Learning Joint Local-Global Iris Representations via Spatial Calibration for Generalized Presentation Attack Detection

Gaurav Jaswal*, Aman Verma*, Sumantra Dutta Roy*, Raghavendra Ramachandra†
*Indian Institute of Technology Delhi, India
†Norwegian University of Science and Technology Norway, Norway
Email: *(gauravjaswal; eey227536; sumantra)@ee.iitd.ac.in, †raghavendra.ramachandra@ntnu.no

***Abstract*—Existing Iris Presentation Attack Detection (IPAD) systems do not generalize well across datasets, sensors and subjects. The main reason for the same is the presence of similarities in bonafide samples and attacks, and intricate iris textures. The proposed DFCANet (Dense Feature Calibration Attention-Assisted Network) uses feature calibration convolution and residual learning to generate domain-specific iris feature representations at local and global scales. DFCANet's channel attention enables the use of discriminative feature learning across channels. Compared to state-of-the-art methods, DFCANet achieves significant performance gains for the IIITD-CLI, IIITD-WVU, IIIT-CSD, Clarkson-15, Clarkson-17, NDCLD-13, and NDCLD-15 benchmark datasets. Incremental learning in DFCANet overcomes data scarcity issues and cross-domain challenges. This paper also pursues the challenging soft-lens attack scenarios. An additional study conducted over contact lens detection task suggests high domain-specific feature modeling capacities of the proposed network.**

***Index Terms*—Feature calibration, iris-spoofing, channel attention.**

I. Introduction

Iris Presentation Attack Detection (IPAD) handles attacks from cosmetic contact lenses, holographic eyes, textured contact lenses, prosthetic eyes, printed iris images, iris videos, paper iris printouts, drug-induced iris manipulation and fake eyeballs [1]–[3]. LivDet-Iris is an international competition series launched in 2013 to assess the current state-of-the-art in IPAD by the independent evaluation of algorithms [2], [4]. In addition, results from the LivDet-Iris 2017 [5] and LivDet-Iris 2020 [4] competitions illustrated that state-of-the-art methods still obtain limited accuracy in IPAD. Under the stringent evaluation protocols introduced in the same, generalization has been observed to be affected [3], [6]. This degradation in performance was eminently observed for unseen-attack and cross-dataset scenarios, where characteristics of testing samples differ from the samples considered for training.

Iris images are characterized by complex iris patterns, as well as morphological artifacts included in the annular structure of the eye, such as eyelashes. Another challenging aspect is that the iris biometric remains concealed within limited boundaries throughout the image. Uneven lighting conditions and occluded image capture further degrade attention to important details. Iris segmentation in unconstrained scenarios is thus a difficult problem. The problems discussed above deepen when IPAD algorithms are evaluated in cross-domain scenarios. Soft-Lens situations bear high visual similarity to normal iris images. Ironically, this aspect has received very limited attention in the literature [7]–[9]. Fig. 1 illustrates various realistic evaluation scenarios such as Inter-Sensor, Cross-Dataset, Inter-Subject and specifically Soft-Lens-as-Attack impends extreme challenges for IPAD algorithms. The following pertinent points arise:

- What domain-specific characteristics accentuate the separation between bonafide and presentation attack samples?
- A consistent IPAD performance even in challenging cross-domain scenarios with drastic training-testing drift.
- Good performance in data-scarce, unconstrained, and zero preprocessing conditions.

We present DFCANet to take advantage of local-global feature learning, and thus generalizing well for unseen and cross-domain IPAD scenarios. It is observed that the iris has similar textures existing at both local and global scales [9]. As illustrated in Fig. 1 there exists a significant difference between the activations of the attack and the bonafide samples at the local and global levels. The proposed DFCANet identifies and accentuates local features and their global context in attacks, differentiating them from bonafide samples. Our model operates on raw iris images without any preprocessing, and has a feature calibration-assisted backbone network for robust IPAD. To validate the efficacy of the proposed model, we perform challenging experiments and extensive ablation studies on seven benchmark data sets, namely IIITD-CSD, Clarkson-15, Clarkson-17, NDCLD-13, NDCLD-15, IIITD-WVU, and IIITD-CLI. Our main contributions are as follows:

- DFCANet extracts domain-specific knowledge in terms of local-global iris patterns using residually connected iris feature calibration, for accentuated discrimination between bonafide and attack images.
- Our extensive experimentation on seven benchmark datasets considers both settings: soft-lens as bonafide (defined well in the literature) and the more realistic soft-lens as an attack. (Normal images and soft-lens cases are visually similar!) DFCANet effectively differentiates normal images from soft or textured contact lenses.
- DFCANet is validated on challenging cross-domain settings, namely, cross-dataset, inter-sensor, and inter-subject settings. DFCANet also achieves good general-

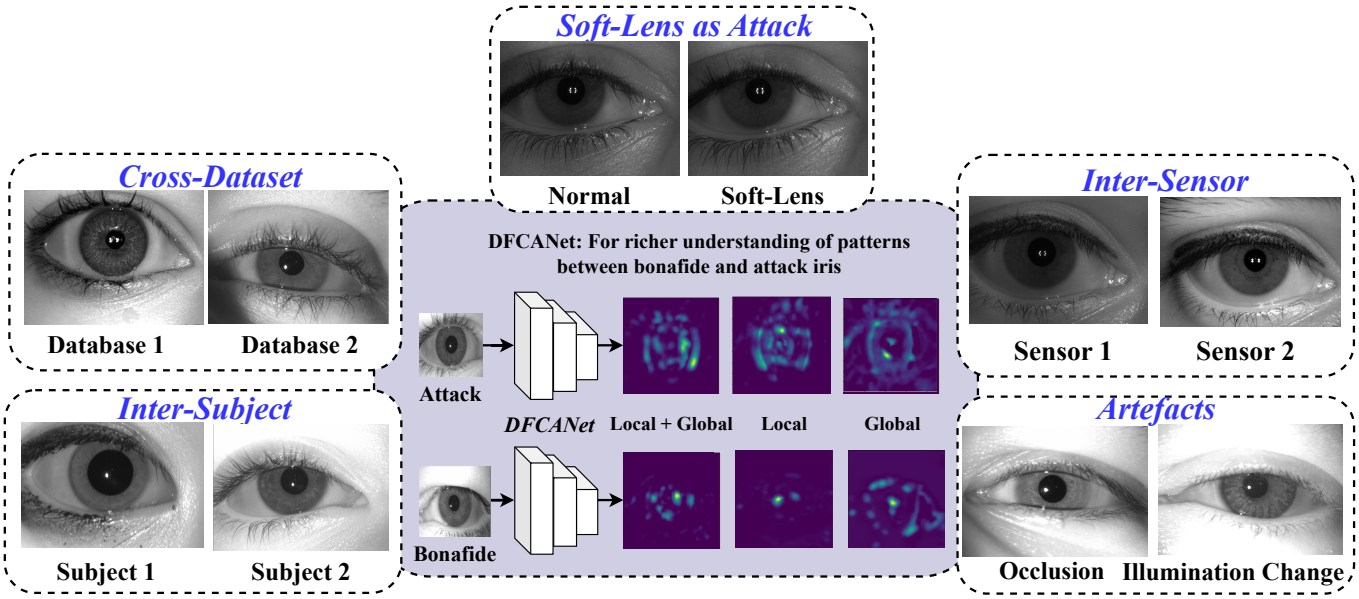


Fig. 1. Iris Presentation Attack Detection involves challenging cross-domain evaluation (Cross-Dataset, Inter-Sensor and Inter-Subject) as well as Soft-Lens situation as attacks and artifacts. The proposed DFCANet generalizes well over these settings by virtue of joint local-global feature modeling.

ization performance in contact lens detection.
- Our incremental learning for low-data regimes works well on the challenging NDCLD-13 dataset.

The remainder of the paper is structured into four main sections. Section II critically reviews the existing work on IPAD with an emphasis on deep learning-based methods. Section III presents the methodology of our approach and includes the details of the network and training considered in this work. Section IV provides details about the data sets, experiments, and testing protocol used for performance evaluation. In Section V, we present the parameters and the runtime analysis. Key findings and discussions are summarized in Section VI.

II. Related Work

In recent years, many studies on iris biometrics have started to employ deep learning schemes [10], [11] and presented remarkable progress in IPAD performance. The hand-crafted features have also shown significant progress in IPAD, specifically for intra-dataset testing scenarios. However, their progress is still far from satisfactory to new application scenarios [12]. The well-known handcrafted features that have been used in IPAD include hierarchical visual codebook [13], local binary pattern [14], weighted local binary pattern [15], spatial pyramidal matching [16]. In contrast, IPAD based on deep learning approaches [8], [17], [18] work well due to their ability to extract highly domain specific iris representations. The first work that proposed a deep architecture for IPAD was called SpoofNet [17]. Recently, [19] attempted to explore iris liveness detection and contact lens identification in near-infrared images. Moreover, some of the schemes explored the combination of hand crafted features with deep learning and achieved good results [7], [11]. Unlike fusion of hand crafted and CNN features, authors in [20] presented a multi-layer deep fusion scheme extracting different level of information from multiple layers of the network. Similarly, in [21], the 2D (textual) and 3D (shape) features of the iris image combined to address the problem of spoof detection. Apart from fusion schemes, most recently, attention based deep learning framework named pixel wise binary supervision network [8] was given to capture fine grained pixel level information that can be emphasized for making accurate IPAD decisions. In [18], authors presented an explainable attention-guided IPAD that can improve both the generalization and explanation capability of existing approaches. The existing IPAD approaches mainly employ LivDet-Iris datasets [5] by performance evaluation of independent algorithms [4]. Authors in [22], presented DensePAD method to detect presentation attacks by utilizing DenseNet-121 architecture. Similarly, [10], also exploited the architectural benefits of DenseNet to propose an IPAD scheme that was tested on four different sub-datasets of the LivDet-Iris 2017 [5] and LiveDet-Iris 2020 [4] datasets. Although their method achieved good results on LivDet-Iris 2017 datasets [5], but performance dropped in the case of cross-dataset scenarios. In recent work [23], authors presented a generalize IPAD to resolve degrade in performance of DNN's against unseen dataset, unseen sensor, and different imaging conditions. Likewise in [24], authors addressed the problem of domain shift in IPAD algorithms under cross dataset scenarios. Some other studies have been developed to create synthetic iris-like patterns using a generative adversarial network [25] with applications to IPAD. However, these methods are not scalable to more than two domains and often show instability in the presence of multiple domains. [26] proposed domain invariant styling n/w to generate good quality iris synthetic images to assist in training IPAD models for better performance. Some recent studies have been reporting survey papers [1], [2], [27], new datasets [28] and conducting IPAD competitions [4], [5],

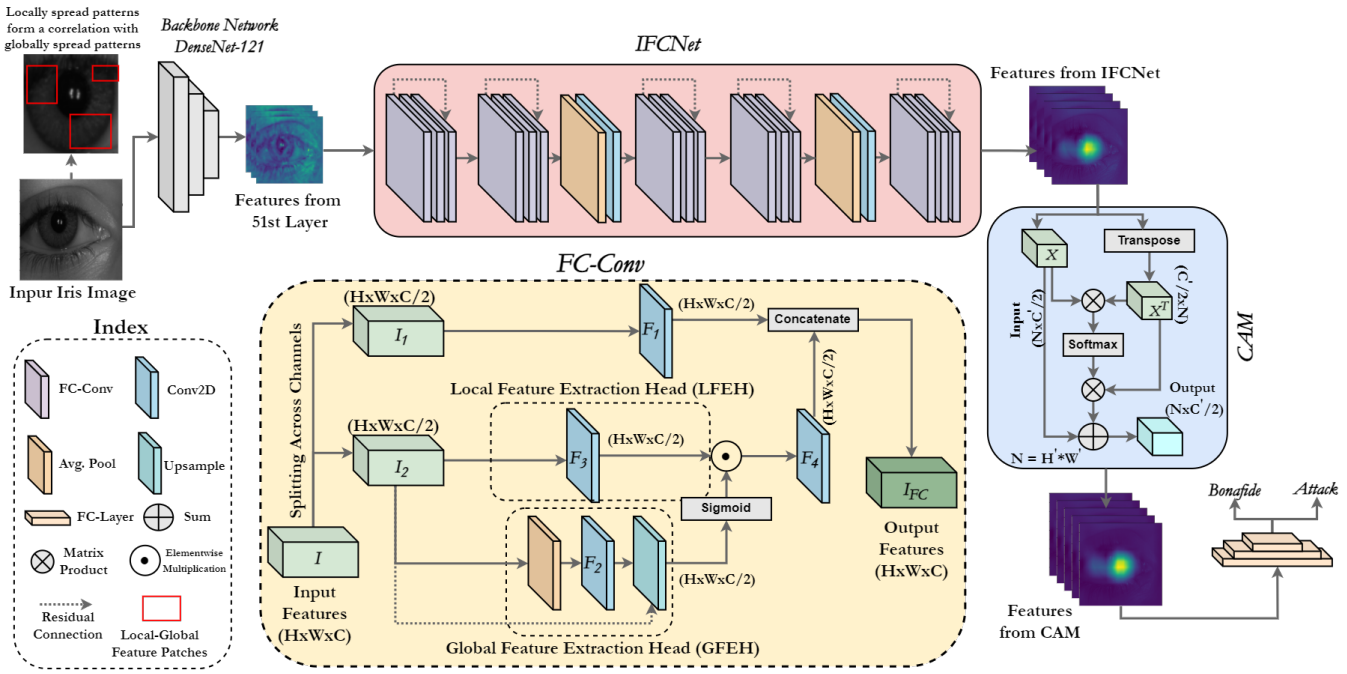


Fig. 2. The proposed DFCANet has three main blocks namely, the DenseNet-121 backbone network, the Iris Feature Calibration Network (IFCNet) (Sec. III-A) and the Channel Attention Module (CAM) (Sec. III-B). A key feature of IFCNet is the 5 Feature Calibration Convolution (FC-Conv) module for local and global feature extraction, which this diagram also shows in detail. The CAM output is used for the bonafide versus attack classification.

[29] that has increased the potential of DNN based IPAD approaches under constrained environments. Recent trends in IPAD are to mix synthetic and actual data to make the network more domain-specific [9], [26], [30], and to improvise performance on unseen domains [31] and open set scenarios [32]. Limited research was presented on gender bias solutions in IPAD [33], [34], which has a lot of potential to advance the state-of-the-art in this domain. In summary, future IPAD systems must be generalized well to attenuate illegal access to biometric systems by leveraging deep learning algorithms.

III. Methodology

In the Section I, we highlighted three design consideration points for IPAD networks, basing on the same we hypothesize the following: i) Iris contains the similar pattern in local neighborhoods. These patterns are spread across the entire iris. ii) These correlated locally spread patterns constitute a globally homogeneous pattern. iii) Jointly at local and global scales, there is a significant difference between activations of attack and bonafide samples (illustrated in Fig. 1). The DFCANet architecture (Fig. 2) accordingly has 3 major components: a DenseNet121 backbone, the Iris Feature Calibration Network (IFCNet), and the Channel Attention Module (CAM).

A. IFCNet and Backbone Network

DFCANet extracts preliminary but generic iris features from a pre-trained backbone network (fine-tuned during training). Instead of other conventional backbone networks (ResNet [35], Inception [36]), the choice of DenseNet-121 [37] has been motivated by its prevalent usage in IPAD literature [10], [18] as well as the benefits of dense feature-flow. The dense connections in DenseNet-121 enable interactions between locally-learnt patterns at lower layers, to be robustly correlated with globally located representations at higher layers. We input $224 \times 224 \times 3$ iris images without segmentation or normalization. Features from the 51^{st} layer are extracted, resulting in output dimensions being $R^{(56\times56\times128)}$. IFCNet (Fig. 2) takes this as its input. IFCNet has 5 FC-Blocks stacked. A FC-Block comprises Feature Calibration Convolutions (FC-Conv) [38] with residual connections.

FC-Conv: Local-Global Features for IPAD Domain-Specificity: FC-Conv (Fig. 2) facilitates the tuning of local features (extracted in a region restricted by the kernel size), in accordance with a global feature hierarchy. The input $I \epsilon R^{H\times W\times C}$ is split across the channel dimensions into two equal halves I_1 and I_2. Height, width and channel dimensions are respectively, H, W and C. The first half operates through a regular 2D-Convolution, while the second half is made to go through both local feature extraction head (LFEH) and global feature extraction head (GFEH) in a parallel manner.

Let F_1 represent the first convolution over I_1. The corresponding output is given by $I_1^{'} \epsilon R^{H\times W\times C/2}$. This convolution learns non-linear relationships across one half of the channels. In LFEH, I_2 is operated using a regular 2D-Convolution F_3. This is a *local convolution* because of its localized receptive field. The resulting output is $I_{2-Local}^{'} \epsilon R^{H\times W\times C/2}$. Simultaneously in GFEH, I_2 is down-sampled using Average-Pooling ($Avg.Pool()$), followed by a 2D-Convolution (F_4) and finally, bilinear upsampling ($Up(.)$). Reduction in spatial dimensions with average pooling lets the model develop representations on the basis of globally compressed information. The output

of GFEH $I^{'}_{2-Global}$ is

$$I^{'}_{2-Global} = Up(F_2(Avg.Pool(I_2))); I^{'}_{2-Local} \epsilon R^{H \times W \times C/2} \quad (1)$$

Next, the output of the upsampling is added with I_2 i.e., the original input to LFEH and GFEH. This ensures stability in flow of gradients as well as highlights important features. The generated output is activated through a sigmoid (σ) to undermine the essential spatial regions with respect to understanding global characteristics. Finally, in the feature calibration step, element-wise multiplication ($\otimes$) of the outputs of GFEH and LFEH is performed. This operation highlights the important local features with respect to the global feature representation. Another 2D Convolution (F_4), succeeds the same, and as a result generates $I^{'}_2 \epsilon R^{H \times W \times C/2}$. The last convolution ($F_4$) facilitates extraction of joint features over locally and globally correlated representations. The entire operation can be mathematically summarized as follows:

$$I^{'}_2 = F_4(I^{'}_{2-Local} \otimes \sigma(I^{'}_{2-Global} + I_2)); I^{'}_2 \epsilon R^{H \times W \times C/2} \quad (2)$$

The final output of FC-Conv is the concatenation of $I^{'}_1$ and $I^{'}_2$ ($Y_{FC} \epsilon R^{H \times W \times C}$). The convolutions F_3 and F_2 are activated by ReLU while the convolutions F_1 and F_4 are activated linearly but are succeeded by batch normalization and ReLU activation. Furthermore, F_1, F_3 and F_4 adhere to the same kernel size k_1. F_2 has a bigger kernel-size k_2 for capturing greater global contexts. Accredited to a limited number of channels per kernel, the parameter efficiency of FC-Conv is significantly higher than regular 2D Convolutions. Hence, FC-Conv allows for stacking multiple layers and in-turn, captures highly non-linear relationships. Table I describes the architecture of IFCNet.

TABLE I
DESCRIPTION OF DIFFERENT PARAMETERS OF IFCNET

Model	Sub-module	k1	k2	Avg. Pooling	o/p channels
FC Block1	FC Conv1	(3,3)	(7,7)	(11,11)	128
	FC Conv2	(3,3)	(7,7)	(11,11)	128
	FC Conv3	(3,3)	(7,7)	(11,11)	128
FC Block2	FC Conv1	(3,3)	(7,7)	(11,11)	128
	FC Conv2	(3,3)	(7,7)	(11,11)	128
	FC Conv3	(3,3)	(7,7)	(11,11)	128
Avg. Pool 2D	Pool Size = (2,2)	-	-	Strides = (2,2)	256
(1×1) Conv2D	-	-	-	-	256
FC Block3	FC Conv1	(3,3)	(5,5)	(9,9)	256
	FC Conv2	(3,3)	(5,5)	(9,9)	256
	FC Conv3	(3,3)	(5,5)	(9,9)	256
FC Block4	FC Conv1	(3,3)	(5,5)	(9,9)	256
	FC Conv2	(3,3)	(5,5)	(9,9)	256
	FC Conv3	(3,3)	(5,5)	(9,9)	256
Avg. Pool 2D	Pool Size = (2,2)	-	-	Strides = (2,2)	-
(1×1) Conv2D	-	-	-	-	512
FC-Block5	FC Conv1	(3,3)	(3,3)	(7,7)	512
	FC Conv2	(3,3)	(3,3)	(7,7)	512
	FC Conv3	(3,3)	(3,3)	(7,7)	512

B. Channel Attention Module (CAM)

Channel attention accentuates feature discriminability through weighing essential channels and suppressing others (and stabilizes training, as well). CAM [18] (Fig. 2) takes in a $X \epsilon R^{N \times C\prime}$ feature map, where $C\prime$ and N respectively represent channel dimensions and spatial axes clubbed in a single dimension. Then, the input and the corresponding transpose are multiplied and have a Softmax activation. For the resulting activation map $U \ \epsilon \ R^{C\prime \ \times \ C\prime}$, each of the elements U_{ij} is a weight measuring the importance of j^{th} channel with respect to i^{th} channel. At broader level, U captures inter-channel interactions. Finally, matrix product between input (X) and activation map (U) is performed. The output is reshaped in image tensor format and the original input feature map is added. The CAM output comprises channels scaled in accordance with their respective distinctiveness.

C. Output Network and Training Details

CAM features are followed by global average pooling and a couple of fully connected layers. During backpropagation, Adam optimizer is employed, while images are augmented with zooming, shearing, spatial shifts and brightness variations. Training is performed for 200 epochs and the model achieving maximum validation (testing) accuracy is chosen and is saved. The proposed framework has been implemented using TensorFlow 2.5 while all the experimentation has been conducted in NVIDIA's RTX-3090 GPU.

IV. EXPERIMENTS AND DISCUSSIONS

In this section, we have conducted experiments under challenging strategies including: i) cross-sensor, ii) intra-sensor and iii) combined sensor. Furthermore, to demonstrate the effectiveness of the IFCNet and CAM structures independently, the ablation experiments are conducted and compared with the proposed framework. Comparisons with other state-of-the-art methods are also conducted. In addition, we introduce baseline experiments based on incremental learning are performed. To have a better check on generalizability, we also performed cross-dataset and contact-lens detection experiments.

A. Datasets and Evaluation Protocols

The proposed framework is evaluated on seven benchmark datasets comprising various types of presentation attacks captured by a variety of sensors. This section can be passed over, without any loss in continuity.

1) IIITD-CLI dataset [39] and Evaluation Protocol: This dataset is composed of 6750 dual-eye ocular images collected from 101 subjects. Two eye-scanners: cogent and vista F2AE have been employed for the task. The dataset comprises contact lens-based attacks i.e. soft and textured lenses. Following the literature - A-PBSNet [8], MVANet [23], ELF [9]], we first exclude soft lenses for conducting comparison with the state-of-the-art. However, we also perform other experiments considering Soft-Lens-as-Attack (SLA). The following instance (SLA) is challenging as it gives a very similar appearance to a bonafide iris image. This scenario was taken into account because of two reasons: i) There is a high possibility that soft-lens can be used as a spoofing medium, ii) Establishing a baseline over a challenging question that can an IPAD algorithm effectively differentiate between bonafide images

and any form (soft or textured) contact lenses. For conducting experiments with this dataset, 50-50% data was split for training and testing in a hold-out fashion.

2) NDCLD'13 dataset [40] and Evaluation Protocol: This dataset comprises two datasets ND-I and ND-II collected from two different sensors, AD100 IrisGuard and LG4000 respectively. Similar to IIIT-CLI, this dataset comprises bonafide iris images and soft, textured contact lens images. Keeping consistent with literature [A-PBSNet [8], MVHF [7]], we have considered textured and soft contact lenses. To be specific, we have considered Soft-Lens-as-Bonafide (SLB) for comparison with state-of-the-art, while for the remaining experiments, we considered Soft-Lens-as-attack (SLA). Both ND-I and ND-II have been provided with a training and testing set. ND-I consists of 600 images in training, and 300 images in testing. For the case of ND-II, there are 3000 images given under training and 1200 images for testing.

3) IIITD-CSD dataset [41] and Evaluation Protocol: IIITD-CSD consists of print, scan, textured and soft contact lens based attacks. It is the largest dataset considered in this study. In total, it consists of 17036 ocular-iris images. For eye scan, Cogent and Vista F2AE sensors had been utilized while for print and scan attacks HP Flatbed Optical Scanner and Cogent-CIS 202 was used. The experimental protocol considered Soft -lens-as-Bonafide (SLB) [DESIST [41], MVHF [7]] for comparison with state-of-the-art. However, for the other experiments soft-lens as attack (SLA) was taken into account. Experimentation in both SLA and SLB involved 50-50% splitting for training and testing set in a holdout fashion.

4) IIITD-WVU dataset [5] and Evaluation Protocol: The IIITD-WVU is a subset of LivDet-2017 Challenge, and presents one of the most challenging cross-dataset scenarios for IPAD algorithm evaluation. The dataset consists of 6250 training iris images captured via Cogent, CIS202 and Vista F2AE sensors in constrained environments. For evaluation, 4209 images are captured for a different set of subjects and also via different sensor (Iris Shield MK2120U). The testing images have been captured in both constrained (indoor) and unconstrained (outdoor) settings to further deepen the challenge. The attacks present in the dataset includes contact lenses and printouts for both real and contact lens iris. Since the dataset does not have soft contact lenses, we have only evaluated the proposed model over the provided testing dataset.

5) Clarkson-17 dataset [5] and Evaluation Protocol: The Clarkson-17 is a subset of LivDet-2017 Challenge, and presents the challenging unseen patterned lens based attacks. The training set consists of three types of iris images: a 2469 number of live images, a 1346 number of printed iris images, and a 1122 number of patterned contact images. There are in total 3158 images from spoof (print and patterned lens attacks) and bonafide classes.

6) NDCLD15 dataset [42] and Evaluation Protocol: This dataset has a total number of 7300 images that were captured by two sensors, IrisGuard AD100 and IrisAccess LG4000. The images were collected under MIR illumination and controlled environments. Alongside NDCLD'13, this is a next data version in NDCLD series which contains three type of iris images, namely: wearing no lenses, soft lenses, and textured lenses. Keeping consistent with literature [42], we have considered 6000 iris images randomly for training and 1300 iris images for testing.

7) Clarkson-15 dataset [5] and Evaluation Protocol: This dataset was captured using two types of sensors: Dalsa and LG sensors, where image capture characteristics differ significantly. There are a total of 1078 live images, 1431 patterned contact lens images, and 1746 printed images are in this dataset. The training subset contains 700 live, 873 patterned contact lens, and 846 printed images. The testing subset contains 378 live, 558 patterned contact lens, and 900 printed images.

B. Performance Metrics

We use state-of-the-art metrics [7], [8] and ISO standards [43]: Attack Presentation Classification Error Rate (APCER), Bonafide Presentation Classification Error Rate (BPCER), and Average Classification Accuracy (AA).

TABLE II
INTRA-SENSOR EXPERIMENTAL RESULTS

Datasets	Training	Testing	AA	APCER	BPCER	ACER
IIITD-CSD	Cogent	Cogent	99.33	2.03	0.31	1.17
	Vista	Vista	99.80	0.40	0.17	0.27
IIITD-CLI	Cogent	Cogent	98.90	1.52	0.87	1.19
	Vista	Vista	99.79	0.41	0.90	0.25
NDCLD13	LG4000	LG4000	95.91	9.75	1.25	5.50
	AD100	AD100	76.00	44.00	14.00	29.00

C. Experimental Results and Discussions

1) Intra-Sensor Experiments: Intra-Sensor experiment involved training the model on data from a sensor, while testing it on the data from the same sensor. Results obtained under this setting has been tabulated in Table II . The proposed DFCANet attains significant performance in all the three datasets. For the Vista trained and Vista tested model on IIITD-CLI dataset, ACER obtained is 0.25% and is the lowest amongst all the models. A comparative performance is observed for the Vista trained and Vista tested model of the IIITD-CSD dataset. In NDCLD'13 dataset, for the model trained-tested on LG4000 dataset, the obtained average accuracy of 95.91% is significant. However, the same model is comparatively less robust in classification of bonafide attempts. In fact, it is evident that when the proposed model is trained and tested on similar sensor, there is low APCER but lower BPCER i.e., the model is more robust for detecting spoofs. From the mentioned observation, it can be inferred that the decision boundaries learnt by the model is more spanning in the attack class's region. A possible reason for the same can be multi-cluster and multi-center data formations of the attack class. It can also be concluded from the table that the model is better for Cogent and Vista sensors, when compared with LG4000 and AD100. A specific reason for the same can be sensing inconsistencies and cross-user evaluation in NDCLD'13 dataset. Particularly, the data from AD100 sensor comprises of just 600 images and

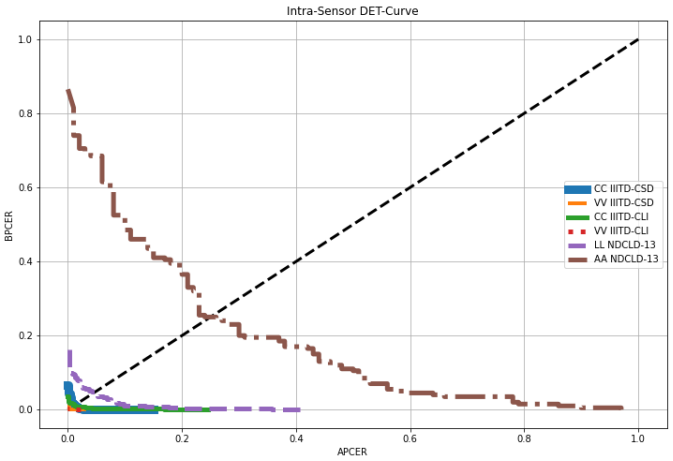


Fig. 3. DET plot for comparing different intra-sensor experiments

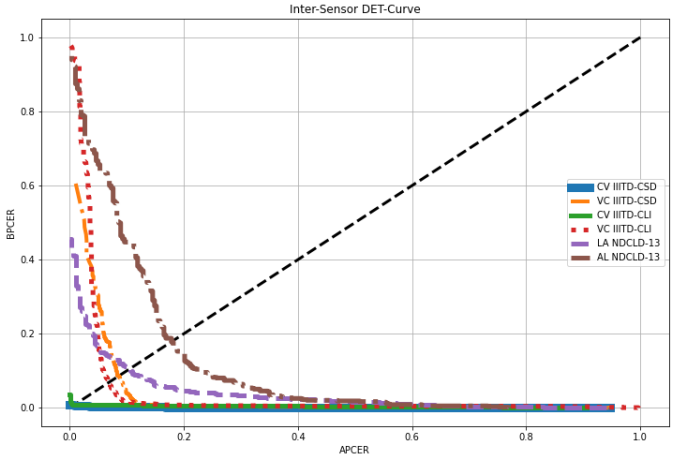


Fig. 4. DET plot for comparing different inter-sensor experiments

TABLE III
INTER-SENSOR EXPERIMENTAL RESULTS

Datasets	Training	Testing	AA	APCER	BPCER	ACER
IIITD-CSD	Cogent	Vista	99.33	0.41	0.72	0.56
	Vista	Cogent	96.49	13.98	0.87	7.42
IIITD-CLI	Cogent	Vista	99.19	0.31	1.04	0.67
	Vista	Cogent	95.52	10.10	1.64	5.87
NDCLD'13	LG4000	AD100	90.22	13.00	8.16	10.58
	AD100	LG4000	85.88	25.74	8.20	16.97

hence in the same case, highly limited training quantities do not match with the architectural depths of the DFCANet.

2) Inter-Sensor Experiments: Inter-sensor experiments have been conducted to validate the performance of the proposed model in environments contrasting to the training. Training-testing strategy in this experiment involved utilization of entire data from one sensor in training while from the other in testing. Table III tabulates performances achieved by various models under this strategy. It can be concluded from the results that DFCANet is robust enough to capture the intricacies in cross sensor settings. Specifically, for the case of IIITD-CLI and IIITD-CSD the performance obtained is encouraging. The models tested on Vista sensor obtain ACER as low as 0.56% and 0.67% respectively. Fig. 4 illustrates the comparative performance of inter-sensor based experimental results using DET curves. Similarly, the average accuracy's attained by the same is significantly higher. The drop in performance for this challenging experimental setup is not stark when compared with the Intra-Sensor experiments. A point to highlight is that unlike Intra-sensor experiments, the model attains superlative performance in IIITD-CSD dataset when compared with IIITD-CLI dataset. It can also be inferred that for most of the models, APCER is relatively higher than the BPCER. For NDCLD'13 dataset, this scenario is quite challenging for the model although the average accuracies obtained are significantly high. For training on AD100 and testing on LG4000, we have used all the 900 examples from AD100, but while testing, only 900 random examples from the LG400 dataset were sampled. In such low-data regime, DFCANet attains significant average accuracy of 85.88%.

TABLE IV
COMBINED-SENSOR EXPERIMENTAL RESULTS

Dataset	AA	APCER	BPCER	ACER
IIITD-CSD	98.02	2.43	1.54	1.99
IIITD-CLI	99.09	1.69	0.51	1.10
NDCLD13	93.00	8.00	6.50	7.25

3) Combined Sensor Experiments: Combined sensor experiments involve training the model on 50% data of available sensors while in testing the remaining 50% of the data is used. These experiments elucidate the overall performance attained by a model in a specific dataset. The results from this experiment have been tabulated in Table IV. Performance of the IIITD-CLI trained model is the most optimal. A low ACER of 1.10% and high accuracy of 99.09% for the same represents the above-forth mentioned. For IIITD-CSD, and NDCLD'13, the respective performances obtained are also significant. Since NDCLD'13 involves cross-subject evaluation, the obtained performance is relatively lower than the other two datasets. There is an overall increment for NDCLD'13 in comparison with inter and intra-sensor experiments, the reason for the same is expanded and more directed training data. Similarly, to previous models (inter-sensor and intra-sensor) APCER remains higher than BPCER. This suggests that the model is more robust in classifying attacks. The inferences drawn from the numerical metrics of this experiment also coincide with the respective DET plot (Fig. 5). The plot highlights significant performance obtained by DFCANet on IIITD-CLI and IIITD-CSD datasets.

TABLE V
INCREMENTAL LEARNING EXPERIMENTAL RESULTS

Experiment	Training	Testing	AA	APCER	BPCER	ACER
Intra-sensor	AD100	AD100	80.33	36.00	11.50	23.75
	LG4000	LG4000	94.66	6.75	5.62	5.68
Inter-sensor	AD100	LG4000	85.44	15.18	14.23	14.70
	LG4000	AD100	90.22	21.60	4.16	12.53
Combined	AD100+ LG4000	AD100+ LG4000	93.93	11.20	3.50	7.35

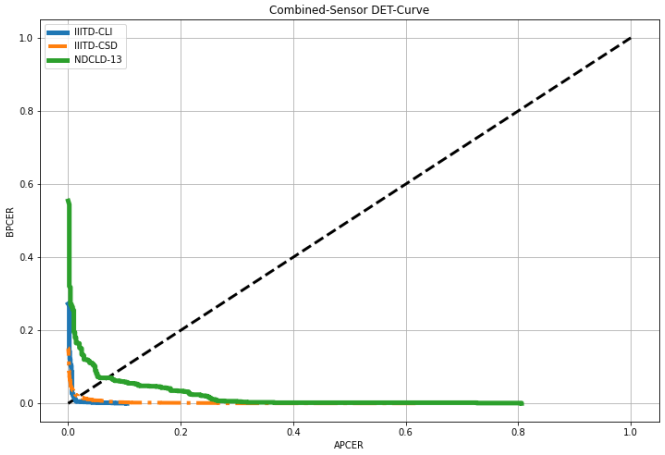


Fig. 5. DET plot for comparing combined sensor experiments

4) Incremental Learning Experiments: To bring about gains in performance for NDCLD'13, we leveraged the IIITD-CLI pre-trained model to fine-tune with NDCLD'13. As both IIITD-CLI and NDCLD'13 are composed of congruent characteristics, we chose the combined sensor model from IIITD-CLI for transfer of its weights. A particular reasoning behind this is that combined sensor model has been trained upon the most diverse dataset; therefore, its generalizability will be very high. Table V presents experimental results. From the same table (row 1), when compared to setting without transfer (Table II, row 6), a prominent gain of 4.33% in average accuracy for the Intra-Sensor experiment of AD100 is observed. Similarly, there is an increase for the combined setting. Then again, there is an incremental gain in ACER of approximately 2. 27% for the Inter-Sensor experiment of AD100 training and LG4000 testing, compared with the non-incremental counterpart (Table III, row 6). These enhancements in performance can be accredited to transfer of iris feature understandings by the IIITD-CLI combined sensor model. Nevertheless, there is a slight decrement in performance for LG4000 model for Intra- and Inter-sensor experimentation. A possible reason can be difference in morphological and textural properties of data in LG4000 and IIITD-CLI. To validate the significance of incremental learning, we perform a t-SNE based analysis in Appendix 1 (Supplementary Material). It can be inferred from the same that for incremental-learning embeddings from bonafide and attack classes are more separated as well as more clustered within the same class

5) Cross-Dataset Experiments: To validate efficacy in the proposed model, we have conducted experiments in the cross-dataset setting. In this set of experiments, the use of NDCLD'13 and IIITD-CLI datasets has been carried out, with the reason being similar characteristics for attacks between the two. The experimental setup involved training the model with complete data of three sensors while testing on the remaining one. Table VI, encompasses the results obtained in this experiment. The highest performance in this setting is achieved when the Vista sensor is evaluated while the other three (Cogent, LG4000, and AD100) are used in training. The following model achieves correct classifications of 98.66% while maintaining a low ACER of 1.21%. It can be concluded from the table that whenever testing is performed on the IIITD-CLI dataset, results are significant, but the challenge remains for testing on any of the NDCLD13 sensors. From the results obtained in Incremental Learning Experiments and Cross-Dataset experiments, it can be inferred that LG4000 has different image properties than the other three. This is not the case with the AD100 sensor, as including it in training helps the model overcome the training subtleties and, in turn, leading to enhanced performance.

TABLE VI
CROSS-DATASET EXPERIMENTAL RESULTS

Training	Testing	AA	APCER	BPCER	ACER
Cogent+Vista+ LG4000	AD100	84.44	37.33	4.66	21.00
Cogent+Vista+ AD100	LG4000	73.76	67.68	5.46	36.22
Cogent+AD100+ LG4000	Vista	98.66	0.94	1.48	1.21
Vista+LG4000+ AD100	Cogent	95.71	6.21	3.20	4.71

D. Ablation Study on use of IFCNet and CAM

In order to validate contribution in performance by each of the individual components of DFCANet, an ablation study has been conducted. We removed one component at a time and trained the model over the challenging NDCLD'13 combined sensor dataset. Results of this study can be referenced from Table VII. It can be inferred from the following table that DFCANet outperforms all other models except the individual IFCNet. The main reason for the same is DenseNet extracting generic but not domain-specific features. The problems becomes more intricate with learning being performed over limited training data from NDCLD'13 dataset. This forbids DenseNet to generate better representations for IFCNet to calibrate at local and global scales. In contrast, the IFCNet operates individually and leverages domain-specific learning to attain high-end performance. However, DFCANet is optimal in terms of average accuracy by 0.60% than the next-best performing model while it attains a significantly low BPCER and APCER. The role of CAM in highlighting the important channels for improving representation can be observed from comparison between DFCANet and DenseNet+IFCNet model. In the same comparison there is a gain of 2.65% in terms of ACER when CAM is introduced in the pipeline. A contrary case to this is also observed when CAM is utilized along IFCNet. To be specific, for this case there is a significant drop in performance (both ACER and Avg. Accuracy) in comparison to DFCANet. This can be accredited to CAM, which emphasizes less attention to more discriminative channels. Few other insightful observations of this ablation study are with regard to properties of IFCNet. Firstly, for IFCNet there is a performance gain of 2.66% and 0.90% in Average Accuracy and ACER, respectively, compared to DenseNet. Secondly, whenever it is removed from the pipeline there is a degradation in performance while increment in ACER. It is

also worthwhile mentioning that when DenseNet is not used in combination with IFCNet there is significant increment in ACER. However, when DenseNet is replaced with ResNet50 [35], a performance drop is observed with respect to both Average Accuracy and ACER. This can be attributed to improved feature flow within DenseNet compared to ResNet.

TABLE VII
ABLATION STUDY ON THE NDCLD'13 DATASET

Model	AA	APCER	BPCER	ACER
DenseNet	92.20	7.00	8.20	7.60
DenseNet + IFCNet	92.40	5.20	8.80	6.99
DenseNet + CAM	90.20	10.20	9.60	9.90
IFCNet	94.66	10.80	2.60	6.70
IFCNet + CAM	90.20	13.60	7.90	10.75
ResNet50 + IFCNet + CAM	92.40	8.00	7.74	7.77
DFCANet	93.00	8.00	6.50	7.25

TABLE VIII
IFCNET AND ATTENTION-BASED MODELS: A COMPARISON

Model	AA (%)	APCER (%)	BPCER (%)	ACER
Comparison with Local and Global IFCNet variants				
IFCNet-local	94.00	5.08	6.50	5.75
IFCNet-global	88.13	24.00	5.80	14.90
Comparative Study over significance of FC-Conv				
Conv2D	87.60	23.00	7.10	15.05
SENet [44]	87.26	27.60	5.30	16.45
CBAM [45]	83.80	29.60	6.50	5.75
GALA [46]	66.66	100.00	0.00	50.00
ViT [47]	79.73	25.80	17.50	21.00
GENet [48]	91.06	19.20	3.80	11.50
IFCNet (Proposed)	**94.66**	10.80	2.60	**6.70**

E. t-SNE Representation and Analysis

Using t-SNE plots, a visual demonstration over decision boundaries and class discrepancies has been presented in Fig. 6. This has been illustrated for the models corresponding intra and inter-sensor experiments. From the trained models, embeddings of the test set were extracted from the last fully connected layer and fed t-SNE. From the plots, stark separative boundaries between bonafide and attack samples can be observed. Both attack and bonafide embeddings are tightly clustered within themselves while their variance is evident. Also, the multi-center nature of the attack samples can be inferred from the plots. The following characteristics of attack samples is induced due to the presence of soft-textured lenses in the class. This elucidates about the added challenge in Soft-Lens-as-Attack (SLA) scenarios. At the same time, the well separated plots give evidence of DFCANet's robustness. As shown, the t-SNE plots in Fig. 6 (a, b, c) of intra-sensor experiments have more distinctive nature in comparison to inter-sensor plots in Fig. 6 (d, e, f). With respect to the datasets, NDCLD'13 appears slightly challenging, because of its cross-user evaluation. Nevertheless, a significant discrepancy between attack and bonafide samples is also observed within both inter- and intra-sensor experiments of NDCLD'13.

F. Ablation Study: Local-Global Feature Calibration

The ablation study is at two levels. First, we analyze the impact of the local and global feature extraction heads of IFCNet. Next, we study the significance of the FC-Conv mechanism in IFCNet (Sec. III-A) over state-of-the-art attention-based modules. For both of these studies, we perform evaluation over the combined-sensor setting of NDCLD'13 dataset while adhering to SLA protocol.

1) Impact of Local and Global Feature Extraction Heads in IFCNet: For this ablation, we consider three variants of IFCNet, i.e., i) IFCNet-Local: It follows the same architecture of IFCNet but from each of the FC-Conv, GFEH is removed. ii) IFCNet-Global: Similar to IFCNet-Local this also encompasses the same architecture of IFCNet but in FC-Conv, LFEH is not present. iii) The proposed IFCNet model (shown in Figure 2). Further, to investigate the quality of the features learned, we compare the feature map outputs of each of these models (see Fig. 7 (a)). Along with this, we also analyze the outputs of Grad-CAM [50] in Fig. 7 (b).

In Table VIII, we have evaluated the results of this ablation, from the same it can be inferred that when both LFEH and GFEH are utilized in FC-Conv, performance gains (in average accuracy) of 0.66% and 6.53% are observed for IFCNet-Local and IFCNet-Global, respectively. Owing to capture of only global information, IFCNet-Global is unable to learn minute morphological patterns and curvilinear features, which are essential for IPAD. On the other hand, high-end performance attained by IFCNet-local can be accredited to its ability of understanding features at local-scales and the hierarchical convolutional architecture. Particularly, this is because these locally-extracted representations are correlated globally with increased receptive field. Next, we analyze the feature maps (7(a)) and from the same following inferences were drawn: The activations of IFCNet are formulated in joint correspondence with a combination of local-global features. For the case of IFCNet-local, at initial layers regions which carry discriminative patterns are highlighted (refer to the first row and fourth row of Figure 7), while for the case of IFCNet-Global, as expected representations are generated with respect to overall spatial contexts (refer to the second row and fifth row of Figure 7). In contrast, IFCNet from the initial layers emphasizes edges and textures. It is also evident that IFCNet-Global learns smoother features which correspond to an overall morphology and spatial-pattern distribution. When considering IFCNet-Local, minute discriminative details are not highly activating for the later layers. A possible reason for the same is IFCNet-Local not being able to correlate multi-scales. In order to strengthen the above-mentioned points, we analyze the gradient-based activated regions in the iris images using Grad-CAM. This allowed us to look into the iris regions that are discriminative for the model. It is clear from Figure 7(b); (row 3 and 6), that IFCNet identifies those regions as discriminative that are of significance between both local and global representations. In contrast, IFCNet-local Figure 7(b); column 1) emphasizes learning features from smaller image patches; therefore, it is unable to correlate the representations between these patches. This makes the model focus on limited spatial contexts. IFCNet-global (in Figure 7(b); column 2) on the other hand, gets activated globally and not to very specific regions. Thus, it find it challenging to emphasize on discriminative but small-sizes image patches.

TABLE IX
INTRA-SENSOR COMPARISON FOR IIITD-CLI DATASET IN TERMS OF AA (IN %)

Sensor	LBP [6]	mLBP [39]	MVANet [23]	DCCNet [19]	APBS [8]	DeepI [49]	ELF [9]	DenseNet	DenseNet (SLA)	IFCNet	IFCNet (SLA)	DFCANet	DFCANet (SLA)
Cogent	77.46	80.87	94.90	98.71	99.57	96.79	95.09	99.47	97.97	98.17	97.22	**99.21**	**98.90**
Vista	76.01	93.97	95.91	99.30	100.00	98.09	97.12	99.38	99.79	99.38	99.66	**99.59**	**99.79**

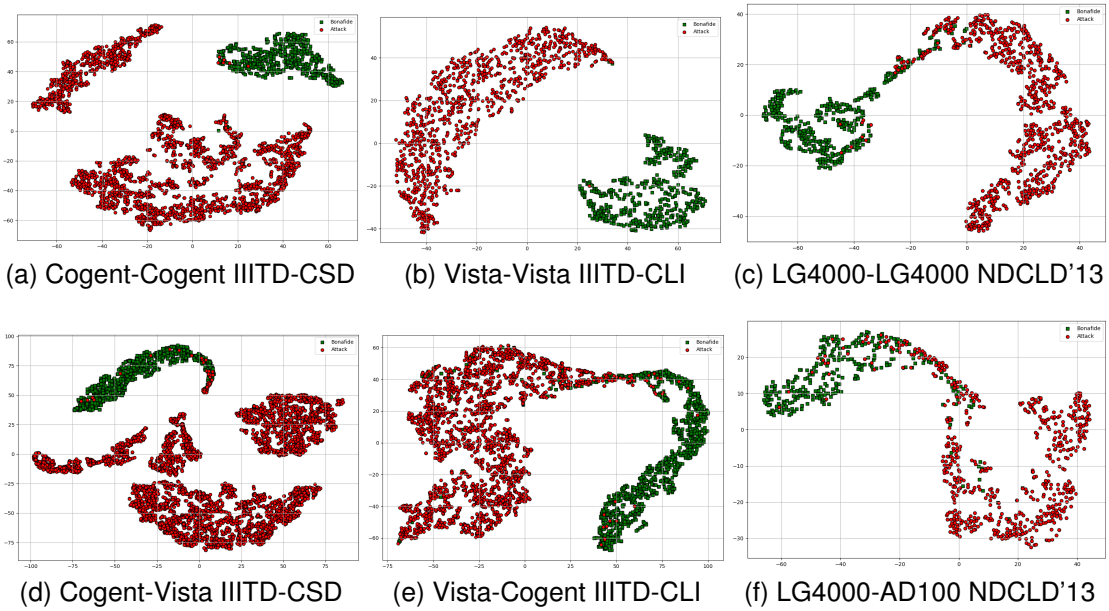


Fig. 6. Illustration of data modeling abilities of proposed DFCANet for Intra-sensor (a, b, c) and Inter-sensor (d, e, f) experiments conducted over all the three datasets: IITD-CSD, IIITD-CLI and NDCLD'13

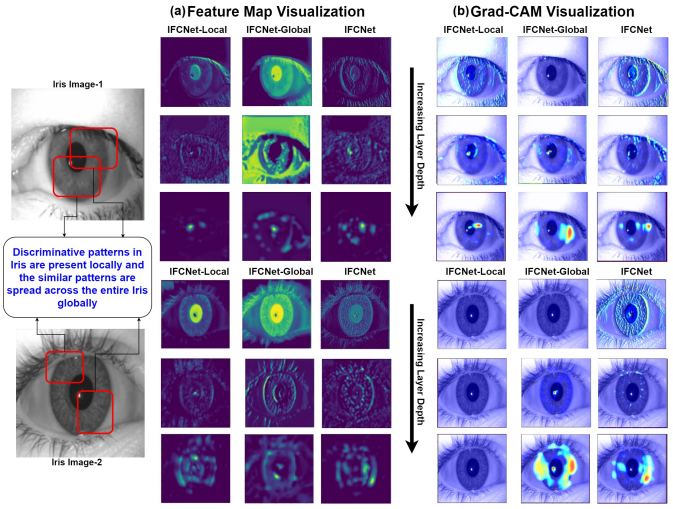


Fig. 7. Feature maps and Grad-CAM outputs for IFCNet, IFCNet-Global and IFCNet-Local (The last row in (a and b) represents final Conv-layer output).

2) FC-Conv: A Comparative Study: In this subsection, we replaced FC-Conv with different modules and the respectively formulated models were compared with IFCNet. Following modules have been used for comparison: i) Vanilla 2D-Convolution (Conv2D), ii) Attention-based: SENet [44], CBAM [45], ViT [47] (we used 4 encoder layers and 8 attention heads) and iii) Local-Global context based: GALA [46], GENet [48]. Obtained empirical results have been tabulated in Table VIII. It is evident from the same that IFCNet outperforms the other models under comparison by at least 3.60% and 4.80% in terms of average-accuracy and ACER respectively. This highlights the importance of FC-Conv. Conv2D model captures only local-contexts, while channel attention based modules like SENet, CBAM emphasizes only on global-contexts. This leads to sub-optimal performance. For ViT performing well in low-data regime is challenging because it lacks inductive-bias. While GALA and GENet operates on local and global contexts, but unlike FC-Conv they does not have channel splitting mechanism which induces diversity between both local and global features. Owing to the same, FC-Conv outperforms them by significant margins. We also illustrate Grad-CAM based comparison of this study in Appendix 2 (Supplementary Material). The same illustrates that FC-Conv-based IFCNet better emphasizes discriminative iris-features and hence outscores other models.

G. Comparative Analysis with State-of-the-Art IPAD Methods

This section presents a comparison of the proposed DFCANet with the state-of-the-art. As mentioned earlier, unlike the standard approaches [7]–[10] placing soft contact lens as a bonafide attempt, we considered it a spoofing attack also.

1) Comparison on IIIT-CLI Dataset: Table IX reports the comparative-results for intra sensor conducted on IIITD-CLI dataset. Average accuracy over Intra-Sensor experiment

TABLE X
COMBINED-SENSOR COMPARISON FOR THE IIITD-CSD DATASET

Method	APCER(%)	BPCER(%)	ACER(%)
wLBP [15]	29.94	14.39	19.66
DESIST [41]	25.58	2.47	14.02
MVHF [7]	1.82	0.72	1.27
YOLO-CNN [51]	0.72	1.81	1.26
DenseNet (SLB)	0.42	3.06	1.74
DenseNet (SLA)	8.11	0.87	4.49
IFCNet (SLB)	3.02	3.99	3.48
IFCNet (SLA)	4.85	2.15	3.50
DFCANet (SLB)	**1.71**	**1.53**	**1.62**
DFCANet (SLA)	**2.43**	**1.54**	**1.99**

of the PAD system has been considered as the comparative metric by most of the previous works, hence we have considered the same. We have made a comparison with respect to two protocols, namely i) Soft- Lens-as-Attack (SLA) and ii) Textured Lens vs Real (as endorsed by [[8], [9], [23]]). Firstly, it is evident from the comparison between SLA and textured lens vs bonafide models that for the case of the Cogent Sensor, SLA is significantly challenging. For the Vista Sensor, performance in both protocols is quite comparable. This highlights the intrinsic challenge in SLA scenarios. Considering comparison with other models in the literature, DFCANet outperforms most of the previous approaches [[6], [19], [23], [39], [49]] by a significant gap. It is also essential to mention that the proposed DFCANet performs at par with the current state-of-the-art [8], under both evaluation protocols. Unlike [8], it does not require binary masks or any other complex training mechanisms. Further, DFCANet outperforms [9] by significant margins. Although [9] highlights rich iris textures, its CNN-based feature extraction process does not involve emphasizing on IPAD domain-specific features (which our DFCANet does). Hence, we conclude high performing capabilities from these comparisons. The high-end performance achieved by the DFCANet can be accredited to its capacity to learn locally inherent iris textures and structure with respect to globally spread patterns. It can also be observed that in both the protocols, IFCNet obtains significant results. This justifies feature calibration convolutions. However, DFCANet outperforms both DenseNet and IFCNet by a margin of almost 0.93% and 1.78% respectively in the case of SLA protocol for cogent sensor, while there are gains also in the Vista Sensor. On an overall basis, the following comparative study reinforces the fact that Cogent sensor's images are more challenging to classify as compared to the Vista case.

2) Comparison on IIIT-CSD Dataset: For comparing the obtained results with previous works [7], [15], [41], [51] on the challenging dataset IIIT-CSD, DFCANet, IFCNet and DenseNet have been evaluated under combined sensor protocol over Soft- Lens-as-Attack (SLA) and Soft-Lens-as-Bonafide (SLB) scenarios. As shown in Table X, the ACER for the proposed DFCANet is as low as 1.62% and 1.99% under the SLB and SLA scenarios respectively. It is evident that the proposed model DFCANet (SLB) performs comparatively with most of the state-of-the-art models (refer Table X). DFCANet achieves 0.89%, 1.21% and 12.5% better performance than [9], [51] and [52] respectively, in terms of average accuracy. The SLA variant of DFCANet is also attains comparable performance with with previous baselines, which are trained over the less challenging SLB protocol. It is worthwhile mentioning that unlike [7], [19], [52], DFCANet does not depend upon handcrafted features or in-depth preprocessing. This grants DFCANet a low inference time and end-to-end pipeline (refer to Section V for DFCANet's inference time analysis.) Compared to individual components, that is, DenseNet and IFCNet, the proposed method DFCANet achieved an overall performance gain (in terms of ACER) of 2.50% and 1.51% for the SLA scenario. While the same are 0.12% and 1.86% for the case of SLB. Another significant observation is that IFCNet (SLB) performed extremely well compared to the previous works. Furthermore, IFCNet (SLA) also outperforms the respective DenseNet (SLA) counterpart by significant margins in ACER. This highlights the prominence of local-global iris feature understanding established by DFCANet and IFCNet. To highlight these results, DET plots for the corresponding experiment have been plotted in Appendix 3 (Supplementary material). The same illustrates several facts: i) DFCANet outperforms its individual components IFCNet and DenseNet, and ii) the SLA protocol is comparatively challenging than the SLB protocol.

3) Comparison on NDCLD'13 Dataset: In NDCLD'13, as defined in [3], [7], [8], we have first compared the performance of DFCANet using the Soft-Lens-as-Bonafide (SLB) protocol. Additionally, we have also taken the results of Soft-Lens-as-Attack (SLA) protocol from DFCANet into account. Table XI tabulates the following comparison, in which state-of-the-art works [3], [7], [8] have been followed with intra-sensor experiments. It is evident from the results that the model considering SLA attempt finds it challenging to differentiate between a soft-lens and corresponding normal images. This fact is further highlighted numerically with a drastic 5. 38% and 29. 00% decrease in ACER for the LG4000 and AD100 sensors, respectively, compared to the proposed model considering SLB. Furthermore, it is also clear that under the following challenging protocol, the proposed DFCANet generalizes well for the LG4000 sensor, but due to limited training data, the quantities for the AD100 sensor performance get degraded. However, in the SLB protocol, the proposed DFCANet achieves state-of-the-art results. For the AD100 sensor, there is a 0.00% error while for the LG4000 error remains as low as 0.12%. Since, the model is designed in accordance to SLA protocol, the performance of the proposed model for IPAD with SLB is sub-optimal and can be enhanced with dedicated hyperparameterization.

When separate components of DFCANet are considered, they also attain comparable performance. Specifically, IFCNet evaluated under both the scenarios of SLB and SLA obtains a performance gain of 32.00% and 0.74% (in ACER) respectively when compared against DenseNet counterpart for AD100 Sensor. Similarly, when APCER and BPCER are considered, a significant bias of the DenseNet model can be observed, while the same is not true for IFCNet. Intrinsic bias of DenseNet towards bonafide class is a result of relatively less domain-specific feature learning while a highly-varying dataset (NDCLD'13) contributes further to the same. IFCNet emphasizes the local-global feature hierarchy and and per-

TABLE XI
INTRA-SENSOR COMPARATIVE IPAD RESULTS ON NDCLD'13. SLB – SOFT-LENS-AS-BONAFIDE, SLA – SOFT-LENS-AS-ATTACK.

Sensor	Metric (%)	wLBP [15]	DESIST [41]	MVHF [7]	A-PBS [8]	MSA [3]	AG-PAD [18]	D-NetPad [10]	DenseNet (SLB)	DenseNet (SLA)	IFCNet (SLB)	IFCNet (SLA)	DFCANet (SLB)	DFCANet (SLA)
LG4000	APCER	2.00	0.50	0.00	0.00	0.00	1.12	1.12	1.12	2.75	0.12	4.50	**0.00**	**9.75**
	BPCER	1.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	7.00	0.25	5.12	**0.25**	**1.25**
	ACER	1.50	0.50	0.00	0.00	0.00	0.56	0.56	0.56	4.87	0.18	4.81	**0.12**	**5.50**
AD100	APCER	9.00	2.00	1.00	0.00	1.00	0.00	0.00	49.00	0.00	0.00	29.00	**0.00**	**44.00**
	BPCER	14.00	1.50	0.00	0.00	0.00	1.00	0.00	16.00	57.00	1.00	26.50	**0.00**	**14.00**
	ACER	11.50	1.75	0.50	0.00	0.50	0.50	0.00	32.50	28.50	0.50	27.74	**0.00**	**29.00**

TABLE XII
PERFORMANCE COMPARISON FOR IIITD-WVU DATASET (IN %)

Metrics	Winner [5]	Spoofnet [53]	Metafusion [11]	DNetPAD [10]	MDCDANet [54]	MLF [20]	MSA [3]	APBS [8]	ELF [9]	FAM [24]	DenseNet	IFCNet	DFCANet
APCER	29.40	0.34	12.32	36.41	17.44	5.39	2.31	8.86	-	1.00	19.65	29.20	**16.80**
BPCER	3.99	36.89	17.52	10.12	12.53	24.79	19.94	4.13	-	12.68	5.73	1.39	**6.75**
ACER	16.70	18.62	14.92	23.27	14.98	15.09	11.13	6.50	3.54	6.84	12.69	15.29	**11.78**

forms well. Compared to previous methods, IFCNet achieves a comparable performance with the state-of-the-art baseline [3], [7], [8], [10], [18], [20]. [10], [18] do not evaluate the IPAD framework on the NDCLD'13 dataset. For a fair comparison of our DFCANet with the state-of-the-art, we reproduced the results of [10], [18] on NDCLD'13 dataset. One can observe from Table XI, that both our proposed models - IFCNet and DFCANet outperform [10], [18] by significant margins. Though [10], [18] try to capture local-global information, they do not distill domain-specific spatial contexts into channel-dimensions (which DFCANet does). Further, since all: [10], [18] and DFCANet follow the same initial DenseNet-based pipeline, the superior performance of DFCANet over these two approaches validates the importance of IFCNet (the feature calibration sub-network). Next, we have illustrated misclassifications by DFCANet in Fig. 8 (a). It is evident from the same that most of the misclassifications in both attack and bonafide are mainly due to SLA. Uneven lighting conditions and morphological artifacts also influence the performance. Thus, in order to robustly model SLA scenarios, more abrupt changes between the boundary of lens and iris must be captured.

4) Comparison on IIITD-WVU Dataset In addition to performing various cross-domain experiments, in this section we evaluate the performance of DFCANet and its sub-modules (IFCNet and DenseNet) over the challenging IIITD-WVU dataset (a subset of LivDet-2017 Challenge). Specifically, this is an ideal dataset for cross-conditions settings as it inculcates cross-subject, cross-sensor and even varying environments for training and testing. The results obtained in the experiment have been illustrated in Table XII. It is evident from the same that DFCANet attains significant performance of 11.78% in terms of ACER. Furthermore, DFCANet outperforms LivDet-2017 benchmark [5] along with D-NetPAD [10], and other well known studies such as [11], [20], [53], [54]. DFCANet also attains highly comparable performance with respect to [3], which relies on highly pre-processed representations which are difficult to realize in unconstrained settings. DFCANet's performance is comparatively inferior to [8], [9], [24], though [8] uses binary supervision masks (restricting its generalizability). [24] uses unsupervised target domain knowledge, while DFCANet follows a simpler end-to-end training and evaluation pipeline. Nevertheless, in contrast to these baselines [8], [9], [24], DFCANet's architecture has been designed over relatively simpler datasets. It generalizes significantly well over challenging cross-domain settings. When DFCANet is compared with its sub-modules, it achieves 3.51% and 0.91% better ACER score than IFCNet and DenseNet respectively. Also, both IFCNet and DenseNet at their end outperform previous baselines [5], [10], [11], [53].This highlights that with capturing local global contexts in iris via IFCNet, DenseNet representations are enriched, and hence DFCANet leads to better performance. To gain further insights upon contributions of each of these modules of DFCANet, corresponding ROC plots have been plotted in Appendix 4 (Supplementary Material). It is evident from the same figure that DFCANet attains lower error rates when compared against both DenseNet and IFCNet. We have also analyzed failure cases in Fig. 8 (b). It is evident from the figure that iris images with rare appearance causes misclassification of bonafide images. While in the attack class, lenses that have lesser intricate patterns are misclassified. Unconstrained capture of iris and other morphological artifacts also poses challenges on the model. These cases motivate requirement of evaluation upon 'in-the-wild' settings.

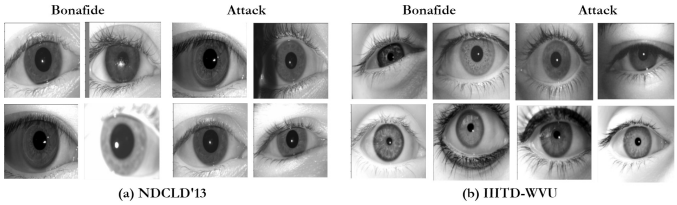


Fig. 8. Misclassified examples from the NDCLD'13 and IIITD-WVU datasets. For NDCLD'13, examples are from combined sensor SLA experiments. Here, attack samples have high visual similarity with bonafide ones. Cross-domain evaluation and imaging artifacts further complicate matters.

5) Comparison on Clarkson-17 Dataset Evaluation on this dataset is characterized by unseen patterned lens attacks. The results obtained in the experiment for the dataset have been summarized in Table XIII. DFCANet attains a relatively

TABLE XIII
PERFORMANCE COMPARISON FOR CLARKSON-17 DATASET (IN %)

Metrics	DNetPad [10]	Spoofnet [53]	PBS [8]	APBS [8]	FAM [24]	Yolo-CNN [51]	EyePAD++ [55]	DenseNet	IFCNet	DFCANet
APCER	5.78	33.00	8.97	6.16	6.10	2.43	7.29	5.25	13.40	**1.81**
BPCER	0.94	20.00	0.00	0.81	0.81	0.00	0.79	3.34	5.55	**5.08**
ACER	3.36	16.50	4.48	3.48	3.45	1.61	3.65	4.29	9.47	**3.44**

lower or comparative ACER in comparison to almost all the well known state-of-the-art frameworks [8], [10], [24], [53]. Besides, DFCANet attains the least APCER score [8], [10], [19], [24], [53], [55]. This indicates our model misclassifies the least number of spoof samples. Both DFCANet and [55] do not involve any preprocessing, but DFCANet attains marginally better performance than [55]. Concretely, [55] achieves high-performance on the basis feature sharing between authentication and IPAD tasks. However, it attains awareness to particular identities (on which is trained on) and hence is not ideal for cross-subject evaluation. DFCANet on the other hand embanks on a domain-specific IPAD design and does not face these challenges. Further, DFCANet also outperforms both IFCNet and DenseNet by respective margins of 6.03% and 0.85% in terms of ACER. Similar results in terms of error rates is also evident from the DET-curve plotted in Appendix 5 (Supplementary Material).

6) Comparison on Clarkson-15 Dataset In this experiment, the presentation attack images acquired by both the sensors are considered together. The obtained results are summarised in Table XIV. It is observed that DFCANet outperforms Yolo-CNN [51] on the Clarkson-15 dataset in AA. When comparison is done for the variants of DFCNet, there are two-fold observations: i) DFCANet attains higher AA than both of its variants. ii) IFCNet on the other hand, outperforms both DFCANet and DenseNet in terms of ACER. These two observations emphasize the importance of local-global feature calibration.

TABLE XIV
PERFORMANCE COMPARISON FOR CLARKSON-15 DATASET (IN %)

Metrics	Yolo-CNN [51]	DenseNet	IFCNet	DFCANet
APCER	2.92	3.17	0.7	**3.04**
BPCER	1.12	5.35	2.2	**1.36**
ACER	2.02	4.26	1.53	**2.20**
AA	97.74	95.09	98.02	**98.29**

7) Comparison on NDCLD-15 Dataset Comparative analysis of the proposed DFCANet with its model counterparts and state-of-the-art [6], [7], [15], [21], [41] has been evaluated on the NDCLD-15 dataset. Note that the evaluation protocol for NDCLD-15 data has not been uniformly adopted in literature. OSPAD [21] considered easier setting of 4,068 images from LG4000 sensor only. While, A-PBS [8] adopted 5-fold cross validation over 7,300 images, and BSIF [42] mentions about 4-Fold Cross-Validation. Thus, it follows that comparison (results tabulated in Table XV) against other state-of-the-art approaches [7], [21] is inadequate. Nevertheless, DFCANet demonstrated a relatively close result with [21], it also significantly outperformed [6], [15], [41] in terms ACER.

TABLE XV
PERFORMANCE COMPARISON (4-FOLD) FOR NDCLD-15 DATASET (IN %)

Metrics	OSPA [21]	LBP [6]	wLBP [15]	DESIST [41]	MHVF [7]	DenseNet	IFCNet	DFCANet
APCER	7.14	6.15	50.58	29.81	1.92	10.29	9.15	7.16
BPCER	8.57	38.70	4.41	9.22	0.39	21.21	19.34	21.35
ACER	7.85	22.43	27.50	19.52	1.15	13.01	12.89	13.13

H. Contact Lens Detection: A Case Study

To further verify generalizability of our model, we took into account the contact lens detection task and compared the proposed model with existing works in the literature [56], [57]. For this part of the study, we have considered NDCLD'13 dataset for training and testing. Data from LG4000 and AD100 were combined and split into two equal halves for the three-class classification. In Table XVI, we have tabulated the results of the comparison for contact lens detection under combined sensor experiments. Since all previous attempts to detect contact lenses have been using class-based matching, [19] involved validation using three class classifications. From Table 12, it is evident that DFCANet outperforms previous methods by 3-8% in performance. It was noticed that the major amount of misclassification in this experiment were between soft contact lenses and bonafide. This further validates the challenge in SLA settings. Nevertheless, the significant performance of DFCANet is a measure of its robustness.

TABLE XVI
COMBINED-SENSOR CONTACT LENS DETECTION: NDCLD'13

Method	ContlensNet [56]	GHCLNet [57]	DCCNet [19]	DFCANet
Avg. Acc.	85.45	87.01	90.04	**93.00**

V. PARAMETER AND DETECTION RUN-TIME ANALYSIS

As a matter of practical interest, it is essential for the IPAD-algorithm to be fast during inference time. Hence, in this section, we analyze the detection run-time and number of parameters for DFCANet and its variants. In Table XVII, detection run-time have been reported in milliseconds for both GPU (Tesla P100 – 16GB) and CPU (Haswell 2.30 GHz, 4 cores, 16 GB). DFCANet requires 40.9 ms and 270.4 ms on GPU and CPU respectively to execute a detection. These time periods are significantly low and are comparable to detection time required by other low-parametric variants such as DenseNet and DenseNet+CAM. In terms of parameter, DFCANet though utilizes the large number of parameters, but trades-off high performance.

TABLE XVII
PARAMETER AND RUN-TIME ANALYSIS OF DFCANET

Model	Network Parameters	Run-time (ms)	
		GPU	CPU
DenseNet	480833	31.6	77.3
DenseNet + IFCNet	14824257	40.6	264.7
DenseNet + CAM	480833	33.4	81.9
CAM	0	-	-
IFCNet	14446081	95.5	3700.0
DFCANet	14824357	40.9	270.4

VI. CONCLUSIONS

Existing IPAD algorithms are burdened with performance penalties in cross-domain scenarios which do not make them ideal for newer formats of iris-based presentation attacks. To this end, this paper proposed IPAD domain specific DFCANet, which emphasizes on learning joint correlations between coherent patterns found at local and global scales. Experiments conducted on the IIITD-CSD, IIITD-CLI, NDCLD-13, NDCLD-15, Clarkson-15, Clarkson-17, and IIITD-WVU datasets show that the DFCANet is not only effective in intra-domain settings but generalizes well enough for various cross-domain scenarios while operating under non-ROI extracted images. We introduce a challenging Soft-Lens-as-Attack (SLA) experiment as a benchmark. To overcome such stark challenges and enhance learning in a low-data regime, an incremental learning based methodology was introduced. However, a subtle research gap remains open. The SLA protocol introduced requires more robust modelling of iris features, and for that in a future work spatio-periodic and local-global features shall be explored. Another limitation of this work is that it does not seek into unseen attack paradigm and 'in-the-wild' settings. These forms of evaluation will be considered along with the extension of the analysis to more challenging attacks and datasets.

REFERENCES

[1] A. Czajka and K. W. Bowyer, "Presentation Attack Detection for Iris Recognition: An Assessment of the State-of-the-Art," *ACM Computing Surveys*, vol. 51, no. 4, pp. 1–35, 2018.

[2] S. Marcel *et al.*, *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection*. Springer, 2019.

[3] M. Fang *et al.*, "Cross-Database and Cross-Attack Iris Presentation Attack Detection using Micro Stripes Analyses," *Image and Vision Computing*, vol. 105, p. 104057, 2021.

[4] P. Das *et al.*, "Iris Liveness Detection Competition (Livdet-Iris)-the 2020 Edition," in *Proc. IEEE International Joint Conference on Biometrics (IJCB)*, 2020, pp. 1–9.

[5] D. Yambay *et al.*, "Livdet Iris 2017—Iris Liveness Detection Competition 2017," in *Proc. IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 733–741.

[6] P. Gupta, S. Behera, M. Vatsa, and R. Singh, "On Iris Spoofing using Print Attack," in *Proc. International Conference on Pattern Recognition (ICPR)*. IEEE, 2014, pp. 1681–1686.

[7] D. Yadav *et al.*, "Fusion of Handcrafted and Deep Learning Features for Large-Scale Multiple Iris Presentation Attack Detection," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR-W)*, 2018, pp. 572–579.

[8] M. Fang *et al.*, "Iris Presentation Attack Detection by Attention-based and Deep Pixel-Wise Binary Supervision Network," in *Proc. IEEE International Joint Conference on Biometrics (IJCB)*, 2021, pp. 1–8.

[9] A. Agarwal, A. Noore, M. Vatsa, and R. Singh, "Generalized Contact Lens Iris Presentation Attack Detection," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 3, pp. 373–385, 2022.

[10] R. Sharma and A. Ross, "D-NetPAD: An Explainable and Interpretable Iris Presentation Attack Detector," in *Proc. IEEE International Joint Conference on Biometrics (IJCB)*, 2020, pp. 1–10.

[11] A. Kuehlkamp *et al.*, "Ensemble of Multi-View Learning Classifiers for Cross-Domain Iris Presentation Attack Detection," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 6, pp. 1419–1431, 2018.

[12] D. Yambay *et al.*, "Review of Iris Presentation Attack Detection Competitions," in *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*. Springer, 2023, pp. 149–169.

[13] Z. Sun, H. Zhang, T. Tan, and J. Wang, "Iris Image Classification based on Hierarchical Visual Codebook," *IEEE Transactions on Pattern Analysis and Machine Intellingence*, vol. 36, no. 6, pp. 1120–1133, 2013.

[14] Z. He *et al.*, "Efficient Iris Spoof Detection via Boosted Local Binary Patterns," in *Proc. International Conference on Biometrics (ICB)*, 2009, pp. 1080–1090.

[15] H. Zhang, Z. Sun, and T. Tan, "Contact Lens Detection based on Weighted LBP," in *Proc. International Conference on Pattern Recognition (ICPR)*, 2010, pp. 4279–4282.

[16] Y. Hu, K. Sirlantzis, and G. Howells, "Iris Liveness Detection using Regional Features," *Pattern Recognition Letters*, vol. 82, pp. 242–250, 2016.

[17] D. Menotti *et al.*, "Deep Representations for Iris, Face, and Fingerprint Spoofing Detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 864–879, 2015.

[18] C. Chen and A. Ross, "An Explainable Attention-Guided Iris Presentation Attack Detector," in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 97–106.

[19] M. Choudhary, V. Tiwari, and U. Venkanna, "Iris Anti-Spoofing through Score-Level Fusion of Handcrafted and Data-Driven Features," *Applied Soft Computing*, vol. 91, p. 106206, 2020.

[20] M. Fang *et al.*, "Deep Learning Multi-Layer Fusion for an Accurate Iris Presentation Attack Detection," in *Proc. International Conference on Information Fusion (FUSION)*, 2020, pp. 1–8.

[21] Z. Fang, A. Czajka, and K. W. Bowyer, "Robust Iris Presentation Attack Detection Fusing 2D and 3D Information," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 510–520, 2020.

[22] D. Yadav *et al.*, "Detecting Textured Contact Lens in Uncontrolled Environment using DensePAD," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR-W)*, 2019.

[23] M. Gupta *et al.*, "Generalized Iris Presentation Attack Detection Algorithm under Cross-Database Settings," in *Proc. International Conference on Pattern Recognition (ICPR)*, 2021, pp. 5318–5325.

[24] Y. Li *et al.*, "Few-Shot One-Class Domain Adaptation based on Frequency for Iris Presentation Attack Detection," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 2480–2484.

[25] N. Kohli *et al.*, "Synthetic Iris Presentation Attack using iDCGAN," in *Proc. IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 674–680.

[26] S. Yadav and A. Ross, "CIT-GAN: Cyclic Image Translation Generative Adversarial Network with Application in Iris Presentation Attack Detection," in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 2412–2421.

[27] A. Morales *et al.*, "Introduction to Iris Presentation Attack Detection," in *Handbook of Biometric Anti-Spoofing*. Springer, 2019, pp. 135–150.

[28] A. Boyd *et al.*, "Comprehensive Study in Open-Set Iris Presentation Attack Detection," *IEEE Transactions on Information Forensics and Security*, 2023.

[29] D. Yambay *et al.*, "Review of Iris Presentation Attack Detection Competitions," in *Handbook of biometric anti-spoofing*. Springer, 2019, pp. 169–183.

[30] A. Agarwal *et al.*, "Enhanced Iris Presentation Attack Detection via Contraction-Expansion CNN," *Pattern Recognition Letters*, 2022.

[31] Y. Li *et al.*, "Single Domain Dynamic Generalization for Iris Presentation Attack Detection," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[32] A. Boyd, K. W. Bowyer, and A. Czajka, "Human-Aided Saliency Maps Improve Generalization of Deep Learning," in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022, pp. 2735–2744.

[33] M. Fang *et al.*, "Demographic Bias in Presentation Attack Detection of Iris Recognition Systems," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2021, pp. 835–839.

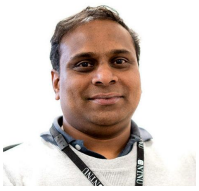
[34] A. Agarwal *et al.*, "Misclassifications of Contact Lens Iris PAD Algorithms: Is It Gender Bias or Environmental Conditions?" in *Proc.*

IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023, pp. 961–970.
[35] K. He *et al.*, "Deep Residual Learning for Image Recognition," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[36] C. Szegedy *et al.*, "Going Deeper with Convolutions," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.
[37] G. Huang *et al.*, "Densely Connected Convolutional Networks," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.
[38] J.-J. Liu *et al.*, "Improving Convolutional Networks with Self-Calibrated Convolutions," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 10 096–10 105.
[39] D. Yadav *et al.*, "Unraveling the Effect of Textured Contact Lenses on Iris Recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 851–862, 2014.
[40] J. S. Doyle, K. W. Bowyer, and P. J. Flynn, "Variation in Accuracy of Textured Contact Lens Detection based on Sensor and Lens Pattern," in *Proc. International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1–7.
[41] N. Kohli *et al.*, "Detecting Medley of Iris Spoofing Attacks using DESIST," in *International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2016, pp. 1–6.
[42] J. S. Doyle and K. W. Bowyer, "Robust Detection of Textured Contact Lenses in Iris Recognition using BSIF," *IEEE Access*, vol. 3, pp. 1672–1683, 2015.
[43] "International Organization for Standardization. iso/iec dis 30107-3:2016: Information Technology – Biometric Presentation Attack Detection," https://www.iso.org/standard/67381.html/.
[44] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7132–7141.
[45] S. Woo *et al.*, "CBAM: Convolutional Block Attention Module," in *Proc. European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.
[46] D. Linsley *et al.*, "Learning What and Where to Attend," *arXiv preprint arXiv:1805.08819*, 2018.
[47] A. Dosovitskiy *et al.*, "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," *arXiv preprint arXiv:2010.11929*, 2020.
[48] J. Hu *et al.*, "Gather-Excite: Exploiting Feature Context in Convolutional Neural Networks," *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 31, 2018.
[49] G. Gautam, A. Raj, and S. Mukhopadhyay, "Deep Supervised Class Encoding for Iris Presentation Attack Detection," *Digital Signal Processing*, vol. 121, p. 103329, 2022.
[50] R. R. Selvaraju *et al.*, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 618–626.
[51] M. Choudhar, V. Tiwari, and V. Uduthalapally, "Iris Presentation Attack Detection based on Best-k Feature Selection from YOLO Inspired RoI," *Neural Computing and Applications*, vol. 33, no. 11, pp. 5609–5629, 2021.
[52] J. McGrath, K. W. Bowyer, and A. Czajka, "Open Source Presentation Attack Detection Baseline for Iris Recognition," *arXiv preprint arXiv:1809.10172*, 2018.
[53] G. Y. Kimura *et al.*, "CNN Hyperparameter Tuning Applied to Iris Liveness Detection," *arXiv preprint arXiv:2003.00833*, 2020.
[54] V. Jain *et al.*, "Robust Iris Presentation Attack Detection through Stochastic Filter Noise," in *Proc. International Conference on Pattern Recognition (ICPR)*, 2022, pp. 1134–1140.
[55] P. Dhar *et al.*, "EyePAD++: A Distillation-based Approach for Joint Eye Authentication and Presentation Attack Detection using Periocular Images," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 20 218–20 227.
[56] R. Raghavendra, K. B. Raja, and C. Busch, "ContlensNet: Robust Iris Contact Lens Detection using Deep Convolutional Neural Networks," in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2017, pp. 1160–1167.
[57] A. Singh *et al.*, "GHCLNet: A Generalized Hierarchically Tuned Contact Lens Detection Network," in *Proc. IEEE International Conference on Identity, Security, and Behavior Analysis (ISBA)*, 2018, pp. 1–8.

Gaurav Jaswal completed his M.Tech and Ph.D. (Electrical Engineering) from NIT Hamirpur. He is currently an i-Hub Faculty Fellow at IIT Mandi, having previously been a Post-doctoral Fellow at IIT Delhi. He is interested in brain-computer interfacing, multimodal biometrics, medical imaging, action/gesture analysis and deep learning.

Aman Verma obtained his B.Tech (Electronics and Telecommunications) from NIT Raipur in 2022. He has been a research intern at MNNIT Allahabad, Manas Lab IIT Mandi, and IIIT-Hyderabad. He is currently pursuing MS-Research in Electrical Engineering at IIT Delhi. His research interests include Deep Learning, Pattern Recognition, Computer Vision, Biometrics and Medical Signal and Image Analysis. He is also a recipient of the AI-Safety Careers Fellowship, IIT Delhi.

Sumantra Dutta Roy completed his M.Tech and Ph.D. (Computer Science and Engineering) from IIT Delhi in 1994 and 2001, respectively. He has been in the Department of Electrical Engineering at I.I.T. Delhi since 2007 (where he is currently a Professor), having previously been in the Department of Electrical Engineering at IIT Bombay (2001-2007). He has also held a visiting position at Hiroshima University. He is a recipient of the 2004 INAE Young Engineer Award (Indian National Academy of Engineering) and the 2004-5 BOYSCAST Fellowship of the Department of Science and Technology, Government of India. He has been an Associate Editor of the Pattern Recognition Letters since 2011. His research interests are in Computer Vision, Machine Learning, Medical Informatics, Biometrics, Video Coding, and Music Information Retrieval.



Raghavendra Ramachandra obtained a Ph.D. in computer science and technology from the University of Mysore, Mysore India and Institute Telecom, and Telecom Sudparis, Evry, France (carried out as collaborative work) in 2010. He is currently a full professor at the Institute of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology (NTNU), Gjøvik, Norway. He is also working as R&D chief at MOBAI AS. He was a researcher with the Istituto Italiano di Tecnologia, Genoa, Italy, where he worked with video surveillance and social signal processing. His main research interests include deep learning, machine learning, data fusion schemes, and image/video processing, with applications to biometrics, multi-modal biometric fusion, human behaviour analysis, and crowd behaviour analysis. He also holds several patents in biometric presentation attack detection and morphing attack detection. He has also been involved in various conference organizing and program committees and has served as an associate editor for various journals. He has participated (as a PI, co-PI or contributor) in several EU projects, IARPA USA and other national projects. He is serving as an editor of the ISO/IEC 24722 standards on multimodal biometrics and an active contributor to the ISO/IEC SC 37 standards on biometrics. He has received several best paper awards, and is also a senior member of IEEE and VP of Finance (IEEE Biometric Council).