Multimedia Tools and Applications (2023) 82:45959–45981
https://doi.org/10.1007/s11042-023-15421-x

IVIDNet: Intelligent iris vitality detection via weighted prediction score level fusion

Palak Verma[1] · **Arvind Selwal**[2] · **Deepika Sharma**[2]

Received: 21 September 2022 / Revised: 24 February 2023 / Accepted: 18 April 2023 /
Published online: 28 April 2023

Abstract
Iris recognition is one of the most widely used human authentication mechanism that has gained huge popularity due to its security and efficiency. However, recently attackers have breached the security of these systems via synthetically generated fake iris traits using a variety of artefacts such as paper printouts, video attacks, and cosmetic lenses. To countermeasure these attacks, an anti-deception sub-system is instilled between sensor and feature extractor modules of the system. These sub-systems intelligently measure the vitality of the presented iris traits and are popularly known as iris vitality detection (IVID) or iris anti-spoofing techniques. Recently, the deep learning-based models are widely deployed for iris anti-spoofing approaches, but these methods result in additional training overhead as well as require larger training dataset. In this paper, we expound a novel iris vitality detection network (IVIDNet) that employs the robust features of two pre-trained deep convolutional neural networks (DCNN). The extracted features are then integrated via a new weighted score-level fusion technique. To demonstrates the efficacy of the proposed technique, experimental analysis is carried out on benchmark iris liveness detection Notre Dame 2017 and Notre Dame cosmetic lenses dataset (NDCLD) 2015. The proposed IVIDNet exhibits excellent performance in the known attack scenario with an average classification error rate (ACER) of 0.60%. Moreover, IVIDNet shows superior performance in the unseen attacks scenario (i.e. cross-database) with an ACER of 9.72%. In overall, the proposed model outperforms other similar iris anti-spoofing approaches and efficient in terms of computational overhead.

Keywords Iris recognition · Spoof attacks · Anti-spoofing techniques · Deep learning · Transfer learning · Score-level fusion

✉ Palak Verma
2022rcs0006@iitjammu.ac.in

[1] Department of Computer Science and Engineering, Indian Institute of Technology Jammu, PO-Nagrota, Jammu, India 181221

[2] Department of Computer Science and Information Technology, Central University of Jammu, Samba, India 181143

1 Introduction

The biometric-based recognition systems are being tremendously utilized for many critical applications such as law enforcement, cell-phone authentication, citizen identification, healthcare, border control, commercial applications and public security. Dominantly, the iris recognition systems are being deployed for this task due to its higher reliability with user's convenience. A recent report [9] indicates that the iris market size is expected to grow from USD 2.3 billion in 2019 to USD 4.3 billion by 2024, at a CAGR of 13.2% during the forecast period. The widespread use of iris recognition technology for identification and verification by government entities is the key driving factor for market growth. Other substantial aspects that are significantly benefiting the iris recognition industry's growth include the rising penetration of iris recognition technology in the consumer electronics vertical and the high demand for iris scanners for access control applications. Even though these systems offer better security to a variety of computing applications, these systems endure a variety of attacks threating the security. Among the eight vulnerable attack points identified by Ratha et al. [11], the presentation or spoof attack is regarded as the most widely and easier attempted at the sensor level. The sensor module of an iris biometric system is mainly compromised by presenting a forged modality using a variety of artefacts such as iris's paper printouts, textured contact lenses, prosthetic eyes, and etc. To alleviate presentation attacks, an IVIDNet [14] also known as a liveness detection module [15] is integrated with the biometric recognition system which serves as a security check. In general, the IVIDNet technique [23] may be considered of as a binary classification problem addressed by computing the difference in micro-textural or image quality features between live and counterfeit iris attributes. Contemporarily, the development of vitality detection network (VIDNet) mechanisms is a prominent field of research, with a slew of contributions aimed at providing efficient liveness detection systems. While VIDNet approaches based on single or multiple image characteristics have yielded accurate detection systems, their performance and generalisation capability to unknown attacks has been restricted. Additionally, choosing the appropriate number and type of image attributes for identifying a given image as genuine or fake is also one of the hefty tasks in conventional VIDNet algorithms. The field of deep learning (DL)-based IVIDNet is emerging as a potential alternate compared to traditional methods in current time. There are several reasons for this, including autonomous deep feature extraction and improved accuracy. However, these strategies have pitfalls, like increased training overhead and the need for a larger training dataset. To address these issues, an emerging paradigm in deep learning is to utilize the knowledge of pre-trained models in a specific domain that may be efficiently translated to build an efficient VIDNet. Transfer learning offers numerous advantages, including reduced training time, better performance (in most cases) and eliminating the need of large training dataset. Although, many researchers have utilized the capabilities of pre-trained models such as AlexNet, VGGNet, and ResNet50 there are still open research issues that are need to be tackled using the most efficient models and utilizing techniques.

Therefore, this work broadly aims to develop an efficient vitality detection technique that is not impacted by the type of fabrication material used to spoof the model. To attain this, we train and test our model on different datasets (cross-datasets). The key contributions of this work may be summarised as follows:

i. We propose a novel technique for iris anti-spoofing using weighted score level fusion of potent deep level features of iris images.

ii. The proposed approach overcome the challenge of scarcity of large amount of iris samples via transfer learning by including various iris artefact.
iii. The IVIDNet approach integrate the merits of two robust and pre-trained deep models to yield a generalized iris spoof detector.
iv. The proposed approach is evaluated on the benchmark iris liveness detection Notre Dame 2017 and NDCLD 2015.
v. The IVIDNet demonstrates superior performance in both known and unknown attack scenarios as well as outperform the similar anti-spoofing mechanisms.

The remainder of the study is organized as follows: Section 2 presents a review on recent advancements in IVIDNet techniques. The framework and algorithms of the proposed IVIDNet approach is illustrated in Section 3. The experimental benchmark datasets along with performance protocol and a detailed experimental analysis are systematically discussed in Section 4. At last, the conclusions as well as the future scope of this work are briefed in Section 5.

2 Related work

Due to the significance of counter mechanisms and current advances in iris-based recognition systems, it is necessary to detect the liveness of the presented characteristic, since intruders may readily impersonate the authentication system using various presenting instruments (PAIs). As a result, the most pressing research problem is to distinguish a live iris from a fake one, as mandated by ISO-Standard IEC 30,107–3 E. Several trends in iris anti-spoofing systems have emerged throughout the years, based on a variety of essential principles. Vitality detection techniques are often classified into two categories based on the type of liveness indicator used, namely (a) hardware-based analysis and (b) data driven-based analysis. For identifying real and fake iris qualities, the former approach uses an extra sensing device in addition to the iris recognition system to assess vitality characteristics including temperature, impedance, image quality, blood cells, and so on. Unlike the former, the data driven-based module analyses the image features of single (static) or multiple (dynamic) iris images using hand-crafted or automated feature extraction-based approaches to detect liveness attributes.

Transfer learning-based techniques are receiving more attention these days as a consequence of their impressive successes in spoof attack detection mechanisms. Our suggested method is based on using the pre-trained models and generating a novel and efficient IVIDNet; hence, our study falls into the automated feature extraction-based iris spoof detector category. As a result, this section's brief literature assessment is confined to pioneer contributions linked to transfer learning-based methods. Beforehand, Ribeiro et al. [12] investigated texture transfer learning for super resolution that is applied to low resolution images. Although, on a subset of the CASIA iris image dataset, the developed technique achieves the best EER of 6.07% in factor 2 when the describable texture dataset (DTD) is used. This work lacks to investigate the integration of best datasets with the enrolling outcomes. Chen and Ross [2] suggested a multi-task iris vitality detector (IVD) system based on a technique for detecting objects. This method is computationally efficient and can be implemented in a real-time setting. However, in instances when the training and test datasets have distinct assaults, the approach is not studied. Gautam and Mukhopadhyay [6] introduced a transfer learning approach that depended on

a pre-trained AlexNet model for feature extraction and dimensionality reduction, followed by principal component analysis (PCA). For classification, a Cubic SVM (cSVM) multi-class model based on error-correcting output code (ECOC) is utilised. This study needs to address efficient comprehension and exploitation of hybrid classifiers, as well as strong feature extraction algorithms in tandem with deep image representation.

Alaslani and Elrefaei [1] proposed an efficient iris authentication system based on transfer learning with CNN. For feature extraction and classification, this method is accomplished by fine-tuning a pre-trained VGG-16 model. The performance of the iris recognition system is assessed using four publicly available databases: IIITD, CASIA-Iris-V1, CASIA-Iris-thousand, and CASIA-Iris-Interval. According to the findings, the proposed technique has a 100% accuracy rate in the instance of IIITD. Minaee and Abdolrashidi [10] provided a deep learning system based on a pre-trained CNN model (ImageNet). The performance of the mechanism is measured on the IITD dataset, and an accuracy rate of 95.5% is measured. Choudhary et al. [3] presented a novel densely connected contact lens detection network (DCLNet) based on DCNN with SVM on top for classification. Other networks have more layers and learning parameters than the DCLNet, which is a densely linked convolutional network with fewer layers and learning parameters. Because of the tight connections between layers, it learns more critical qualities. The experimental findings show that when compared to the state-of-the-art (SOA), the suggested technique improves the CCR by up to 4%. Normalization, on the other hand, may be inferred to degrade the model's accuracy in the vast majority of circumstances. Therar et al. [21] used the multimodal biometric real-time approach IrisConvNet based on the architecture of a deep learning model for instances of a person's left and right irises. The CNN and transfer learning techniques are deployed to produce specific features that are fed into a multi-class SVM algorithm for feature extraction and classification. IrisConvNet performance is evaluated using two publicly available datasets: IITD and CASIA-Iris-V3. IITD has a 99% accuracy rate for both the left and right iris, whereas CASIA-Iris-V3 has a 94% and 93% accuracy rate for the left and right iris, respectively. Sardar et al. [13] proposed a deep Interactive Squeeze Expand Unet (ISqEUNet) model with interactive learning to reduce training time while enhancing storage efficiency by minimising the number of involved parameters. NICE.I has a mean true positive rate (mTRP) of 0.983% and a mean error rate (MER) of 0.261%, according to the results of three publicly accessible datasets.

Another IVD solution based on multi-layer fusion is propounded by Fang M. et al. [5]. Two level fusion i.e. feature level and score level is done on the feature extracted from the last several convolution layers. Although, result shows that multi-layer fusion technique performs better as compare to the best single layer feature extractor using pre-trained VGG-16, but while trained from scratch this technique perform well only on larger dataset such as the IIITD-WVU database in comparison to the Notre Dame database. Recently, Tapia J. et al. [20] deliberated a two stage serial framework for PAD focused on detecting bonafide images. For this approach the largest iris PA database by combining several other databases is developed and model is tested when trained from scratch and using fine-tuning. Although comparable results were obtained in known environment the performance of proposed two stage networks is not measured in unknown attack scenarios.

Based on the comparative analysis of several TL-based IVD as specified in Table 1, it can be inferred that in most of the techniques a pre-trained model on ImageNet is deployed. The reason behind this is, it consists of over 14 million images of roughly 20,000 categories and training a new model using this may reduce the overall training time. Moreover, IITD iris anti-spoofing dataset is widely used anti-spoofing dataset in these approaches.

Table 1 A summary of various transfer learning enabled IVD mechanisms

Paper id	Year	Author(s)	Pre-trained model	Classifier	Dataset	Performance	
						Known	Unknown
t1	2017	Ribeiro et al. [12]	Single-Image Super Resolution (SRCNN) and deeper CNN VDSR	CNN based on transfer learning fir image classification	Texture databases, natural image databases, iris databases	In case of EER the best result for: factor 2 (115X115) is when DTD is used = 6.07% and for factor 4 (57X57) is when bicubic interpolation is used	_
t2	2018	Chen and Ross [2]	Darknet-19 model trained on ImageNet	MT-PAD with no fully connected layer	ND-Contact-13, CASIS iris, LivDet iris-15 Warsaw, LivDet iris-17 clarkson, BERC-Iris Fake	BPCER: ND-Contact, CASIS iris = 0.5%, LivDet iris-15 Warsaw = 0%	BPCER for Cross sensor: BERC = 1.66%, LivDet iris-17 clarkson = 20.44%
t3	2018	Gautam and Mukhopadhyay [6]	Transfer learning relying on (AlexNet) pre-trained DCNN	cSVM learner-based ECOC multiclass model	IIITD, ND	Multi sensor evaluation: Total CCR in case of: IIITD combined = 81.40%, ND combined = 86.33%	_
t4	2019	Alaslani and Elrefaei [1]	The Softmax classifier in pre-trained VGG -16	Transfer learning with pre-trained CNN and then static formula or SVM for classification	IIITD, CASIA-Iris -thousand, CASIA-IrisV1 and CASIA-Iris-Interval	Recognition accuracy: IIITD iris = 100%, CASIA-Iris thousand = 95%, CASIA-IrisV1 = 98.3%, and CASIA-Iris-Interval = 91.6%	_

Table 1 (continued)

Paper id	Year	Author(s)	Pre-trained model	Classifier	Dataset	Performance	
						Known	Unknown
t5	2019	Minaee and Abdolrashidi [10]	ResNet50 trained on ImageNet	Transfer learning model using deep residual CNN	IIT Delhi	Accuracy rate = 95.5%	_
t6	2019	Choudhary et al. [3]	DenseNet121	DCLNet based on DCNN with an additional SVM classifier	IIITD-CLI, Notre Dame (ND)-13	CCR for multi sensor validation: ND combined = 96.04%, IIITD combined = 94.93%,	_
t7	2020	Therar et al. [21]	CNN with transfer learning	IrisConvNet with softmax classifier and multi-class SVM classifier	IITD and CASIA-IrisV3	Accuracy rate: IITD = 99% (left and right iris both), CASIA-IrisV3 = 94% and 93% (left and right iris respectively)	_
t8	2020	Sardar et al. [13]	Pre-trained model onto the dataset CASIA-Iris-V4-Interval	Deep ISqEUNet model with interactive learning	CASIA-Iris-V4-Interval, IITD, NICE.I	NICE.I: mTRP = 0.983% MER = 0.261%	_
t9	2020	Fang M. et al. [5]	Pre-trained VGG-16 and trained from scratch MobileNet V3	Multi-layer fusion (on feature level and score level)	LivDet-iris 2017 (Notre Dame and IIITD-WVU datasets)	Notre Dame (HTER): VGG-16 = 2.31% MobileNet = 8.89% IIITD-WVU (HTER): VGG-16 = 20.88% MobileNet = 15.09%	_
t10	2022	Tapia J. et al. [20]	MobileNet V2 and a model trained from scratch	Two stage classification using MobileNetv2a and MobileNetv2b	LivDet-Iris 2020	ACER = 29.78%	_

Besides, the accuracy rate for IITD dataset in transfer learning-based approaches ranges from 81.40% to 100%.

3 The proposed approach

The deep learning-based approaches are advantages for capturing similarities among adjacent pixel values to safeguard against spoof attacks. To achieve better vitality detection results, it is imperative to train an appropriate IVID model that is based on significant features extracted from an adequate number of relevant images. In order to address specific concerns of existing SOA approaches, such as increased training overhead and the need of larger dataset, we provide a transfer learning-based IVID technique that significantly improves overall performance. The approach is based on weighted fusion of the predictions of two pre-trained model namely InceptionNet [18] and VGG-19 [17]. In the following subsections, we describe the underlying idea of IVIDNet framework, algorithms, and weighted score level fusion of the outcomes of various models.

3.1 The IVIDNet framework

Extracting deep level features to design a robust IVD that performs well in unknown assaulting scenarios is one of the critical issues in CNN-based methods. To address these issues, a recent deep learning paradigm is to use the knowledge of pre-trained models in a specific domain that may be efficiently translated to construct an efficient IVD. To this end, we present an efficient framework for IVIDNet to accomplish the task of the anti-spoofing sub-module as illustrated in Fig. 1. The suggested framework's main premise is to work in two stages, which include training and testing procedures.

The training stage broadly comprises a series of activities that are applied to iris dataset such as pre-processing, deep feature extraction, building a model by fusing the predictions of two pre-trained models, and parameter of the target models. The goal of testing stage is to evaluate the IVIDNet model's correctness by validating it on a randomly selected set of images encompassing a variety of sensors and datasets. The detailed explanation of these phases is discussed in the following subsections.

3.1.1 Pre-processing

The key objective of image pre-processing is to get ideal data with the fixed region and high-quality that eliminates undesired distortions. The acquired iris images are usually of low quality as they are captured under different environmental conditions through a variety of sensing device. To enhance and prepare these images for DL-based authentication models, the dataset is subjected to a series of pre-processing operations [22]. First, the region of interest is segmented from the iris images to remove any extraneous background information. The coloured images are then converted to grey scale to reduce the computing complexity. Indeed, in our approach, the colour feature is not necessary to discriminate between the classes of fake and live modalities as it provides additional information that adds unnecessary complexity and takes up more memory space. The next phase resizes the iris images to a dimension of 224×224 to achieve uniformity. Thereafter, to overcome the problem of lack of inadequate size of anti-spoofing dataset augmentation operations are

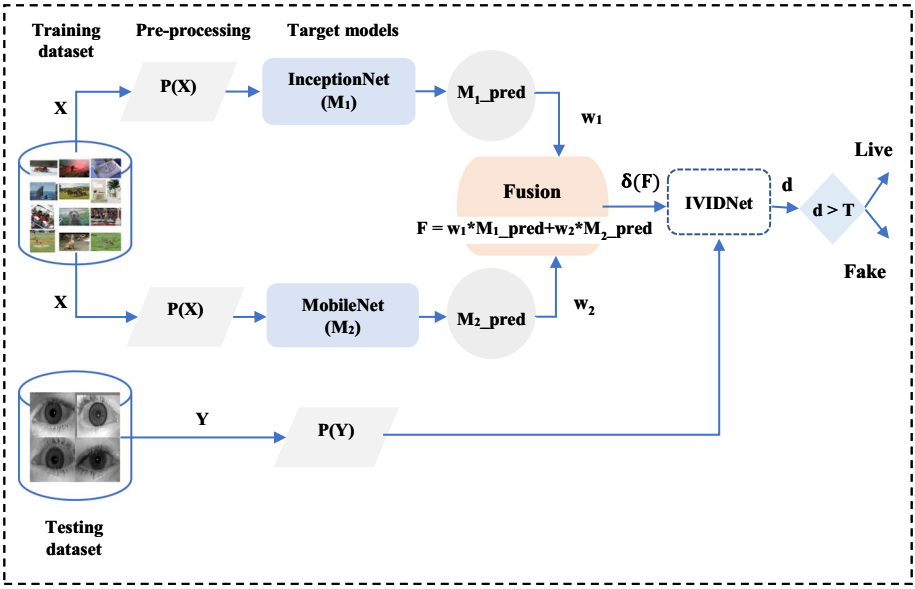


Fig. 1 The framework of our proposed IVIDNet technique

performed. Further, we use feature scaling technique to standardize the independent features present in the data in a fixed range.

3.1.2 Deep feature extraction

Deep feature extraction is the process of extracting image features from the deep layers of a CNN, and the features extracted are known as deep features. This procedure entails first providing the input data to the pre-trained CNN, and then obtaining the relevant activation values from the fully connected layer, which is usually present at the network's end, or the various pooling layers present at different levels. These features extracted from the iris images are used for correct authentication purposes. The fundamental approach which is adopted to extract deep image features along with their pseudo-codes is described in following subsections.

a **Customizing InceptionNet model**

The fundamental way of improving the efficiency of deep neural networks is by increasing their size. This entails expanding the network's depth (number of levels) as well as its breadth. This is a simple and safe technique to train higher-quality models, especially if a large amount of labelled training data is available. However, there are two main downsides to this easy method. The bigger the network, the more parameters it has, which makes it more prone to overfitting, especially if the number of labelled samples in the training set is restricted. This can become a substantial bottleneck, since the creation of high-quality training sets can be difficult. Another issue of uniformly increased network size is the significantly increased use of computational resources. Even inside the convolutions,

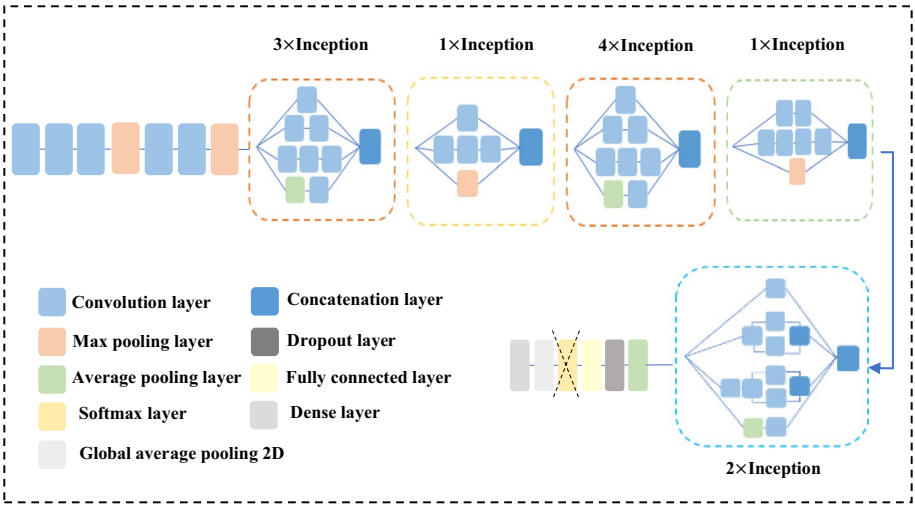


Fig. 2 A generic architecture of customized InceptionNet V3 model

Table 2 A description of customized InceptionNet model

Sr. no	Layer (Type)	Out shape	Param #
1	inception_v3 (Functional)	(None, 5, 5, 2048)	21,802,784
2	global_average_pooling2d (GlobalAverage-Pooling2D)	(None, 2048)	0
3	dense (Dense)	(None, 1)	2049
	Total params: 21,804,833 Trainable params: 2,049 Non-trainable params: 21,802,784		

the fundamental way to solve both difficulties would be to move from fully connected to sparsely connected architectures. An inception network is a deep neural network with an architectural design that consists of repeating components referred to as inception modules as illustrated in Fig. 2. One of the most appealing features of this architecture is that it allows for a large increase in the number of units at each level without an uncontrollable blow-up in computing complexity.

Another practical benefit of this design is that it follows the intuition that visual input should be processed at several scales before being aggregated so that the following step may abstract features from multiple scales simultaneously. The improved utilization of computational resources allows for increasing both the width of each step as well as the number of stages without getting into computational difficulties.

Another way to avail use of the inception architecture is to create slightly inferior, but computationally less expensive variants of it. Further to utilize the capabilities of pre-trained InceptionNet in building the IVIDNet model the customization is done as shown in Table 2. A global average pooling 2D layer is added to the functional Inception V3 model. Finally, a dense layer is added to classify the images as real or fake.

b. **Customizing MobileNet model**

A building block for mobile models are becoming increasingly efficient. As an effective substitute for traditional convolution layers, MobileNet V1 proposed depth-wise separable convolutions. By separating spatial filtering from the feature generation process, depth-wise separable convolutions efficiently factorise conventional convolution. In order to benefit of the low rank nature of the problem, the following generation MobileNet V2 [8] included the linear bottleneck and inverted residual method to construct even more efficient layer structures. In order to improve the expressiveness of non-linear per channel transformations, this structure internally extends to a higher-dimensional feature space while maintaining a compact representation at the input and output. By adding lightweight attention modules based on squeeze and excitation into the bottleneck structure, MnasNet was further built upon the MobileNet V2 framework. To achieve the most effective models for MobileNet V3 [7], it combines these layers as building blocks.

Additionally, layers are improved by enhanced swish nonlinearities. It employs the hard sigmoid in place of the sigmoid, which is used in the nonlinearities of squeeze, excitation, and swish and can be computationally inefficient as well as difficult to maintain accuracy in fixed point arithmetic. Through this procedure, two new MobileNet models: MobileNet V3-Large and MobileNet V3-Small that are oriented toward high and low resource use cases, respectively, are released. Compared to MobileNet V2, MobileNet V3-Large improves ImageNet classification accuracy by 3.2% while lowering latency by 20%. Also, in comparison with MobileNet V2 model with comparable latency, MobileNet V3-Small is 6.6% more accurate.

Figure 3 shows the generic architectural view of MobileNet V2 model. Further to utilize the capabilities of pre-trained MobileNet in building the IVIDNet model the customization is done as shown in Table 3. A global average pooling 2D layer is added to the functional MobileNet model. Finally, a dense layer is added to classify the images as real or fake.

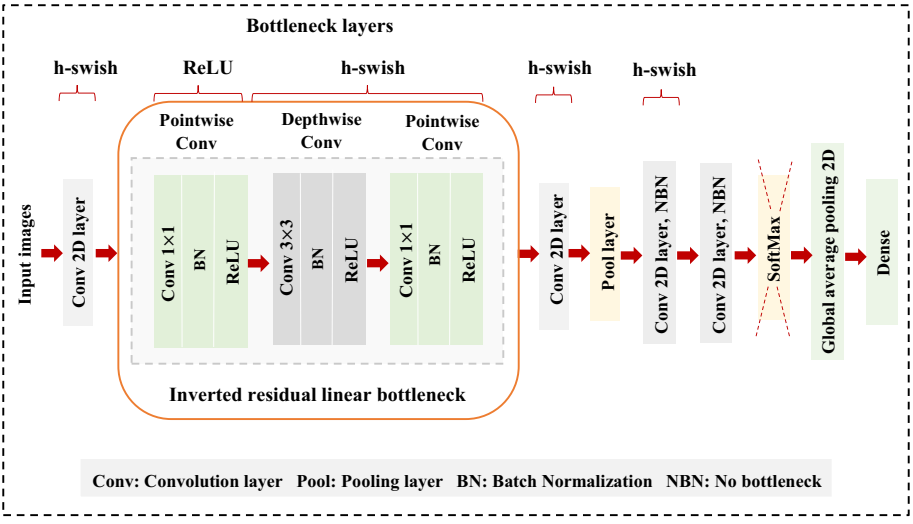


Fig. 3 An illustration of customized MobileNet V2 model

Table 3 A description of customized MobileNet V2 model

Sr. no	Layer (Type)	Out shape	Param #
1	mobilenetv2_1.00_224 (Functional)	(None, 7, 7, 1280)	2257984
2	global_average_pooling2d (GlobalAverage-Pooling2D)	(None, 1280)	0
3	dense_(Dense)	(None, 1)	1281
	Total params: 2,259,265 Trainable params: 1,281 Non-trainable params: 2,257,984		

3.1.3 Weighted score level fusion

In general, a score level fusion is a process of integrating the prediction probabilities of two or more models together. We employed a weighted score level fusion that is one step ahead to the score level fusion, here we specify weights to each model and then fuse the probabilities to get better results.

```
final_results = []
for i in range (len(results_m1)):
     prediction_m1, prediction_m2 = results_m1[i], results_m2[i]
     final_prediction = (0.4 * prediction_m1) + (0.6 * prediction_m2)
     final_results.append(final_prediction[i])
```

First the image features are extracted by using the fine-tuned models, the next task involves fusion of predictions probabilities stated by the models. Figure 4 displays the process of computing the weighted score level fusion for a given set of iris images.

This module further enhances the abilities of the model as the power two models is transferred to one model to efficiently do the classification process. Here the weights are first chosen at random and the one pair that gives the best accuracy is further used to efficiently build the module.

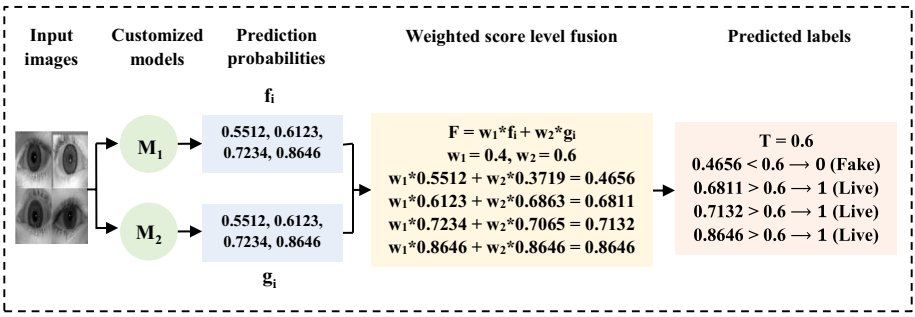


Fig. 4 An example of our proposed weighted score level fusion in IVIDNet

3.1.4 Decision module

The last module of IVIDNet model is the decision module. The output of the fusion module is considered and evaluated against the pre-defined threshold. The samples with prediction probabilities greater than the threshold are considered to be as a live samples and others are considered as fake samples.

3.2 Proposed algorithm

A training algorithm used for learning of VID model is depicted in Fig. 5. Initially, a training set of iris images of size 'X' is chosen form the anti-spoofing dataset 'D_t'.

These acquired images are usually of low quality to enhance and prepare these images for further processing, the dataset is subjected to a series of pre-processing operations. At first the $'\nabla'$ operator is used to covert RGB to grayscale to reduce the computing complexity. Then to obtain uniformity in our model images are resized by $'\mu'$ to the size of 224 X 224. Further augmentation operation is $'\vartheta'$ is performed to artificially increase the amount of data. Finally, images are rescaled ∂ to ensures optimal comparisons across data acquisition methods and texture instances.

After pre-processing of images both the models are fine-tuned the dense $'\lceil'$ layer is dropped and a new 'n' output layer is added to the model 'Q'. Afterwards, the customized models M_1 and M_2 are trained on all the instances of database D_k. The results of the models are integrated as $'\varphi'$. Finally, IVIDNet is built as $'\delta(\varphi)'$ by hyper tuning with optimal parameters. A similar set of steps are followed to pre-process the testing images 't' for validation of IVIDNet model as shown in Fig. 6. Then the testing of

Algorithm 1: To build an IVIDNet model using pre-trained models.

Input: D_t of iris images (Live and Fake)
Output: IVIDNet
Begin
  Let there are X = |Dt| number of mixed sets of training images
    for i = 1 to t do
      Gi (x, y) ← ϰ (Xi (x, y)) // RGB to grayscale images
      Ti (x, y) ← μ (Gi (x, y)) // Resize input images to 224 x 224
      Si (x, y) ← ϑ (Ti (x, y)) // Data augmentation
      Zi (x, y) ← ∂ (Si (x, y)) // Rescaled input images
    end
    for j = 1 to 2 do // Customize the models Mj
      Q ← (Mj – ɗ) // Drop the output layer
      Q ← ∯ (Q) // Freeze the layers
      Q ← (Q + n) // Add the new output layer
    end
    for k = 1 to i do // Train IVIDNet using pre-trained models
      Pick all the instances of database Dk
      Train the models M1 using Dk s.t M1 ← I (Dk)
      Pick all the instances of database Dj
      Train the models M2 using Dk s.t M2 ← V (Dk)
    end
      φ ← ∑ (M1, M2)
  IVIDNet ← δ (φ) // Hyper-tuning with optimal parameters
end

Fig. 5 The learning algorithm for our proposed IVIDNet

models M_j is performed using 'x_test' and prediction sets are generated as 'M_1_pred' and 'M_2_pred. Thereafter we define the weights as w_1 and w_2 and calculate the D_ score as $w_1 * M_1$_pred + $w_2 * M_2$_pred. At last we compare the D_score and defines the labels of the sample as live or fake.

4 The IVIDNet learning and validation

The learning of our IVID approach by using a training algorithm is depicted in Fig. 5. Initially, a training set of iris images of size 'X' is chosen from the benchmark anti-spoofing dataset (D_t). For a model to perform effectively, the training dataset should be large enough to encompass samples from all of the labelled classes as well as sensors. Following that, the training images are pre-processed using the fundamental image processing operations to obtain standardized set of images. Let S_i (x, y) is the ith image of D_t obtained after applying pre-processing operations on corresponding input image X_i (x, y). Thereafter, the selected base models are customized and a new set of layers are added on top of the base models. In the next step we train the new layers on the dataset. Afterwards, a weighted score level fusion model is build using the predictions of two pre-trained models. Finally, the trained model is hyper-tuned to various parameters with proper experimentation at an appropriate search space. Figure 6 lists the steps involved for testing algorithm of the IVIDNet method. The trained IVIDNet is validated by presenting images from the testing datasets. To build the appropriate feature sets, a similar sequence of steps is conducted to test images, such as image pre-processing. Finally, the classification of the test samples is carried out by IVIDNet by assigning a class label as either live or fake.

4.1 Experimental analysis

In this section, we assess the effectiveness of our approach as a vitality detection mechanism. To begin, we provide a brief overview of datasets and evaluation methodologies, which are used as a standard criterion for assessing performance. The IVIDNet is then fine-tuned, and the resulting model is tested on two publicly available datasets, Notre Dame-17 and Notre Dame-15. The proposed method is also evaluated in cross-database scenarios to determine the technique's generalization capability. Finally, the performance of the IVIDNet model is compared against the related state-of-the-art IVID approaches.

4.1.1 Evaluation datasets

An iris anti-spoofing database represents the systematic collection of iris information mainly used for developing and evaluating the iris VID algorithms. An adequate size of database corresponding to different iris sensing technologies and fabrication materials are required to assess these algorithms. For evaluation purposes two benchmark iris anti-spoofing datasets, i.e., LivDet 2017 Notre Dame [24] and NDCLD 2015 [4] are used. The details of these datasets are summarized in Table 4.

Algorithm 2: To validate the IVIDNet models.

Input: Test iris images as Y_t (x, y)
Output: Class label (Fake or Live)

```
Begin
    for i = 1 to t do
        Gi (x, y) ← ϒ (Yi (x, y)) // RGB to grayscale images
        Ti (x, y) ← μ (Gi (x, y)) // Resize input images to 224 x 224
        Zi (x, y) ← ∂ (Ti (x, y)) // Rescaled input images
    end
    for j = 1 to i do // Test the IVIDNet model using pre-trained models
        M1_pred ← M1 (x_test) // Customized InceptionNet model
        M2_pred ← M2 (x_test) // Customized MobileNet model
    end
        w1 = m // Weight for M1
        w2 = n // Weight for M2
        T = 0.6 // pre-defined threshold
    D_score = w1 * M1_pred + w2 * M2_pred
    if D_score > T
        D_score ← Class 1 // Live
    else
        D_score ← Class 2 // Fake
end
```

Fig. 6 The validation algorithm for our proposed IVIDNet

4.1.2 Performance protocols

For performance evaluation, we select the overall protocol related to metrics and appropriate dataset selection. We utilize the training images from both the datasets to learn the IVIDNet and testing samples are selected across various domains to compute the performance.

The model is evaluated in terms of five standard performance metrics, namely attack presentation classification error rate (APCER), bona-fide presentation classification error rate (BPCER), average classification error rate (ACER), average classification accuracy (ACA) and receiver operating characteristic (ROC). A detailed description of various protocols is listed in Table 5.

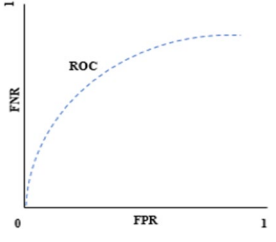
4.2 Experimental setup

Once our dataset for the IVIDNet model is prepared, we can perform the various experiments.

Table 4 An outline of LivDet 2017 Notre Dame and LivDet 2015 NDCLD datasets

S.No	Dataset	Wavelength	Image size	Sensors used	Live	Fake	Types of fakes
1	LivDet 2017 Notre Dame	NIR	640×480	IrisGuard AD100 and IrisAccess LG4000	2400	2400	CL
2	NDCLD 2015	NIR	640×480	IrisGuard AD100 and IrisAccess LG400	_	4068	TCL, SCL

Table 5 An analysis of various metrics used for the evaluation of IVIDNet

S.No	Acronym	Metric	Description	Formula
1	APCER	Attack presentation classification error rate	The ratio of attack presentation with the identical PAI which are misclassified as bonafide presentations to the total number of samples	$\text{APCER} = \frac{\text{FP}}{\text{TN+FP}}$
2	BPCER	Bona-fide presentation classification error rate	The ratio of bonafide presentations presented to the IVD subsystem which are misclassified as attack presentations in a particular situation to the total number of samples	$\text{BPCER} = \frac{\text{FN}}{\text{TP+FN}}$
3	ACER	Average classification error rate	It is the average of APCER and BPCER with a predefined threshold	$\text{ACER} = \frac{\text{APCER+BPCER}}{2}$
4	ACA	Average classification accuracy	It is used to measure the correct classification performance of the given algorithm	$\text{ACA} = \frac{\text{TP+TN}}{N}$
5	ROC curve	Receiver operating characteristic curve	It is a graphical representation of performance of any classification model at all classification thresholds	

4.2.1 Hyper-parameter tuning

The anti-spoofing model's detection accuracy may be significantly impacted by the hyper-parameter settings. Using a Meta-process, the ideal hyperparameters are tuned for each dataset. For IVIDNet, the hyper-parameter such as number of epochs, learning rate, activation function, batch size, and optimizer are chosen. Table 6 contains the outcomes of our proposed model's hyper-parameter settings when trained and tested on Notre-Dame 2017 iris dataset.

In each step of model's tuning the optimum value of hyper-parameter is chosen and fixed for the next step. The process recurs until the whole set of optimal values are acquired. The optimal parameters are further used to evaluate the performance of IVIDNet model in different scenarios.

4.2.2 Performance with different pre-trained models

Pre-trained model is a saved network that was previously trained on a large dataset such as ImageNet, typically for a large-scale image classification task. Further these models could either be used as it is or we can transfer the knowledge to customize it to perform related tasks. To build an IVIDNet we have customized the most efficient models among all the models trained for image classification problems. To confirm that the best models are selected we have compared five well known pre-trained model that are publicly available: InceptionNet, EfficientNet [19], VGG-16, VGG-19, and MobileNet. The comparative analysis of various fine-tuned model when trained and validate on Notre Dame 2017 iris anti-spoofing dataset is summarized in Table 7.

Table 6 Performance evaluation of IVIDNet at different parameter settings

S.No	Hyper-parameters	Search space	Parameter setting	Model performance (%)	Selected value
1	Epochs	[20, 30, 40, 50, 60]	20	96.25	40
			30	98.95	
			40	99.33	
			50	98.89	
			60	99.00	
2	Learning rate	[0.1, 0.01, 0.001]	0.1	99.50	0.1
			0.01	99.44	
			0.001	98.67	
3	Optimizer	['SGD', 'Adam', 'RMSprop']	SGD	99.39	Adam
			Adam	99.50	
			RMSprop	99.06	
4	Activation function	['tanh', 'ReLU', 'sigmoid']	tanh	49.91	sigmoid
			ReLU	50.08	
			sigmoid	99.50	
5	Batch size	[32, 64]	32	99.39	32
			64	99.22	

Table 7 Comparison of performance among various fine-tuned models

S.No	Model	Dataset	Accuracy (%)	Val accuracy (%)	Loss (%)	Val loss (%)
1	VGG-16	Notre Dame-17	79.40	77.08	99.77	58.39
2	VGG-19	Notre Dame-17	92.65	96.25	18.94	8.43
3	InceptionNet V3	Notre Dame-17	96.27	98.33	68.97	38.25
4	EfficientNet V2	Notre Dame-17	66.25	50.00	383.32	218.33
5	MobileNet V2	Notre Dame-17	99.69	100	2.41	0.000067804

The accuracy rate is measured at the selected values of hyper-parameter from the preceding step. From the comparison it can be clearly inferred that MobileNet V2 and InceptionNet V3 are the two models with maximum training and validation accuracy rate. The performance measure of these two models with the help of graphs is depicted in Fig. 7.

4.2.3 Performance of IVIDNet before and after fusion

A solitary fine-tuned model could be seen as a potent and accurate tool for efficient image classification purposes. Fusion of more than one such model could further influence the accuracy of a classifier. This step could either increase the accuracy or increase the complexity of resultant model. Table 8 shows the performance measure of a sole fine-tuned models and weighted fused IVIDNet.

From the outcomes we can infer that in comparison to the sole models the accuracy of the fused IVIDNet model is increased in both testing in known and unknown scenario of

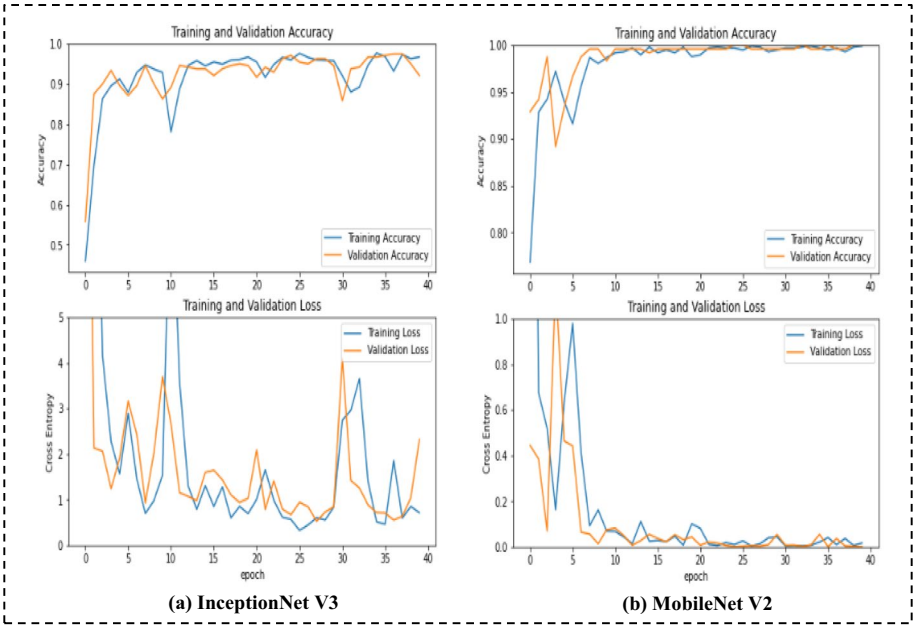


Fig. 7 A graphical representation of InceptionNet and MobileNet performance

Table 8 Performance measure of customized models before and after fusion

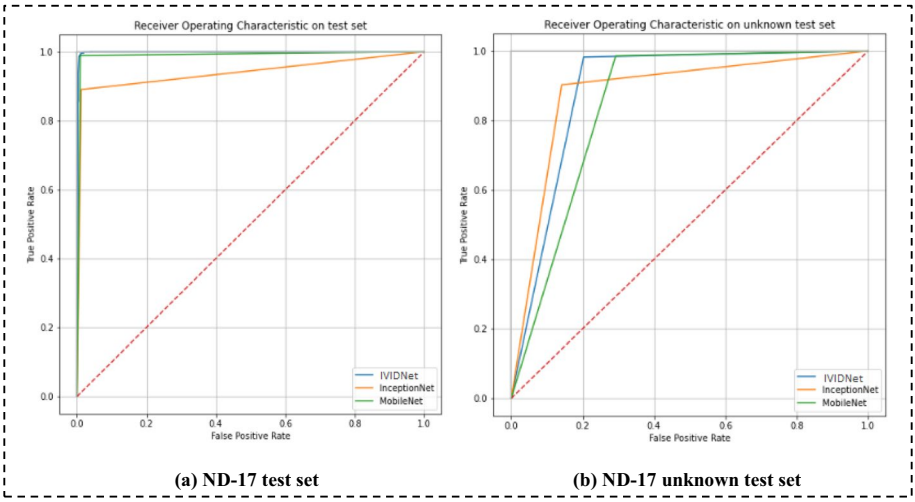
S.No	Models	Training dataset	Testing dataset (Notre Dame-17)		APCER (%)	BPCER (%)	ACER (%)	ACA (%)
			Test	Unknown test				
1	InceptionNet V3	Notre Dame-17	✓	×	4.03	0.68	2.35	97.57
2	InceptionNet V3	Notre Dame-17	×	✓	13.70	10.10	11.90	88.02
3	MobileNet V2	Notre Dame-17	✓	×	1.51	0.22	0.86	99.11
4	MobileNet V2	Notre Dame-17	×	✓	23.13	1.97	12.55	84.54
5	IVIDNet	Notre Dame-17	✓	×	0.87	0.33	0.60	99.39
6	IVIDNet	Notre Dame-17	×	✓	17.30	2.15	9.72	88.90

the Notre Dame 2017 datasets. It is a graphical representation of performance of any classification model at all classification thresholds could be represented with the help of ROC curve. Figure 8 illustrates the ROC curve for IVIDNet model.

4.2.4 Performance at varying weights

The proposed model is based on the idea of weighted score level fusion of prediction from the two customized models. Here 'w1' and 'w2' corresponds to the weights assigned to InceptionNet V3 and MobileNetV2 respectively.

Table 9 depicts the contrast among the performance measure of the model at different values of weights. The performance is measured on the training set of Notre Dame 2017 iris anti-spoofing dataset. Based on the classification accuracy of the final IVIDNet model the optimal value for weights could be 0.4 and 0.6 for InceptionNet V3 and MobileNet V2 model respectively. Figure 9 shows the ROC curve of IVIDNet with different values of weights.



(a) ND-17 test set **(b) ND-17 unknown test set**

Fig. 8 The ROC curve of IVIDNet tested on known and unknown set

Table 9 Comparative analysis of IVIDNet at varying weights

S.No	Weights		APCER (%)	BPCER (%)	ACER (%)	ACA (%)
	w1	w2				
1	0.7	0.3	4.03	0.68	2.35	97.57
2	0.6	0.4	2.37	0.56	1.46	98.51
3	0.5	0.5	0.44	0.76	0.60	99.39
4	0.4	0.6	0.87	0.33	0.60	99.39
5	0.3	0.7	1.51	0.22	0.86	99.11

4.2.5 Performance with cross dataset scenarios

An IVD's efficiency in terms of generalizability to unknown assaults is a critical part. We thus conduct an experiment to assess how well our technique performs in a cross-database scenario. Cross-dataset testing, is where a model is trained on one set and evaluated on distinct datasets entailing iris artefacts created by different spoofing materials, is used to extend the IVD approach across unknown threats. In this test, we used images from a set to train our model and samples from another set to test it. Table 10 presents the results at cross-dataset evaluations.

The contrast among the performance of the model when trained on Notre Dame 2017 and tested on NDCLD 2015 at different values of epochs is summarized in Table 10. From there we can be inferred that the IVIDNet achieves an accuracy rate of 89.63% on 40 epochs at cross dataset evaluation.

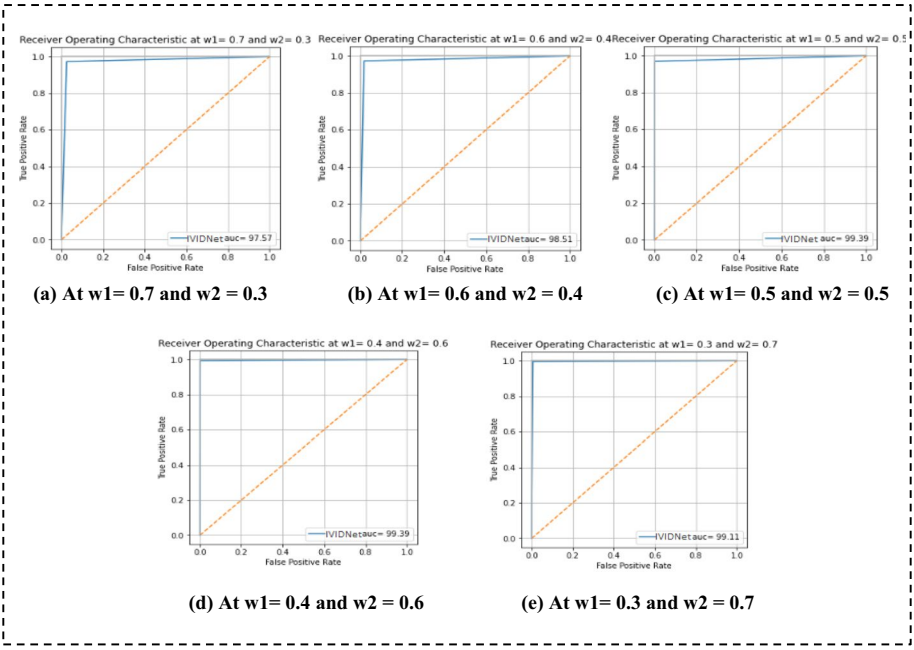


Fig. 9 The ROC curve of IVIDNet by varying weights

Table 10 Cross-dataset performance of the IVIDNet

S.No	Epochs	Training dataset	Testing dataset	InceptionNet performance (%)	MobileNet performance (%)	IVIDNet performance (%)
1	10	Notre Dame-17	NDCLD-15	74.53	72.72	81.38
2	15	Notre Dame-17	NDCLD-15	77.79	74.12	83.01
3	20	Notre Dame-17	NDCLD-15	77.97	80.93	86.23
4	25	Notre Dame-17	NDCLD-15	79.20	82.54	87.52
5	30	Notre Dame-17	NDCLD-15	79.50	81.26	88.09
6	35	Notre Dame-17	NDCLD-15	81.09	84.09	88.71
7	40	Notre Dame-17	NDCLD-15	82.78	86.35	89.63

4.3 Comparison with SOA techniques

Several machine and deep learning-based solutions are presented in the literature to address the issue of iris anti-spoofing. Since our method is transfer learning-based and grounded on the concepts of weighted fusion of prediction values of two fine-tuned model at score level. In order to assess the IVIDNet's effectiveness, we contrast it with comparable SOA techniques that are based on similar techniques. Table 11 depicts the contrast between our proposed approach and a multi-layer fusion technique trained and evaluated on Notre Dame dataset using a pre-trained model.

The comparison clearly indicates that our proposed model performs more efficiently compared to the multi-layer fusion method.

The comparison of proposed IVIDNet with SOA iris spoof detection mechanisms is briefly discussed in Table 12. The outcomes indicate that our approach performs well in known environment with ACA of 99.39% and shows descent results in unknown attack scenarios.

5 Conclusions

This research work has presented an efficient and novel iris liveness detection mechanism that fuses the robust features of two pre-trained DCNN models (InceptionNet V3 and MobileNet V2). The IVIDNet has been evaluated by conducting a series

Table 11 A comparison of IVIDNet with related approach

S.No	Comparison basis	Multi-layer fusion [5]	IVIDNet (ours)
1	Key concept	Multi-layer fusion (at 4 layers)	Weighted score level fusion
2	Pre-trained model	VGG-16	InceptionNet V3 + MobileNet V2
3	Datasets	Notre Dame 2017 and IIITD-WVU	Notre Dame 2017 and NDCLD-15
4	Testing dataset	Notre Dame 2017	Notre Dame 2017
5	Testing accuracy (%) on ND-17 (test)	APCER = 2.72	APCER = 0.87
		BPCER = 1.89	BPCER = 0.33
		ACER = 2.31	ACER = 0.60
6	Testing accuracy (%) on ND-17 (unknown-test)	-	APCER = 17.30
		-	BPCER = 2.15
		-	ACER = 9.72

Table 12 IVIDNet's comparison with SOA TL-based anti-spoofing approaches

S.No	Technique	Datasets	Models performance	
			Known	Unknown
1	Chen and Ross [2]	ND-Contact-13, CASIS iris, LivDet iris-15 Warsaw, LivDet iris-17 clarkson, BERC-Iris Fake	BPCER: ND-Contact, CASIS iris = 0.5%, LivDet iris-15 Warsaw = 0%	BPCER for Cross sensor: BERC = 1.66%, LivDet iris-17 clarkson = 20.44%
2	Alaslani and Elrefaei [1]	IIITD, CASIA-Iris -thousand, CASIA-IrisV1 and CASIA-Iris-Interval	Recognition accuracy: IIITD iris = 100%, CASIA-Iris thousand = 95%, CASIA-IrisV1 = 98.3%, and CASIA-Iris-Interval = 91.6%	_
3	Choudhary et al. [3]	IIITD-CLI, Notre Dame (ND)-13	CCR for multi sensor validation: ND combined = 96.04%, IIITD combined = 94.93%,	_
4	Therar et al. [21]	IITD and CASIA-IrisV3	Accuracy rate: IITD = 99% (left and right iris both), CASIA-IrisV3 = 94% and 93% (left and right iris respectively)	_
5	Sardar et al. [13]	CASIA-Iris-V4-Interval, IITD, NICE.I	NICE.I: mTRP = 0.983% MER = 0.261%	_
6	Tapia J. et al. [20]	LivDet-Iris 2020	ACER = 29.78%	_
7	IVIDNet (proposed)	Notre Dame 2017, NDCLD 15	Accuracy (ACA) trained and tested on Notre Dame 2017 (test set) = 99.39	Cross dataset: Accuracy (ACA) trained on Notre Dame 2017 and tested on Notre Dame (unknown-test set) 2017 = 88.90

of experiments on benchmarks anti-spoofing datasets and the empirical results proves the effectiveness of our approach. Besides, known attack scenarios the IVIDNet shows promising performance in unknown attack environment covering cross-database scenario. The IVIDNet offers several merits as compared to which other counterparts that include: (i) usage of robust deep level features from pre-trained models (ii) works well with smaller datasets (iii) lower training overhead due to using pre-trained models. However, the proposed IVIDNet approach has not been evaluated in cross-material and cross-sensor environments. The future scope of this work may include the performance evaluation of IVIDNet on some more recent iris anti-spoofing datasets such as LivDet 2017, LivDet 2020, IIITD, and etc. An additional future work is to evaluate IVIDNet in new unknown attack environments particularly unknown fabrication materials used for creating iris artefacts and sensors. The proposed approach may be extended for anti-spoofing techniques in other biometrics such as fingerprint, face, palm prints, etc.

Acknowledgements The iris LivDet datasets Notre Dame 2017 and NDCLD 2015 employed in this research is kindly provided by Notre Dame University for which authors express their gratitude.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interest All the authors declare that they do not have any conflict of interest.

References

1. Alaslni MG, Elrefaei LA (2019) Transfer Learning with Convolutional Neural Networks for IRIS Recognition. Int J Artif Intell Appl 10(5):49–66. https://doi.org/10.5121/ijaia.2019.10505
2. Chen C, Ross A (2018) A Multi-Task Convolutional Neural Network for Joint Iris Detection and Presentation Attack Detection. (March)
3. Choudhary M, Tiwari V, Venkanna U (2019) An approach for iris contact lens detection and classification using ensemble of customized DenseNet and SVM. Futur Gener Comput Syst 101:1259–1270. https://doi.org/10.1016/j.future.2019.07.003
4. Doyle JS, Bowyer KW (2015) Robust Detection of Textured Contact Lenses in Iris Recognition Using BSIF. IEEE Access 3:1672–1683. https://doi.org/10.1109/ACCESS.2015.2477470
5. Fang M, Damer N, Boutros F, Kirchbuchner F, Kuijper A (2020) Deep learning multi-layer fusion for an accurate iris presentation attack detection. Proc 2020 23rd Int Conf Inf Fusion, FUSION 2020. https://doi.org/10.23919/FUSION45008.2020.9190424
6. Gautam G, Mukhopadhyay S (2018) Contact Lens Detection using Transfer Learning with Deep Representations. Proc Int Jt Conf Neural Netw 2018-July:1–8. https://doi.org/10.1109/IJCNN.2018.8489590
7. Howard A et al (2019) Searching for mobileNetV3. Proc. IEEE Int. Conf. Comput. Vis., vol. 2019-Octob, pp. 1314–1324. https://doi.org/10.1109/ICCV.2019.00140
8. Howard AG et al (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. [Online]. Available: http://arxiv.org/abs/1704.04861
9. Iris Recognition Market by Component (Hardware, and Software), Product (Smartphones, Scanners), Application (Identity Management and Access Control, Time Monitoring, E-payment), Vertical, and Region - Global Forecast to 2024. 2019. https://www.marketsandmarkets.com/Market-Reports/iris-recognition-market-141994093.html#:~:text=. The global iris recognition market, 13.2%25 during the forecast period
10. Minaee S, Abdolrashidi A (2019) DeepIris: Iris recognition using a deep learning approach. [Online]. arXiv preprint arXiv:1907.09380. http://arxiv.org/abs/1907.09380

11. Ratha NK, Connell JH, Bolle RM (2001) An analysis of minutiae matching strength. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2091(May 2015):223–228. https://doi.org/10.1007/3-540-45344-x_32
12. Ribeiro E, Uhl A (2017) Exploring Texture Transfer Learning via Convolutional Neural Networks for Iris Super Resolution. Lect Notes Informatics (LNI) Proc - Ser Gesellschaft fur Inform (00736) 0–4. https://doi.org/10.23919/BIOSIG.2017.8053505
13. Sardar M, Banerjee S, Mitra S (2020) Iris Segmentation Using Interactive Deep Learning. IEEE Access 8:219322–219330. https://doi.org/10.1109/ACCESS.2020.3041519
14. Sharma D, Selwal A (2021) On data-driven approaches for presentation attack detection in Iris recognition systems. In: Singh PK, Singh Y, Kolekar MH, Kar AK, Chhabra JK, Sen A (eds) Recent Innovations in Computing. ICRIC 2020. Lecture Notes in Electrical Engineering, vol 701. Springer, Singapore. https://doi.org/10.1007/978-981-15-8297-4_38
15. Sharma D, Selwal A (2021) An intelligent approach for fingerprint presentation attack detection using ensemble learning with improved local image features, no. 0123456789. Springer US
16. Sharma D, Selwal A (2022) HyFiPAD : a hybrid approach for fingerprint presentation attack detection using local and adaptive image features. Vis Comput 38:2999–3025. https://doi.org/10.1007/s00371-021-02173-8
17. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–14
18. Szegedy C et al (2015) Going deeper with convolutions. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 07-12-June:1–9. https://doi.org/10.1109/CVPR.2015.7298594
19. Tan M, Le QV (2019) EfficientNet: Rethinking model scaling for convolutional neural networks, 36th Int. Conf. Mach. Learn. ICML 2019. 2019-June: 10691–10700
20. Tapia JE, Gonzalez S, Busch C (2022) Iris Liveness Detection Using a Cascade of Dedicated Deep Learning Networks. IEEE Trans Inf Forensics Secur 17:42–52. https://doi.org/10.1109/TIFS.2021.3132582
21. Therar HM, Mohammed LDEA, Ali APDAJ (2021) Multibiometric System for Iris Recognition Based Convolutional Neural Network and Transfer Learning. IOP Conf Ser Mater Sci Eng 1105(1):012032. https://doi.org/10.1088/1757-899x/1105/1/012032
22. Verma P, Selwal A, Sharma D (2022) "An exploration of pre-processing approaches for iris spoof detectors," in 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES). pp. 1–8. https://doi.org/10.1109/CISES54857.2022.9844401
23. Verma P, Selwal A, Sharma D (2022) A survey on data-driven iris spoof detectors: state-of-the-art, open issues and future perspectives. Multimed Tools Appl. https://doi.org/10.1007/s11042-022-14014-4
24. Yambay D et al (2018) LivDet iris 2017 - Iris liveness detection competition 2017. IEEE Int. Jt. Conf. Biometrics, IJCB 2017. 2018-Janua: 733–741. https://doi.org/10.1109/BTAS.2017.8272763