each category:
 - **Textured Contacts on Printed Eyes:** patterned contact lenses added on top of the printed eye images.
 - **Textured Contacts on Doll Eyes:** patterned contact lens put on the iris area of plastic doll eyes.
 - **Clear Contacts on Printed Eyes:** transparent contact lenses added on top of the printed eye images.
 - **Eye Dome on Printed Eyes:** transparent 3D plastic eye domes added on top of the printed eye images.
 - **Doll Eyes:** Fake eyes of two different types – Van Dyke Eyes (has higher iris quality details) and Scary eyes (plastic fake eyes with simple pattern on iris region); different color variation of both types of fake eyes were included.
- **Cadaver Eyes:** The Warsaw-BioBase-Post-Mortem-Iris v3.0 dataset [20, 24] encompasses a total of 1,094 NIR images (collected with an IriShield M2120U handheld iris recognition camera) and 785 visible light images (obtained with Olympus TG-3) collected from 42 post-mortem subjects, and is fully subject-disjoint from previ-

ous Warsaw post-mortem dataset publicly available before LivDet-Iris 2020 competition [22, 23]. Data collection sessions were organized accordingly with medical staff and cadaver availability and ranges from several hours after demise up to 369 hours postmortem. For the purpose of LivDet-Iris 2020 competition, only NIR images are employed. This data collection had institutional review board clearance and the ethical principles of the Helsinki Declaration were followed by the data acquisition staff.

4.3. Experimentation Platform

LivDet-Iris 2020 used the benefits of the Biometrics Evaluation and Testing (BEAT) platform for the competition. BEAT is a solution for open access, scientific information sharing and re-use including data and source code while protecting privacy and confidentiality. It allows easy online access to experimentation and testing in computational science. The platform allows access and comparison of different experimentation and results. [1, 2]. Participants were encouraged to submit their algorithm through this platform. However, we also accepted the submissions that were sent to us for evaluation as executables for Windows, Linux, or Mac OS. Alternatively, we accepted also codes in Python (v.3.7) or MATLAB (2019a or above). All submission options were equivalent in terms of participation. There are several advantages of conducting this edition of LivDet-Iris in the BEAT platform:

Table 2: Test Dataset Summary

Class	Presentation Attack Instruments	Sample Count	Sensor
Live	-	5,331	LG 4000, AD 100, Iris ID iCAM7000
Spoof	Printed Eyes	1,049	Iris ID iCAM7000
Spoof	Textured Contact Lens	4,336	LG 4000, AD 100, Iris ID iCAM7000
Spoof	Electronic Display	81	Iris ID iCAM7000
Spoof	Fake/Prosthetic/Printed Eyes with Add-ons	541	Iris ID iCAM7000
Spoof	Cadaver Iris	1,094	IriTech IriShield

- **Privacy:** The algorithms submitted by the participants remain invisible to everybody except the participant. Similarly, any data uploaded to the BEAT platform also remain inaccessible to any user of the platform. In particular, this allowed us to share the test data in an anonymous and reproducible manner that otherwise could not be shared, due to sponsor restrictions.
- **Re-submission:** The participants can make multiple submissions before the deadline but do not have access to results until after their final submission.
- **Continuity:** This platform will serve as an iris PAD “on-going” benchmark after LivDet-Iris 2020 is concluded, since the test data and protocol are planned to be retained on BEAT and available for executing algorithms by all interested researchers.

4.4. LivDet Iris 2020 Competition Algorithms

All teams were given the opportunity to submit a description of their submitted algorithm, and two such descriptions are provided below. One team elected not to provide a description.

USACH/TOC Team: For this competition an algorithm was presented based on a multilabel CNN network that has been used to detect printed images and patterned contact lenses. The SMobileNet and FMobilNet models are both based on MobilenetV2. SMobileNet was trained from scratch to detect the presence of patterned contact lenses in the iris image area. FMobilNet was trained using fine-tuning with average and max pooling options, in order to detect the printed images of the whole image by identifying the physical source of the image. Finally, a multi-output classifier was developed in order to identify fake or live or real images. This option allowed the team to create a lightweight classifier to be implemented in a mobile iris recognition camera such as Gemini Iritech.

FraunhoferIGD Team: The algorithm starts by finding special local-features in the investigated image. These local-features are clustered into a number of classes. Moreover, an image patch is extracted from the area around each of these local-features. For each of the clusters, a classifying network is used to determine the origin of the patch (belonging to this specific cluster) as a bonafide image or an attack image. After that, a logistic regression model was trained for each cluster class. This logistic regression

takes the classification probability from the network and the cluster class reliability and results in a final bonafide/attack classification score. All the classification scores produced by the logistic regressions from different patches are fused using a simple mean-rule. The algorithm uses the K-means approach to build the local-feature clusters and calculate the class reliability. The used patch size is 64×64 pixels. The classification network used is trained from scratch and is based on the MobileNetV3-Small [12] neural network architecture.

4.5. Baseline Algorithms

All organizing teams who contributed to the baseline and open-source algorithm performance evaluation had access to LivDet Iris 2020 test dataset. However, as a declaration, every team that contributed a baseline/open source algorithm verified that they did not use the data from the test dataset as part of training.

Notre Dame PAD Algorithm: The implemented solution extends the methodology proposed by Doyle and Bowyer [3] and the feature extraction is based on Binary Statistical Image Features (BSIF) proposed by Kannala and Rahtu [17]. In this method, the calculated “BSIF code” is based on filtering the image with n filters of size $s \times s$, and then binarizing the filtering results with a threshold at zero. Hence, for each pixel n binary responses are given, which are in the next step translated into a n -bit grayscale value. The histograms resulting from gray-scale BSIF codes are later normalized to a z -score and used as texture descriptors with the number of histogram bins equal to 2^n . We use a *Best Guess* segmentation technique to select a region of interest. A separate set of three classifiers (SVM, RF, and MLP) was trained on the Notre Dame LivDet-Iris 2017 dataset for each feature set (n, s pair). Since not all the classifiers have the same strength, a subset of the strongest classifiers was selected through testing on the Clarkson LivDet-Iris 2017 dataset and majority voting is applied to these selected classifiers to come up with a final decision.

MSU PAD Algorithm 1: The proposed algorithm, namely TL-PAD [6, 5], operates on the cropped iris regions and offers a simple and fast solution. It utilizes the pre-trained ImageNet model to initialize the weights and then performs transfer learning. First, an off-line trained iris detector [6] was used to obtain a rectangular region encom-

passing the outer boundary of the iris. Then, the iris region was automatically cropped based on the estimated rectangular coordinates. Finally, the cropped iris region was input to a CNN to train the iris PA detection model. MobileNetV2 was used as the backbone network with squeeze-and-excitation module applied to the last convolution layer to recalibrate channel-wise features. The training was fine-tuned on an existing ImageNet model, by leveraging extensive data augmentation schemes, including rotation, shift and flip operations, to name a few. The learning rate was set to 0.0001 and Adam optimizer was used. The algorithm was trained on a proprietary dataset comprising of 12,058 live images and 10,622 PA images.

MSU PAD Algorithm 2: MSU second baseline method is a variant of D-NetPAD [19] whose base architecture is Dense Convolutional Network 161 (DenseNet161) [14]. The input to the model is a cropped iris region resized to 224×224 . The model weights are first initialized by training on the ImageNet dataset [8] and then fine-tuned using bonafide iris and PA samples. Fine-tuning was performed with a proprietary dataset, NDCLD-2015 [21] and Warsaw PostMortem v3 dataset. The proprietary dataset consists of 19,453 bonafide irides and 4,047 PA samples. PA samples include 51 kindle display attacks, 1,005 printed eyes, 1,804 artificial eyes, and 1,187 cosmetic contact lenses. From the NDCLD-2015 dataset, 2,236 cosmetic contact lenses images were used for the training, and from the Warsaw PostMortem v3 dataset, 1,200 cadaver iris images from the first 37 cadavers were used. The architecture consists of 161 convolutional layers integrated into four Dense blocks, and three Transition Layers lie between the Dense blocks. The last layer is a fully connected layer. A detailed description of the architecture is provided in [14]. The learning rate used for the training is 0.005, the batch size is 20, the number of epochs is 50, the optimization algorithm is stochastic gradient descent with a momentum of 0.9, and the loss function used is cross-entropy.

4.6. Open Source Algorithms

For completeness, three open-source iris PAD algorithms available today in the public domain (in addition to the baseline algorithms) are also evaluated. All three algorithms are trained on a subset of the 2017 LivDet-Iris competition data, constructed such that 100 samples are taken from each unique combination of data provider, image label, and dataset partition (*e.g.* one possible combination would be Notre Dame, contact lens, and the training set).

RegionalPAD: Hu *et al.* [13] investigate the use of regional features in iris PAD. Features are extracted from local neighborhoods based on spatial pyramid (multi-level resolution) and relational measures (convolution on features with variable-size kernels). Several feature extractors such as Local Binary Patterns (LBP), Local Phase Quanti-

zation (LPQ), and intensity correlogram are investigated. In our experiments, we use the three-scale LBP-based feature, since it achieves the best performance as pointed out by the original authors.

SIDPAD: Gragnaniello *et al.* [9] proposes that the sclera region also contains important information about iris liveness. Hence, the authors extract features from both the iris and sclera regions. The two regions are first segmented and scale-invariant local descriptors (SID) are applied. A bag-of-feature method is then used to summarize the features. A linear Support Vector Machine (SVM) is used to perform final prediction. We refer to this method as SIDPAD.

DACNN: Gragnaniello *et al.* [10] incorporates domain-specific knowledge of iris PAD into the design of their model. With the domain knowledge, a compact network architecture is obtained and regularization terms are added to the loss function to enforce high-pass / low-pass behavior. The authors show that the method can detect both face and iris spoofing attacks. We refer to this method as DACNN.

5. Results and Analysis

This section discusses the performance of the algorithms in 3 categories: (1) LivDet-Iris 2020 competitors, (2) baseline algorithms, and (3) open-source algorithms. The performance has been evaluated based on APCER for each of the five PAIs. APCER and BPCER are evaluated at the threshold of 0.5, which was announced prior to the competition. A summary of the error rates for all 3 categories is provided in Table 3. The ROCs shown for all PAIs broken by algorithm category (competitors, open-source and baselines) are shown in Figure 2. The ROCs for individual PAIs for all nine methods are depicted in Figure 3. Below we discuss the performance of algorithms in three groups of methods.

LivDet Iris 2020 Competitors: Team USACH/TOC was determined as the winner based on lowest ACER = 29.78%, very closely followed by Team FraunhoferIGD with ACER = 30.14%, and Competitor-3 with an ACER of 49.06%. The winning team's method achieved also the lowest BPCER = 0.46% out of all nine algorithms in the three categories. This aligns well with the operational goal of PAD algorithms to correctly detect bonafide presentations (*i.e.*, and not to contribute to system's False Rejection Rate) and capture as many attacks as possible. The three algorithms had variable performance for each type of PAI. The algorithm offered by USACH/TOC was specifically tuned for printed eyes (PE) and trained from scratch to detect textured contact lenses (CL) (as explained in Sec. 4.4), but performed best for the electronic display (ED) PAI achieving APCER=9.87%, which is lower than all competing algorithms (53.08% and 83.95%) by a large margin. Algorithm offered by Fraunhofer IGD, performed best in three categories: printed eyes (APCER = 14.87%),

Table 3: Error Rates (%) for all algorithms calculated at a threshold of 0.5, corresponding to each PAI (ACPER) and the overall performance (ACER)

Method category	Algorithm	APCER					Overall Performance		ACER
		PE	CL	ED	F/P	CI	APCERaverage	BPCER	
Livet Iris 2020 Submissions	Team: USACH/TOC	23.64	66.01	9.87	25.69	86.10	59.10	0.46	29.78
	Team: FraunhoferIGD	14.87	72.80	53.08	19.04	0	48.68	11.59	30.14
	Competitor-3	72.64	43.68	83.95	73.19	89.85	57.8	40.31	49.06
Baselines*	ND PAD**	55.95	50.74	35.80	43.25	92.59	57.21	0.71	28.96
	MSU PAD Algorithm 1	14.96	2.23	23.45	10.90	0	4.67	0.56	2.61
	MSU PAD Algorithm 2	2.38	3.85	1.23	0.18	0.18	2.76	1.61	2.18
Open Source	DACNN**	54.53	45.94	75.31	41.22	97.99	55.2	16.39	35.8
	SIDPAD**	8.48	52.19	1.24	17.93	99.82	49.85	39.96	44.9
	RegionalPAD**	92.18	67.62	96.29	70.79	6.49	62.42	23.80	43.11

* Authors had access to the entire LivDet-Iris 2020 test dataset, but did not use it as part of training. ** These methods were **not** trained on all categories of PAIs present in the LivDet-Iris 2020 test dataset. **PE**: Printed Eyes; **CL**: Textured Contact Lens; **ED**: Electronic Display; **F/P**: Fake/Prosthetic/Printed Eyes with Add-ons; **CI**: Cadaver Iris.

fake/prosthetic eyes (APCER = 19.04%) and cadaver irises (perfect detection of all cadaver samples). However, the detection of bonafide samples was worse (BPCER = 11.59%) compared to the winner (BPCER=0.46%). The algorithm offered by Competitor-3 was the lowest for the contact lenses (CL) category with APCER = 43.68%. It is important to note that all of these results are based on independent evaluation, the competitors did not have an access to test data, and trained their algorithms on data not necessarily representing all PAIs present in the test data. It demonstrates the difficulty of open-set iris PAD.

Baseline Algorithms: The results of baseline algorithms are included to additionally demonstrate how a good representation of PAIs during training is important. The baseline offered by the University of Notre Dame was trained solely on live samples and contact lens PAI. It is thus not surprising to see an overall performance (across all PAIs) to be close to the competition winner. The weak performance on CL PAI category also suggests lower generalization capabilities of this image texture-based method onto unknown contact lens brands and patterns. In contrast, two baselines developed by Michigan State University offer the best performance out of all 9 algorithms. This could be due to the use of a more comprehensive training set to design the methods. In particular, MSU PAD Algorithm 2 resulted in a weighted APCER of 2.76% at a BPCER of 1.61%.

Open Source Algorithms: All three algorithms lacked balance in the performance between bonafide and attack samples, and achieved high BPCER ranging between approximately 16% and 40%. The SIDPAD algorithm, which considers both the sclera and iris portions of the eye in their algorithm design to detect presentation attacks, performed well for three PAIs in comparison to other algorithms: printed eyes (PE) with APCER of 8.48%, electronic display (ED) with an APCER of 1.24% and fake/prosthetic/printed

eyes with add-ons with APCER of 17.93%. However, SIDPAD demonstrated limited accuracy of bonafide detections (approx. 40% of BPCER) and failed in recognizing cadaver irises. The RegionalPAD algorithm, which considered three-scale LBP based features, achieved a low error rate only with cadaver iris (CI) attack type with APCER of 6.49%, while it failed to detect reliably printed eyes and electronic display attacks. The DACNN algorithm, even though it demonstrated a relatively low ACER in the open-source category, it presented limited capability of detecting all PAIs, and also relatively high BPCER = 16.39%. This may suggest that some of the older iris PAD methods, available in the public domain, may offer lower accuracy when applied to currently observed attacks.

6. Conclusions

The LivDet-Iris 2020 featured multiple new additions to the evaluation of iris presentation attack detection: (a) employed three novel PAIs (cadaver iris, fake/prosthetic/printed eyes with various add-ons, and electronic display), (b) introduced BEAT platform for the competition thereby facilitating privacy, re-submission and continuity of algorithm evaluation in the public domain, and (c) provided a comparative analysis of the nine state-of-art methods (three competition algorithms, three baseline algorithms and three open-source algorithms). The winning entry performed with an ACER of 29.67% (BPCER = 0.46 and APCER averaged over all PAIs = 59.10%). However, two baseline algorithms from MSU resulted in the best performance.

We note a degradation in the best overall performance of the winning entry in LivDet-Iris 2020 compared to the previous competitions organized in 2013, 2015 and 2017 (as shown in Table 1). This degradation can be attributed to multiple factors:

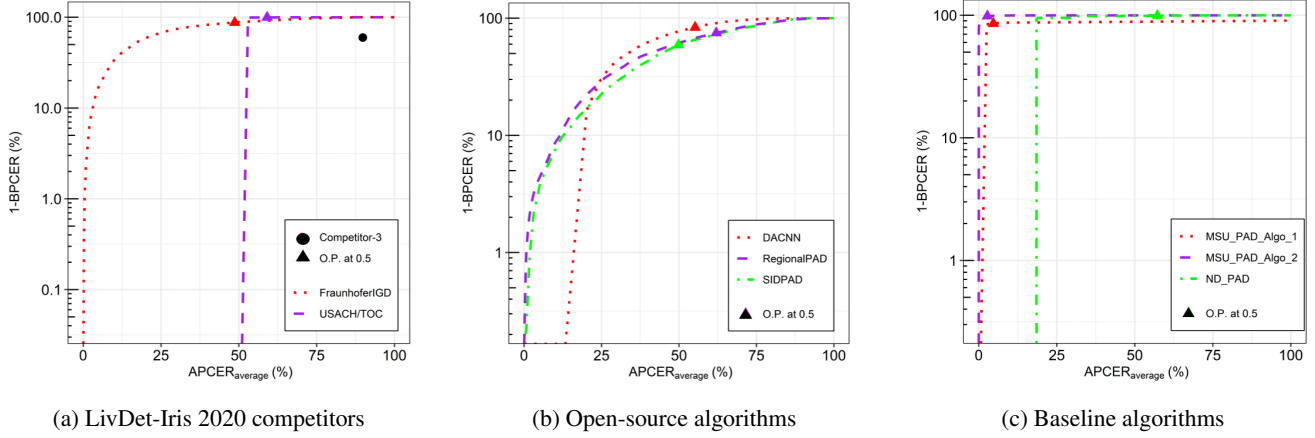


Figure 2: ROC curves for all nine algorithms presenting the overall performance on samples representing all five PAIs. The overall APCER is evaluated based on ($APCER_{average}$). The operating point (“O.P. at 0.5”) used to rank participants of this LivDet-Iris competition is marked by a \blacktriangle on each curve.

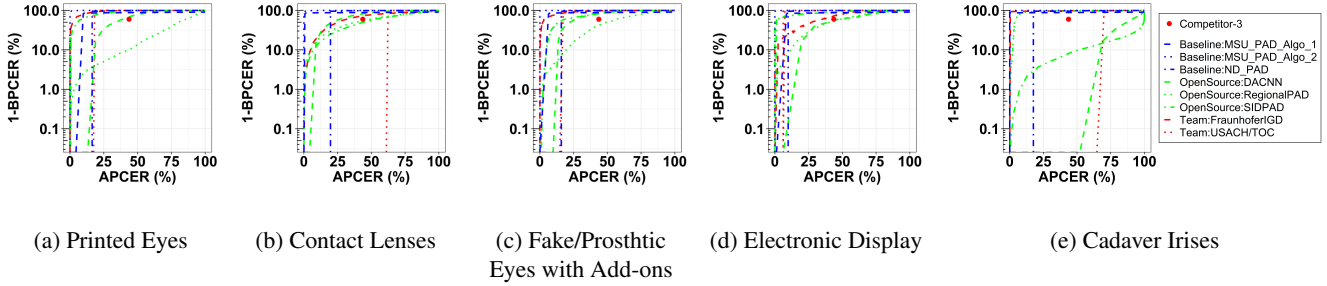


Figure 3: Same as in Fig. 2, except that the performance of all nine methods is presented separately for each PAI.

- highly increased complexity in the test dataset: five different PAI categories were employed in the competition this year compared to two PAIs in previous editions;
- introduction of novel attack types with limited or no access to large-enough public datasets for a few PAIs;
- no specific training dataset was offered, and that design choice was left to be decided by competitors;
- the results could reflect variability between the training and the test datasets in terms of environmental factors, sensors, quality of PAIs, and the use of “unseen” PAIs.

The results from this competition indicate that iris PAD is still far from a fully solved research problem. Large differences in accuracy among baseline algorithms, which were trained with significantly different data, stress the importance of access to large and diversified training datasets, encompassing a large number of PAIs. We believe that this competition, and the benchmark now available to researchers via the BEAT platform, will contribute to our efforts as a biometric community in winning the PAD arms race.

7. Acknowledgement

This material is based upon work supported in part by the National Science Foundation under Grant No. #1650503 and the Center for Identification Technology Research. MSU’s and CU’s research is supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2017 - 17020200004. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

- A. Anjos, L. El Shafey, and S. Marcel. BEAT: An open-science web platform. In *Thirty-fourth International Conference on Machine Learning*, 2017. 2, 4
- A. Anjos, L. El Shafey, and S. Marcel. Beat: An open-source web-based open-science platform, Apr. 2017. 2, 4

- [5] C. Chen and A. Ross. Exploring the Use of IrisCodes for Presentation Attack Detection. In *IEEE International Conference on Biometrics Theory, Applications and Systems*, pages 1–9. IEEE, 2018. 5
- [6] C. Chen and A. Ross. A multi-task convolutional neural network for joint iris detection and presentation attack detection. In *IEEE Winter Applications of Computer Vision Workshops*, pages 44–51, 2018. 5
- [7] A. Czajka and K. W. Bowyer. Presentation attack detection for iris recognition: An assessment of the state-of-the-art. *ACM Computing Surveys (CSUR)*, 51(4):1–35, 2018. 1, 2
- [8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009. 6
- [9] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva. Using iris and sclera for detection and classification of contact lenses. *Pattern Recognition Letters*, 82:251–257, 2016. 6
- [10] D. Gragnaniello, C. Sansone, G. Poggi, and L. Verdoliva. Biometric spoofing detection by a domain-aware convolutional neural network. In *12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 193–198, 2016. 6
- [11] S. Hoffman, R. Sharma, and A. Ross. Convolutional neural networks for iris presentation attack detection: Toward cross-dataset and cross-sensor generalization. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1620–1628, 2018. 1
- [12] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam. Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1314–1324, October 2019. 5
- [13] Y. Hu, K. Sirlantzis, and G. Howells. Iris liveness detection using regional features. *Pattern Recognition Letters*, 82:242–250, 2016. 6
- [14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017. 6
- [15] ISO/IEC 30107-3. Information technology – Biometric presentation attack detection – Part 3: Testing and reporting, 2016. 2
- [16] A. K. Jain, K. Nandakumar, and A. Ross. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, 79:80–105, 2016. 1
- [17] J. Kannala and E. Rahtu. BSIF: Binarized statistical image features. In *The 21st International Conference on Pattern Recognition (ICPR)*, pages 1363–1366, 2012. 5
- [3] K. W. Bowyer and J. S. Doyle. Cosmetic contact lenses and iris recognition spoofing. *Computer*, 47(5):96–98, May 2014. 5
- [4] A. Boyd, Z. Fang, A. Czajka, and K. W. Bowyer. Iris presentation attack detection: Where are we now? *arXiv preprint arXiv:2006.13252*, 2020. 1, 2
- [18] A. F. Sequeira, H. P. Oliveira, J. C. Monteiro, J. P. Monteiro, and J. S. Cardoso. Mobilive 2014 – mobile iris liveness detection competition. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–6, 2014. 2
- [19] R. Sharma and A. Ross. D-NetPAD: An Explainable and Interpretable Iris Presentation Attack Detector. *IEEE International Joint Conference on Biometrics (IJCB)*, 2020. 6
- [20] M. Trokielewicz, A. Czajka, and P. Maciejewicz. Post-mortem iris recognition with deep-learning-based image segmentation. *Image and Vision Computing*, 94:103866, 2020. 4
- [21] University of Notre Dame. *The Notre Dame Contact Lens Dataset – NDCLD 2015*. available at: <https://cvrl.nd.edu/projects/data>. 6
- [22] Warsaw University of Technology, Poland. *Warsaw-BioBase-PostMortem-Iris-v1.1* database. available at: <http://zbum.ia.pw.edu.pl/EN/node/46>, 2016. 4
- [23] Warsaw University of Technology, Poland. *Warsaw-BioBase-PostMortem-Iris-v2.0* database. available at: <http://zbum.ia.pw.edu.pl/EN/node/46>, 2018. 4
- [24] Warsaw University of Technology, Poland. *Warsaw-BioBase-PostMortem-Iris-v3.0* database. available at: <http://zbum.ia.pw.edu.pl/EN/node/46>, 2020. 4
- [25] D. Yambay, B. Becker, N. Kohli, D. Yadav, A. Czajka, K. W. Bowyer, S. Schuckers, R. Singh, M. Vatsa, A. Noore, D. Gragnaniello, C. Sansone, L. Verdoliva, L. He, Y. Ru, H. Li, N. Liu, Z. Sun, and T. Tan. LivDet-Iris 2017 – Iris Liveness Detection Competition 2017. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 733–741, 2017. 2, 3
- [26] D. Yambay, J. S. Doyle, K. W. Bowyer, A. Czajka, and S. Schuckers. LivDet-Iris 2013 – Iris Liveness Detection Competition 2013. In *IEEE International Joint Conference on Biometrics*, pages 1–8, 2014. 1, 2, 3
- [27] D. Yambay, B. Walczak, S. Schuckers, and A. Czajka. LivDet-Iris 2015 – Iris Liveness Detection Competition 2015. In *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pages 1–6, 2017. 2, 3