IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 May 2023, accepted 8 June 2023, date of publication 13 June 2023, date of current version 19 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3285826

RESEARCH ARTICLE

A Novel Deep Learning Architecture With Image Diffusion for Robust Face Presentation Attack Detection

MADINI O. ALASSAFI[1], MUHAMMAD SOHAIL IBRAHIM[2], IMRAN NASEEM[3,4,5], RAYED ALGHAMDI[1], REEM ALOTAIBI[1], FARIS A. KATEB[1], HADI MOHSEN OQAIBI[1], ABDULRAHMAN A. ALSHDADI[6], AND SYED ADNAN YUSUF[7]

[1]Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 25732, Saudi Arabia
[2]College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China
[3]School of Electrical, Electronic, and Computer Engineering, The University of Western Australia, Perth, WA 6907, Australia
[4]College of Engineering, Karachi Institute of Economics and Technology, Karachi 75190, Pakistan
[5]Research and Development, Love for Data, Karachi 75600, Pakistan
[6]College of Computer Science and Engineering, Department of Information Systems and Technology, University of Jeddah, Jeddah 21959, Saudi Arabia
[7]Research and Innovation Department, Intelexica Pvt. Ltd., SO15 2RZ London, U.K.

Corresponding author: Muhammad Sohail Ibrahim (msohail@zju.edu.cn)

This work was supported in part by the Institutional Fund Projects under Grant IFPRC-044-611-2020; and in part by the Ministry of Education and King Abdul Aziz University, Jeddah, Saudi Arabia.

ABSTRACT Face presentation attack detection (PAD) is considered to be an essential and critical step in modern face recognition systems. Face PAD aims at exposing an imposter or an unauthorized person seeking to deceive the authentication system. Presentation attacks are typically made using a fake ID through a digital/printed photograph, video, paper mask, 3D mask, and make-up etc. In this research, we propose a novel face PAD solution using an interpolation-based image diffusion augmented by transfer learning of a MobileNet convolutional neural network. The proposed interpolation-based image diffusion method and face PAD approach, implemented in a single framework, shows promising results on various anti-spoofing databases. The experimental results illustrate that the proposed face PAD method shows superior performance compared to most of the state-of-the-art methods.

INDEX TERMS Anti-spoofing, deep learning, face liveness detection, face presentation attack detection, interpolation-based image diffusion.

I. INTRODUCTION

Biometric authentication systems such as face recognition and fingerprint have gained a lot of popularity over the past decade due to the increased security and reliability compared to the conventional password-based authentication systems. Face recognition, in particular, due to its non-intrusive nature, high accuracy, and usability has led to its applications in various domains including surveillance [1], classroom attendance systems [2], school examination monitoring [3], mobile phone unlock [4], and access control systems [5] to name a few. This has been made possible due to the availability of large databases and greater computing power due to which, many deep learning-based automatic face recognition systems have been reported to achieve accuracy of over 99% [6]. However, face authentication systems suffer from an intrinsic drawback of false acceptance which can entail a security risk due to the possibility that the system security can be vulnerable under a spoofing attack by a malicious adversary/imposter attempting to spoof the face recognition system.

Face spoofing attacks or face presentation attacks are referred to attacks where an adversary obtains unauthorized access to a face recognition system by camouflaging/imposing as an authorized person. Acquiring facial images using social media has also enabled the attackers to spoof the face recognition systems using a variety of attacks such as print photo attacks, recorded facial videos, and 3D

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

mask attacks. Therefore, the demand of efficient face anti-spoofing systems or face presentation attack detection (PAD) systems has also risen to alleviate the spoofing risks to the face recognition applications.

During the past decade, a large number of face PAD methods have been reported. These methods put their prime focus towards the exploration of efficient features for face PAD. Such methods can be divided into different categories such as distortion, motion, texture, and deep learning-based face PAD techniques. Each of the aforementioned techniques have made important contributions in the performance of face PAD systems. Distortion-based techniques focus on utilizing the distortion sensitive features to perform face anti-spoofing. Motion based techniques focus on extracting motion features such as blinking of the eyes, optical flow, head movement, and lip movement to perform face liveness detection. Texture-based face PAD techniques aim to adopt texture features such as local binary patterns (LBP) to detect face spoofing. Deep learning-based techniques perform anti-spoofing by learning the feature representation using deep neural networks.

Early face PAD solutions rely heavily on the motion-based features such as movement of the head, blinking of the eyes, and lip movement to detect 2D face spoofing attacks. In [7], a face detection technique leveraging mouth localization and utilizing lip motion analysis for face detection and liveness detection by incorporating AdaBoost and SVM, respectively, was presented. The authors in [8] performed liveness detection using the difference of optical flow fields generated by the movement of 3D objects (such as a face) and 2D planes (such as a printed photograph). A face liveness detection technique that analyzes the correlation between the background and the fore-ground using optical flow is presented in [9]. Motion-based face PAD techniques also use eye-blinking movements [10], [11], dynamic facial textures [12], nose and ear movement [13] among other facial movement cues.

Most of the face PAD schemes presented in the literature focus on the detection of replay and print spoofing attacks, which can be addressed using the texture and color features. A number of early works incorporate hand-crafted features such as color texture features and local binary patterns (LBP) [14], [15], [16], histogram of oriented gradients (HOG) [17]. Other texture-based methods employed scale-invariant feature transform [4], [18], speeded-up robust feature (SURF) [19], optical flow and texture analysis [20], etc.

Despite the recent trend towards detection of 3D mask attacks [21], most works in literature focus on the detection of 2D spoofing attacks such as print photo attack, replay attacks, masking attacks, etc. This is due to the reason that most 3D face PAD approaches rely on the introduction of additional hardware (cameras for multiple channels e.g. thermal, NIR, etc.) and also due to the computational complexity of such approaches, it makes such approaches infeasible for application in low-cost systems. Wenyun Sun et. al. [22] proposed to revisit the face PAD task using local label supervision where the pixel-level spoofing cues are classified using the spoof fore-ground, genuine fore-ground, and the genuine background by a depth-based fully convolutional network (FCN). A face PAD method that utilized a shape-from-surface algorithm to extract intrinsic properties such as depth, reflectance, and albedo of the faces followed by a novel shallow CNN architecture to learn useful features for the face PAD task is presented in [23]. The work in [24] presented an approach that uses a mix of real-word face images and deep convolutional autoencoder generated images followed by a CNN feature fusion layer to balance the fusion, in an adaptive fashion, of the two images to efficiently perform face anti-spoofing. A CNN-based face PAD approach that learns the dynamic disparity maps based hand-crafted features within the network using an additional disparity layer in the custom CNN architecture is presented in [25]. The use of multiple channels such as near-infrared (NIR), thermal, etc., besides the visual spectra for face PAD has also been reported in the literature. The authors of [26] presented a multi-channel CNN-based technique that uses four channels namely color, NIR, depth, and thermal to address the 2D and 3D face PAD problem. A hybrid approach that uses a region based CNN and an improved Retinex-based LBP in a cascade fashion to perform face PAD is presented in [27].

In many deep learning and computer vision applications, transfer learning has been readily used for the extraction of deep convolutional features. This has been made possible with the availability of very deep CNN architectures where complex and discriminative features can be extracted using pre-trained models or fine-tuning. These pre-trained CNN architectures provide excellent feature extraction capabilities. While most of the existing works design and train the CNN models from scratch using face PAD datasets, such models usually suffer from overfitting due to the unavailability of large training datasets. In order to address such overfitting problems and enhance the performance of computer vision and deep learning tasks, the use of pre-trained models and fine-tuning deep CNN models from large image classification databases such as ImageNet [28], has been actively reported in literature. In this regard, a two-stream CNN-based face PAD technique which leverages a pre-trained ResNet18 model to learn the features from RGB space and an illumination-invariant space to be provided to an attention-based feature fusion mechanism for efficient face PAD performance, is presented in [29]. The use of pre-trained deep CNN models in face PAD approaches that use multiple channels such as RGB, depth, NIR, thermal, etc. as an input to a pre-trained multi-channel CNN (MC-CNN) network followed by a small network consisting of a few layers to perform the classification of spoof vs. real faces have also been presented in the literature [26], [30]. The hybrid approach presented in [27] also uses deep features extracted from a pre-trained VGG16 model which has been used as a base network, as well as illumination features extracted using an improved Retinex algorithm for the face anti-spoofing task. The work presented in [31] also proposed to

use a pre-trained InceptionV4 model as a liveness feature extraction network in their face PAD framework.

In this context, however, most of the studies do not explore the potential of some state-of-the-art deep CNN models for e.g. MobileNet [32] in the face PAD problem. In this research, we propose a framework for efficient deep learning-based face PAD for 2D attacks. In the proposed approach, a diffusion method is implemented that uses an interpolation-based method to perform image diffusion to enhance the distinguishing features in the real and spoof images. Since image diffusion is a process to introduce smoothness into an image, and it can be viewed as applying a Gaussian filter on an image [33]. Therefore, this research presents an interpolation-based technique to generate diffused images. The diffused images are then fed to a deep CNN followed by a detection model which classifies real and spoof images. This study presents an end-to-end approach where diffusion and the deep CNN are combined into a single model.

The rest of the paper is organized as: the proposed method is outlined in section II, the details of datasets and performance metrics are provided in section III followed by the experimental results and conclusion in sections IV and V respectively.

II. PROPOSED METHOD

In this paper, we propose a face PAD approach by designing an interpolation-based image diffusion mechanism followed by a deep CNN-based face PAD network using a pre-trained MobileNet [32] as our base model.

A. IMAGE DIFFUSION

Image diffusion is a process where input images are smoothed either at a constant rate (linear diffusion) [33], where the smoothness is achieved at a constant rate throughout the image, or in a nonlinear fashion where the important image features such as edges are retained in the diffused image [34], [35]. In computer vision applications, linear diffusion is among the oldest and most investigated partial differential equation method which can be seen as an evolution process where an image is diffused/smoothed in all directions at a constant rate. Such diffusion processes tend to suppress the finer scale structural details in the image subjected to diffusion.

Contemporary image diffusion schemes include linear and non-linear diffusion models. Gaussian smoothing is considered the most popular diffusion scheme among the linear diffusion schemes [36]. Among the non-linear diffusion schemes, the Perona-Malik diffusion [34] or commonly known as anisotropic diffusion is considered the most widely used image diffusion technique in image processing. Other non-linear diffusion techniques include continuous diffusion filtering, semi-discrete diffusion filtering, and discrete diffusion filtering schemes [36]. Other image diffusion models include hybrid image diffusion [37], modified Perona-Malik diffusion model [38], [39], etc.

The proposed diffusion strategy is carried out as a pre-processing step where the captured frames are subjected to compression using inter-area interpolation scheme. During the compression stage, the inter-area interpolation technique calculates the ratios of the output image height and width to input image height and width termed as $scale_x$ and $scale_y$. The product of these ratios is then used to calculate $Area = scale_x \times scale_y$. Then depending on the value of the ratios, the number of corresponding pixels from each channel are selected to form an array of pixels for each channel. These selected pixel arrays are then added and divided by $Area$ to calculate the corresponding output image pixels. The values of ratios can either be integer or fractional. If the ratios are non-integer, then the sum is taken as a weighted sum and the value of weight for each pixel depends on the percentage of a pixel in a particular pixel array. The weight for each pixel is calculated as the ratio of the percentage of the pixel being in the pixel array to $Area$.

The compressed frames are then expanded keeping the pixel-area relationships into consideration thereby resulting in smooth/diffused images. The proposed image diffusion method enhances the class discrimination cues in the images. This is due to the reason that, in general, image diffusion is mainly performed for noise removal while preserving the important information such as lines, edges, and other content that is vital for the interpretation of the image [34]. In general, noise reduction can remove significant information from an image. Therefore, using inter-area interpolation, where pixel-area relationships are used for re-sampling can remove the noise content as well as preserve the useful information in the image and generates images free from Moires' patterns.

To verify the effectiveness of the proposed diffusion method, in Fig. 1 PCA embedding for CASIA-FASD [40] database using the proposed diffusion is visualized in comparison with anisotropic diffusion [34]. The green regions in the figure represent the real class samples while the red regions represent the attack/spoof class samples. It can be observed that the PCA embedding of the proposed technique show better class separability compared to that for anisotropic diffusion (i.e. less overlap between the red and green regions). It is also noteworthy that the class separability shown in PCA embedding has been achieved in image domain which can essentially aid in CNN training to obtain superior classification ability. A comparison of different pre-processing methods detailing different diffusion techniques is discussed in section IV-A.

B. DEEP CNN-BASED FACE PAD MODEL

The proposed face PAD approach harnesses the concept of transfer learning using a pre-trained deep CNN. The methodology is motivated by the fact that the available face PAD datasets are typically not sufficient for training a deep CNN from scratch. Transfer learning is a technique where the learned knowledge from a deep network trained on one task is passed on to another network [41]. Transfer learning can also overcome the overfitting problem caused by insufficient

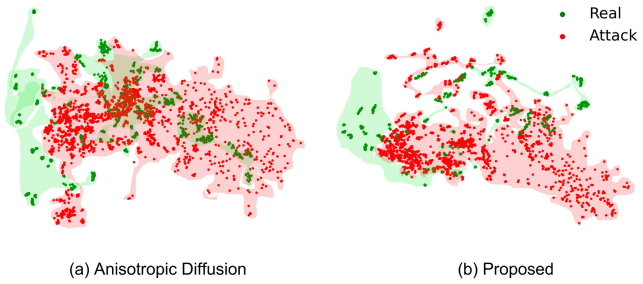

FIGURE 1. PCA embeddings for CASIA-FASD database, (a) Anisotropic diffusion, (b) Proposed diffusion method.

and imbalanced training data. Transfer learning can be used in two ways; (1) as a feature extractor [42] and (2) fine tuning the source model [43].

For the proposed face PAD approach, we used a pre-trained MobileNet [32] deep CNN architecture that was trained on ImageNet dataset [28]. We used the MobileNet architecture in a fine-tuning setting, where pre-trained weights of ImageNet are used for fine-tuning the model on the face PAD databases. The use of transfer learning a deep CNN model is motivated by the strong feature extraction capability as well as the reduced computational requirements of such training methods. An experiment for the performance evaluation of the two transfer learning strategies mentioned above. The findings of the experiment suggest that fine tuning a pre-trained model results in better performance. The details of the experiment are presented in Section IV-B. The bock diagram of the proposed deep CNN-based face PAD network is presented in Fig. 2.

The MobileNet model is based on depth-wise separable convolution [44] where a standard convolution is divided into a depth-wise convolution and piece-wise convolution. The filtering and combining of inputs into an output are done in a single step in standard convolution, whereas the depth-wise separable convolution splits this task into two layers thereby significantly reducing the model size and the number of computations performed in the model. A standard convolution takes $D_A \times D_A \times M$ input features and yields an output feature map of the dimensions $D_A \times D_A \times N$, which results in the number of computations as:

$$D_k \cdot D_k \cdot M \cdot N \cdot D_A \cdot D_A \tag{1}$$

where $D_k \times D_k$ is the kernel size, M is the number of input channels, N is the number of output channels, and $D_A \times D_A$ is the size of the feature map. 3×3 depth-wise separable convolutions are used in MobileNet architecture which can reduce the computations substantially. Using depth-wise separable convolution results in a computational cost of

$$M \cdot D_k \cdot D_k \cdot D_A \cdot D_A + N \cdot M \cdot D_A \cdot D_A \tag{2}$$

which can be simplified as

$$M \cdot D_A \cdot D_A \cdot (D_k \cdot D_k + N) \tag{3}$$

TABLE 1. Proposed face PAD model architecture.

Layer Type / Strides	Input Shape	Output Shape
Diffusion	$30 \times 40 \times 3$	$244 \times 244 \times 3$
Conv2D / 2	$244 \times 244 \times 3$	$112 \times 112 \times 32$
Conv2D Depth-wise / 1	$112 \times 112 \times 32$	$112 \times 112 \times 32$
Conv2D Point-wise / 1	$112 \times 112 \times 32$	$112 \times 112 \times 64$
Zero Padding	$112 \times 112 \times 64$	$113 \times 113 \times 64$
Conv2D Depth-wise / 2	$113 \times 113 \times 64$	$56 \times 56 \times 64$
Conv2D Point-wise / 1	$56 \times 56 \times 64$	$56 \times 56 \times 128$
Conv2D Depth-wise / 1	$56 \times 56 \times 128$	$56 \times 56 \times 128$
Conv2D Point-wise / 1	$56 \times 56 \times 128$	$56 \times 56 \times 128$
Zero Padding	$56 \times 56 \times 128$	$57 \times 57 \times 128$
Conv2D Depth-wise / 2	$57 \times 57 \times 128$	$28 \times 28 \times 128$
Conv2D Point-wise / 1	$28 \times 28 \times 128$	$28 \times 28 \times 256$
Conv2D Depth-wise / 1	$28 \times 28 \times 256$	$28 \times 28 \times 256$
Conv2D Point-wise / 1	$28 \times 28 \times 256$	$28 \times 28 \times 256$
Zero Padding	$28 \times 28 \times 256$	$29 \times 29 \times 256$
Conv2D Depth-wise / 2	$29 \times 29 \times 256$	$14 \times 14 \times 256$
Conv2D Point-wise / 1	$14 \times 14 \times 256$	$14 \times 14 \times 512$
$5\times$ Conv2D Depth-wise / 1 $5\times$ Conv2D Point-wise / 1	$14 \times 14 \times 512$ $14 \times 14 \times 512$	$14 \times 14 \times 512$ $14 \times 14 \times 512$
Zero Padding	$14 \times 14 \times 512$	$15 \times 15 \times 512$
Conv2D Depth-wise / 2	$15 \times 15 \times 512$	$7 \times 7 \times 512$
Conv2D Point-wise / 1	$7 \times 7 \times 512$	$7 \times 7 \times 1024$
Conv2D Depth-wise / 1	$7 \times 7 \times 1024$	$7 \times 7 \times 1024$
Conv2D Point-wise / 1	$7 \times 7 \times 1024$	$7 \times 7 \times 1024$
Dense / 1	50176	512
Sigmoid	512	1

where it can be seen that the computations are decreased nearly by a factor of N as compared to that of standard convolutions presented in (1).

The proposed face PAD method uses the MobileNet pre-trained model as a base model. The proposed face PAD model architecture is presented in Table. 1. The top layers of the MobileNet base model are interchanged by a simple MLP classifier network containing two fully-connected layers with 512 and 1 units respectively. All layers are followed by batch normalization and activated by ReLU activation, except for the final fully-connected layer where a Sigmoid activation is used for binary classification. We used Adam optimizer with a learning rate of 1×10^{-4} for model compilation.

III. MATERIALS AND METHODS

A. DATASETS

The proposed deep learning framework was extensively evaluated on four standard datasets namely: (1) Replay-Attack, (2) Replay-Mobile, (3) CASIA-FASD, and (4) ROSE-Youtu. Each dataset is described below.

1) REPLAY-ATTACK DATASET

The Replay-Attack dataset [45] is a 2D face presentation attack dataset consisting of 1300 video clips (9-15 seconds duration) of photo and video attack attempts for 50 clients. The video clips are shot using different cameras and under different controlled and adverse lighting conditions, an example of real and attack video frames under adverse and

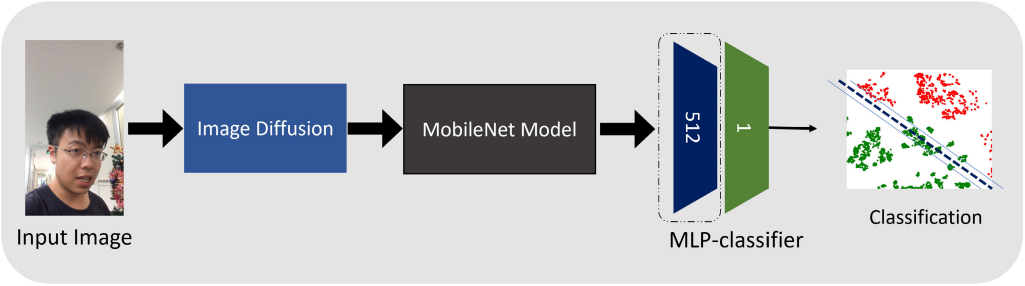

FIGURE 2. Proposed face PAD network.

controlled lighting is presented in Fig. 3. The first, second, and third columns represent the video frames for a real user, attack using fixed stand, and attack using hands to hold the spoofing device respectively. The first row represents the video frames in adverse lighting and the second row represents the video frames in controlled lighting conditions. All the videos are recorded at a frame rate of 25FPS, and have a resolution of 320 × 240. The dataset is divided into training, testing, and development set. The dataset contains 4 real and 20 attack videos per client/subject. The training and development set each contain 15 subjects with 360 video clips while the test set consists of 20 subjects with 480 video clips. The subjects present in one set do not appear in any other sets. The dataset details are enlisted in Table. 2.

2) REPLAY-MOBILE DATASET

The Replay-Mobile dataset [46] was developed for face recognition and face PAD in 2016. It contains 1030 video clips of photo and video attacks of 40 subjects. These video clips were recorded on different mobile devices under different lighting conditions. The dataset is divided into train, test, and development sets and the subjects present in one set do not appear in any other sets. The details of Replay-Mobile dataset are enlisted in Table. 2. Some examples of the extracted frames from the Replay-Mobile dataset are presented in Fig. 4.

3) CASIA-FASD DATASET

The CASIA-FASD dataset [40] developed for face anti-spoofing was released in 2012. CASIA-FASD is a small dataset containing diverse photo and video attacks of different image qualities for 50 subjects. Each subject in the dataset contains 3 genuine video clips and 9 attack video clips. Hence the dataset contains 600 video clips for 50 subjects. The dataset is divided into 20 subjects for training and 30 subjects for testing, whereas, each subject appears only in one of the sets. Some examples of the extracted frames from the CASIA-FASD dataset are presented in Fig. 5.

4) ROSE-YOUTU DATASET

ROSE-Youtu face liveness detection dataset [47], [48] is a comprehensive face anti-spoofing database. It covers a variety of lighting conditions, attack types, and camera

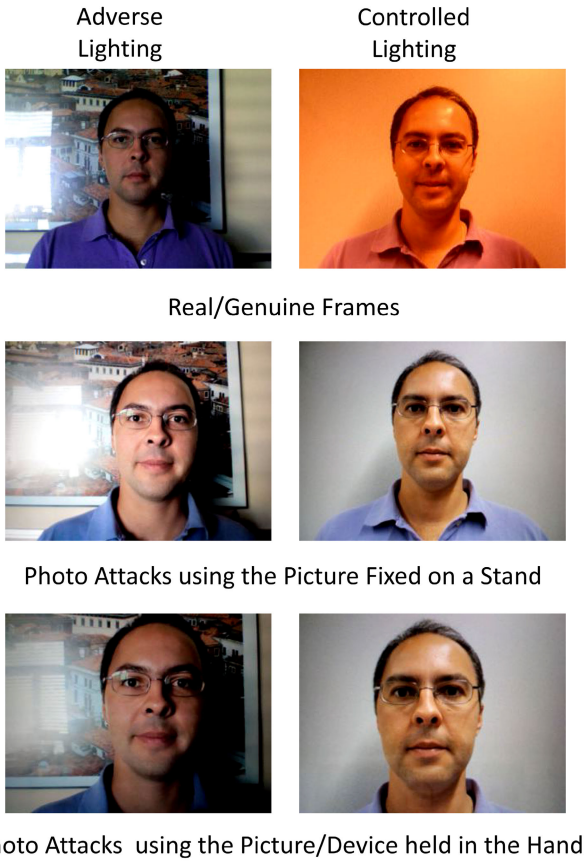

FIGURE 3. Example of real and attack video frames under adverse (first row) and controlled (second row) lighting in replay-attack dataset.

models. The database consists of 3350 video clips of 20 subjects. For each subject, there are around 150 ∼ 200 video clips with an average duration of 10 ∼ 12 seconds. There are three type of spoofing attacks covered in this database including video replay attack, print photo attack, and masking (paper mask) attack. The performance on ROSE-Youtu database is usually measured in equal error rate (EER). Some examples of the extracted frames from the ROSE-Youtu dataset are presented in Fig. 6.

B. PERFORMANCE EVALUATION METRICS

Since face PAD is essentially a classification problem, therefore, standard threshold dependent performance evalu-

FIGURE 4. Sample frames from real and attack video clips from replay-mobile database.

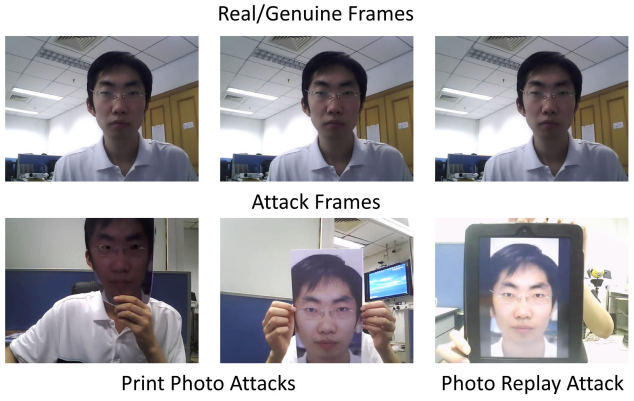

FIGURE 5. Sample frames from real and attack video clips from CASIA-FASD database.

ation parameters such as sensitivity (Sen), specificity (Spe), Youden's index (YI), and F1-score are reported in this paper. Besides these metrics, face liveness classifiers or commonly called face PAD methods can also be evaluated on the basis of the classification accuracy achieved by the algorithm [31]. Each of these parameters can be calculated using the following equations:

$$Sen = \frac{TP}{TP + FN} \tag{4}$$

$$Spe = \frac{TN}{TN + FP} \tag{5}$$

$$YI = Sensitivity + Specificity - 1 \tag{6}$$

$$F1 - Score = 2 \times \frac{Prec \times Recall}{Prec + Recall} \tag{7}$$

$$Prec = \frac{TP}{TP + FP} \tag{8}$$

TABLE 2. Dataset description.

Dataset	Year	Number of Subjects	Real / Attack
Replay-Attack [45]	2012	50	200 / 1000
Replay-Mobile [46]	2016	40	390 / 640
CASIA-FASD [40]	2012	50	150 / 450
ROSE-Youtu [47]	2018	20	1000 / 2350

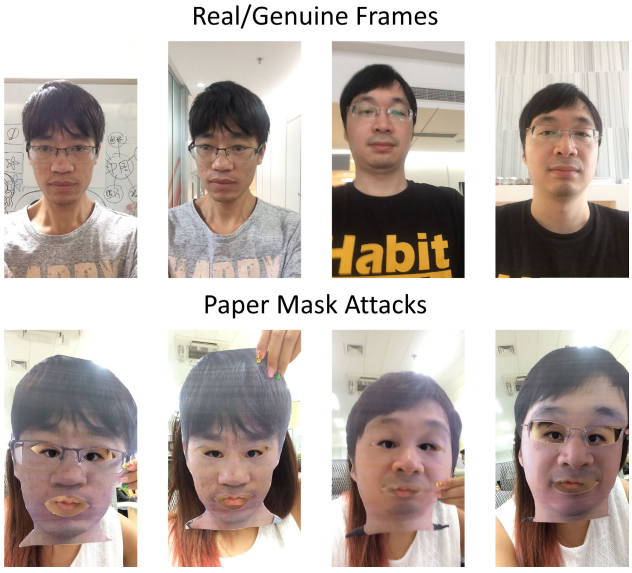

FIGURE 6. Sample frames from real and attack (Masking attack) video clips from ROSE-Youtu dataset.

where TP, FP, TN, and FN represent true positive, false positive, true negative, and false negative, respectively. Thus, sensitivity represents the fraction of real/genuine face images correctly detected/classified as genuine faces while specificity presents the fraction of spoof/attack faces correctly classified as attack faces. The Youden's index integrates the sensitivity and specificity measures in a way that emphasize both the sensitivity and specificity. The value of Youden's index ranges between 0 and 1, with 0 being the worst result and 1 being the perfect value indicating no false positives and false negatives. F1 score is also a popular measure of test accuracy and it represents the harmonic mean of the precision (Prec) and recall/sensitivity.

Besides these measures, most face PAD literature also reported the half total error rate (HTER), which is defined by the following equation:

$$HTER = \frac{FAR + FRR}{2} \tag{9}$$

where FAR and FRR are the false acceptance rate and false rejection rate defined by the following equations respectively:

$$FAR = \frac{FP}{FP + TN} \tag{10}$$

$$FRR = \frac{FN}{FN + TP} \tag{11}$$

HTER is used as a performance metric in most face anti-spoofing literature due to the reason that in most face PAD databases, there is a significant imbalance between the real and attack classes. This imbalance is mainly due to the difficulties in obtaining real data as compared to the imposter/attack data. Since FAR and FRR are database distribution independent metrics and the HTER is the average of the both, therefore, it is convenient to use HTER for performance comparison than the standard classification error metrics for face PAD methods.

For the performance evaluation on ROSE-Youtu and CASIA-FASD database, it is recommended to use equal error rate (EER) instead of HTER. EER can be defined by the following equation. Both the HTER and EER are reported as a percentage in literature.

$$EER = \frac{FP + FN}{TP + FP + FN + TN} \tag{12}$$

Among the various performance evaluation metrics, one of the most important metric to evaluate the classification performance of a binary classification system or any classifier in general is the receiver operating characteristics (ROC) curve. This is often called area under the curve - receiver operating characteristics (AUC-ROC) or simply area under the receiver operating characteristics (AUROC) curve. ROC curve is a probability curve and AUC represents the degree or extent of class separability. Therefore, the higher the AUC, the better the classifier is at discriminating class 0 from class 1. The ROC curve is a graph of sensitivity or true positive rate (TPR) against false positive rate (FPR) (where $FPR = 1 - specificity$).

C. TRAINING SETUP

Since face anti-spoofing datasets, in general, are imbalanced, i.e. the number of real video clips are less than the number of attack videos, therefore, for Replay-Attack and Replay-Mobile datasets, we randomly selected 25 frames of each video in the real training, development, and testing sets. Whereas for the attack videos, we randomly selected 10 frames from each video in the training, development, and testing sets. To keep the number of samples/images in both the classes balanced, we captured 40 frames from the real video clips in CASIA-FASD database and 15 frames from the respective attack video clips. Similarly, for ROSE-Youtu database 6 random frames were selected from each real video clip, while 2 frames were randomly selected from the attack video clips to maintain a balance between the two classes. The frames were captured with the dimensions of 30×40 for all the datasets used in this study except CASIA-FASD, where the captured frame dimensions were kept 224×224.

The proposed model was trained for 50 epochs and we used early stopping with the patience of 10 epochs to avoid over-fitting, thereby stopping the model training if the performance stopped improving. We used validation loss as the metric for early stopping, and the best model was saved to perform testing/inference on test set. The code for training and testing of the proposed method on CASIA-FASD database has been made available in the authors' github repository (https://github.com/mhdshl/Face-PAD-transfer-learning-diffusion).

TABLE 3. Performance of the proposed method evaluated on standard threshold-dependent metrics for Replay-Attack, Replay-Mobile, CASIA-FASD, and ROSE-Youtu datasets.

Dataset	Replay-Attack	Replay-Mobile	CASIA-FASD	ROSE-Youtu
Accuracy (%)	99.93	99.04	99.90	95.04
Sensitivity	0.9980	1.0	0.9902	0.9542
Specificity	1.0	0.9771	0.9709	0.9473
Youden Index	0.9980	0.9771	0.9612	0.9015
F1-Score	0.9990	0.9917	0.9738	0.9461

IV. EXPERIMENTAL RESULTS

Extensive experiments were conducted to evaluate the efficacy of the proposed framework. For the proposed approach, five independent simulation runs were performed to calculate the standard threshold-dependent metrics and the results for accuracy, sensitivity, specificity, Youden's index, F1-Score, etc., for all the datasets are reported in Table. 3. The results clearly show the superior discriminating capability of the proposed approach in terms of these standard threshold-dependent metrics.

To verify the robustness of the proposed approach, data visualization was performed for all four databases using PCA embedding, refer to Fig. 7. The PCA plots of the diffused input images and the plots of latent feature space of the proposed face PAD network are presented. The latent features are extracted from the MLP network hidden layer with 512 nodes. The data visualization exercise not only verifies the effectiveness of the proposed image diffusion scheme, but also shows the discriminating capability of the proposed face PAD model. The latent feature space PCA embeddings presented in Fig. 7 show the strong discriminating capability of the proposed face PAD technique.

We also calculated AUC and plotted the AUC-ROC curves for the performance evaluation of the proposed technique on each dataset. The AUC-ROC curves for Replay-Attack, Replay-Mobile, ROSE-Youtu, and CASIA-FASD datasets are presented in Fig. 8a, 8b, 8c, and 8d respectively. The AUC-ROC curves and the AUC values presented in Fig. 8 showcase the promising face PAD performance of the proposed approach with an AUC of 0.9995, 0.9918, 0.9496, and 0.9989 units for Replay-Attack, Replay-Mobile, ROSE-Youtu, and CASIA-FASD databases respectively.

A. COMPARISON OF DIFFERENT PRE-PROCESSING SCHEMES

In order to evaluate the effectiveness of the proposed diffusion technique and its utility in model training, an experiment is designed to evaluate the performance of the proposed face PAD model on ROSE-Youtu database with different pre-processing schemes including: (1) model trained without any diffusion in pre-processing module, (2) model trained

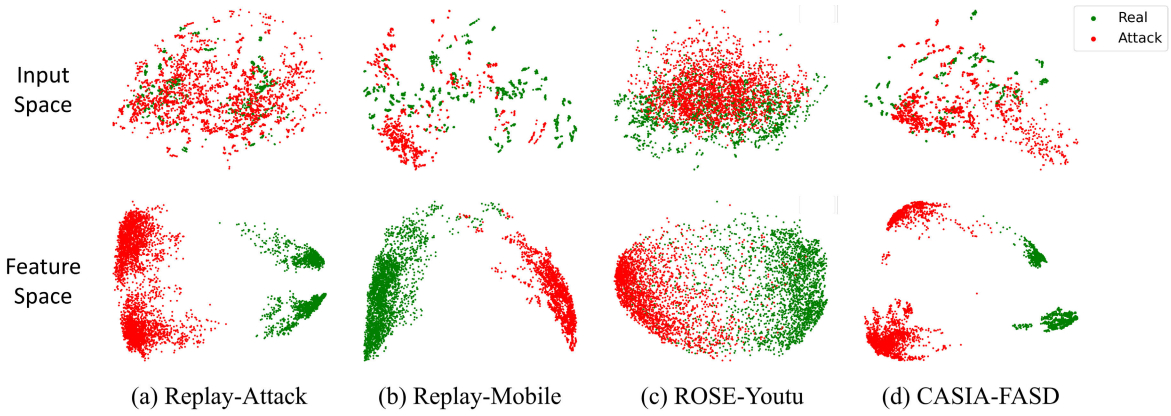

FIGURE 7. PCA embeddings of input space vs. latent feature space, (a) Replay-Attack database, (b) Replay-Mobile database, (c) ROSE-Youtu database, (d) CASIA-FASD database.

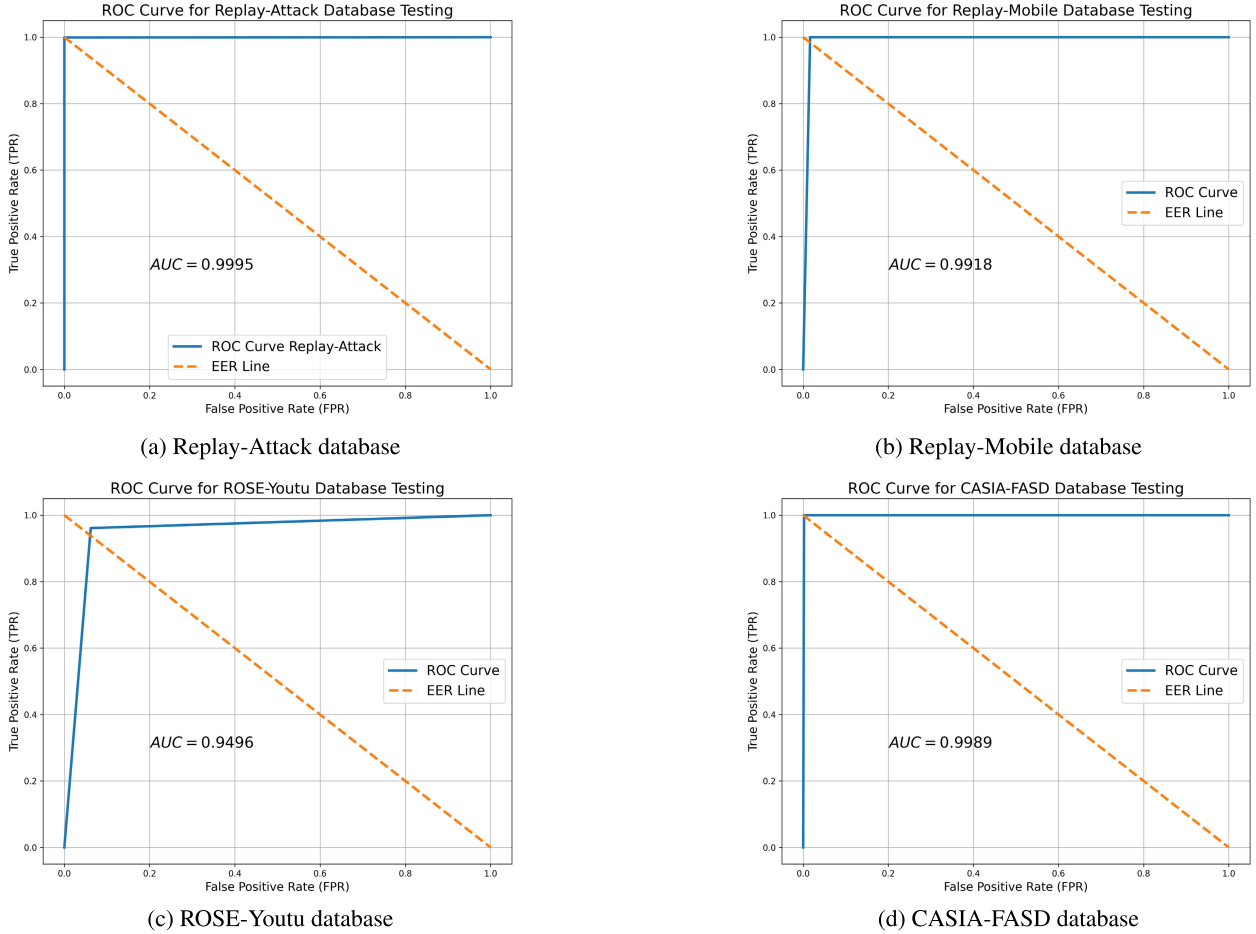

FIGURE 8. Receiver operating characteristics (ROC) curves. (a) AUC-ROC curve for training and testing on Replay-Attack database ((AUC = 0.9995, EER(%) = 0.06). (b) AUC-ROC curve for training and testing on Replay-Mobile database (AUC = 0.9918, EER(%) = 0.96). (c) AUC-ROC curve for training and testing on ROSE-Youtu database (AUC = 0.9496, EER(%) = 4.95). (d) AUC-ROC curve for training and testing on CASIA-FASD database (AUC = 0.9989, EER(%) = 0.09).

with Perona-Malik anisotropic diffusion in pre-processing module, (3) model training done with Gaussain filtering-based diffusion in pre-processing module, and (4) model trained with the proposed diffusion scheme in pre-processing module. The training setup for these experiments is kept the same as outlined in the previous section.

For Perona-Malik anisotropic diffusion, the parameters such as number of iterations, conduction coefficient, stability constant, step size (distance between the adjacent pixels), etc., are kept the same as the default values. Similarly, for diffusion using Gaussian filtering, the standard deviation for Gaussian filter is also kept as the default value. Table 4 presents the

TABLE 4. Comparison of the performance of the face PAD model with different pre-processing schemes on ROSE-Youtu database.

Pre-Processing Scheme	Accuracy (%)	HTER / EER (%)
No Diffusion	90.99	8.56 / 9.01
Perona-Malik Diffusion [34], [35]	79.26	17.03 / 20.74
Gaussian Filtering [36]	90.52	8.9 / 9.48
Proposed	95.04	4.92 / 4.95

TABLE 5. Performance comparison for different learning schemes on ROSE-Youtu database.

Performance Metric	Scheme-1	Scheme-2	Scheme-3
Accuracy (%)	89.35	91.91	95.04
Sensitivity	0.8692	0.9392	0.9542
Specificity	0.9216	0.8958	0.9473
Youden Index	0.7908	0.8350	0.9015
F1-Score	0.8975	0.9257	0.9461
HTER (%)	10.46	8.25	4.92

TABLE 6. Comparative test results for replay-attack dataset.

Method	HTER (%)
Diffusion Speed [49] (2015)	12.5
FASNet [50] (2017)	1.20
Diffusion-CNN [51] (2017)	10.0
SfSNet [23] (2020)	3.1
InceptionV4 [31] (2020)	13.54
SCNN [31] (2020)	7.53
SE-ResNet-18 [52] (2020)	3.3
CompactNet [53] (2020)	0.7
VGG16+GMM [54] (2021)	1.46
GoogleNet+GMM [54] (2021)	3.76
WA (PSO+PS) [55] (2021)	0.0
WA (GA+MMS+PS) [56] (2022)	0.0
Proposed Method	0.09

TABLE 7. Comparative test results for replay-mobile dataset.

Method	HTER (%)
SR Arashloo et. al. [57] (2020)	6.7
InceptionV4 [31] (2020)	5.94
SCNN [31] (2020)	4.96
VGG16+GMM [54] (2021)	17.21
MKL [57] (2021)	6.7
GoogleNet+GMM [54] (2021)	13.56
WA (PSO+PS) [55] (2021)	5.85
WA (GA+MMS+PS) [56] (2022)	5.12
Proposed Method	1.14

accuracy and HTER/EER scores of the models trained with each of the pre-processing schemes detailed above.

TABLE 8. Comparative test results for ROSE-Youtu dataset.

Method	HTER/EER (%)
LPQ (HSV) [16] (2016)	30.4/-
LPQ (YCbCr) [16] (2016)	27.6/-
Wavelet [58] (2017)	26.6/-
Ensemble of Classifiers [59] (2019)	9.3/-
SE-ResNet-18 [52] (2020)	-/8.0
WA (PSO+PS) [55] (2021)	5.61/-
FASNet [60] (2021)	8.57/-
ResNet50+GMM [54] (2021)	14.69/-
WA (GA+MMS+PS) [56] (2022)	5.12/-
Fatemifar et al. [61] (2022)	6.34/-
Proposed Method	4.92/4.95

TABLE 9. Comparative test results for CASIA-FASD dataset.

Method	HTER/EER (%)
Deep Metric Learning [62] (2019)	16.74/-
Motion Pattern [63] (2020)	17.81/-
Noise Pattern [63] (2020)	13.33/-
Decision Fusion [63] (2020)	10.54/-
GFA-CNN [64] (2020)	-/ 8.3
S-CNN [65] (2021)	-/ 0.69
Proposed Method	0.09/0.09

The results presented in Table 4 clearly show that the proposed pre-processing scheme achieves 42.5%, 71.1%, and 44.7% improvement in HTER score compared to no diffusion, Perona-Malik diffusion, and Gaussian filtering-based diffusion respectively.

B. COMPARISON OF DIFFERENT LEARNING SCHEMES

In this section, an experiment to evaluate the performance of different learning schemes is conducted. Similar to the previous experiment, ROSE-Youtu database is used in this experiment to compare the performance of the two transfer learning schemes discussed in Section II as well as a learning scheme where the MobileNet architetcure is trained from scratch. The network architecture and the training setup is kept the same for all the learning schemes. For simplicity, the learning schemes have been denoted as: Scheme-1: where the MobileNet architecture is loaded without any weights and trained from scratch, Scheme-2: where pre-trained ImageNet weights are loaded and the layers of the MobileNet model are frozen (i.e. the MobileNet model is used in feature extraction setting), Scheme-3: pre-trained ImageNet weights are loaded and the model is trained in a fine-tuning setting. It is noteworthy that in Scheme-2, only the fully-connected layers are trained while in Scheme-3, the layers of MobileNet as well as the fully-connected layers are fine-tuned. The performance has been evaluated using the

TABLE 10. Cross-database performance (in-terms of HTER(%) of the proposed approach and the baseline approaches). Train dataset → Test dataset (Replay-Attack: RA, ROSE-Youtu: RY, CASIA-FASD: CF).

Method	RA → RY	RA → CF	RY → RA	RY → CF	CF → RA	CF → RY	Mean HTER (%)
Tzeng et al. [66] (2017)	50.0	49.8	34.6	28.7	41.8	31.4	39.38
Li et al. [47] (2018)	40.1	12.3	38.8	30.1	39.3	31.6	32.03
Wang et al. [67] (2019)	41.7	41.5	30.3	34.1	17.5	29.4	32.42
Proposed Method	33.27	23.09	8.43	29.78	28.89	32.03	25.91

performance evaluation metrics presented in Section III and the results are presented in Table 5.

The comparative results presented in Table 5 clearly indicate that the learning Scheme-3 (Fine-tuning) shows superior results compared to the learning Scheme-1 and Scheme-2 in-terms of almost every performance evaluation metric. In particular, the HTER results show 52.96% and 40.36% improvement in Scheme-3 as compared to Scheme-1 and Scheme-2 respectively. Therefore, transfer learning in a fine-tuning setting is selected for the proposed scheme.

C. COMPARISON WITH STATE-OF-THE-ART FACE PAD METHODS

The performance of our proposed face PAD approaches using our interpolation-based image diffusion method was compared with the state-of-the-art methods on Replay-Attack and Replay-Mobile dataset.

We compared our method with different CNN and diffusion-based face PAD methods and the comparative results are presented in Table 6 and Table 7 for Replay-Attack and Replay-Mobile dataset, respectively. It is evident from the results that the proposed approach shows superior performance compared to most of the competing approaches. For the Replay-Attack database, for instance, the proposed approach outperforms most of the contemporary methods by a high margin and achieves comparable HTER performance with the methods presented in [55] and [56]. Similarly, as presented in Table 7, the proposed method outperforms the contemporary approaches on Replay-Mobile database.

The results for ROSE-Youtu and CASIA-FASD datasets are presented in Tables 8 and 9 respectively. The proposed method achieves an HTER/EER (%) score of 4.92/4.95 for ROSE-Youtu database which outperforms the contemporary methods presented in the literature. Similarly for CASIA-FASD, the proposed method achieves an HTER value of 0.09% outperforming the other methods presented in the results.

D. CROSS-DOMAIN PERFORMANCE EVALUATION

Extensive experiments were conducted to perform cross-domain performance evaluation of the proposed face PAD technique. A number of experiments were performed where the proposed face PAD model was trained on Replay-Attack training dataset and tested on ROSE-Youtu, and CASIA-FASD datasets. In a similar setting, we trained the proposed model on individual datasets and performed testing on the others and report HTER(%) values. Table 10 presents the cross-database testing results in-terms of HTER. The results presented in the table exhibit the robustness and versatility of the proposed technique in a cross-database testing scheme. The overall HTER results show that the proposed method achieves better mean HTER compared to the face PAD techniques presented in [47], [66], and [67] by 34.25%, 19.11%, and 20.08% respectively.

For training on Replay-Attack and testing on CASIA-FASD, the proposed method outperforms [66] and [67] by an HTER margin of 53.63% and 44.36%. The cross-database performance is found inferior to [47]. For training on Replay-Attack and testing on ROSE-Youtu, the proposed approach shows superior performance in terms of cross-database HTER and achieves 33.46%, 17.03%, and 20.21% gain as compared to [47], [66], and [67] respectively. For training on ROSE-Youtu and testing on Replay-Attack, the proposed method outperforms [47], [66], and [67] by an HTER margin of 75.63%, 78.27%, and 72.18% respectively. For testing on CASIA-FASD, the proposed technique shows improvement in HTER margin by 1.06% and 12.67% compared to the techniques presented in [47] and [67]. Lastly, for the case of training on CASIA-FASD and testing on the Replay-Attack database, the proposed method outperforms [47] and [66] by an HTER margin of 30.88% and 26.49% respectively. The cross-database testing of a model trained on CASIA-FASD and tested on ROSE-Youtu also shows comparable performance.

V. CONCLUSION

In this paper, a hybrid face PAD approach is proposed which incorporates the notion of interpolation-based image diffusion with the transfer learning of a MobileNET CNN. The proposed framework has shown promising results on Replay-Attack, Replay-Mobile, CASIA-FASD, and ROSE-Youtu databases attaining the highest accuracy and HTER of 99.93% and 0.09%, 99.04% and 1.14%, 99.90% and 0.09%, and 95.04% and 4.92%, respectively. The proposed method also demonstrated superior performance in cross-domain evaluation as well. The applications of such face PAD approaches are vast and for our future prospects, we aim to combine our face PAD method with a face recognition and gesture recognition system for student attendance and

examination monitoring in an educational setting to provide a combined deep learning-based framework for the day-to-day activities carried out in schools. We also aim to improve the cross-domain performance of the proposed method in our future works by leveraging the proposed method in an unsupervised learning scheme to perform domain adaptation for face PAD across various complex face PAD databases.

REFERENCES

[1] P. Rasti, T. Uiboupin, S. Escalera, and G. Anbarjafari, "Convolutional neural network super resolution for face recognition in surveillance monitoring," in *Proc. Int. Conf. Articulated Motion Deformable Objects*. Cham, Switzerland: Springer, 2016, pp. 175–184.

[2] S. Lukas, A. R. Mitra, R. I. Desanti, and D. Krisnadi, "Student attendance system in classroom using face recognition technique," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2016, pp. 1032–1035.

[3] A. Fayyoumi and A. Zarrad, "Novel solution based on face recognition to address identity theft and cheating in online examination systems," *Adv. Internet Things*, vol. 4, no. 2, pp. 5–12, 2014.

[4] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 10, pp. 2268–2283, Oct. 2016.

[5] H. Lee, S.-H. Park, J.-H. Yoo, S.-H. Jung, and J.-H. Huh, "Face recognition at a distance for a stand-alone access control system," *Sensors*, vol. 20, no. 3, p. 785, Jan. 2020.

[6] P. J. Grother, M. L. Ngan, and K. K. Hanaoka, "Ongoing face recognition vendor test (FRVT) part 2: Identification," Nat. Inst. Standards Technol. (NIST), USA, Tech. Rep. 8238, 2018.

[7] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in 'liveness' assessment," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 548–558, Aug. 2007.

[8] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *Proc. Int. Conf. Image Anal. Signal Process.*, 2009, pp. 233–236.

[9] A. Anjos, M. M. Chakka, and S. Marcel, "Motion-based counter-measures to photo attacks in face recognition," *IET Biometrics*, vol. 3, no. 3, pp. 147–158, Sep. 2014.

[10] G. Pan, Z. Wu, and L. Sun, "Liveness detection for face recognition," in *Recent Advances in Face Recognition*. Rijeka, Croatia: InTech, 2008, pp. 109–124.

[11] K. Patel, H. Han, and A. K. Jain, "Cross-database face antispoofing with robust feature representation," in *Proc. Chin. Conf. Biometric Recognit.* Cham, Switzerland: Springer, 2016, pp. 611–619.

[12] R. Shao, X. Lan, and P. C. Yuen, "Deep convolutional dynamic texture learning with adaptive channel-discriminability for 3D mask face anti-spoofing," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 748–755.

[13] K. Kollreider, H. Fronthaler, and J. Bigun, "Evaluating liveness by face images and the structure tensor," in *Proc. 4th IEEE Workshop Autom. Identificat. Adv. Technol. (AutoID)*, 2005, pp. 75–80.

[14] T. D. F. Pereira, A. Anjos, J. M. D. Martino, and S. Marcel, "*LBP—TOP* based countermeasure against face spoofing attacks," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2012, pp. 121–132.

[15] T. D. F. Pereira, A. Anjos, J. M. D. Martino, and S. Marcel, "Can face anti-spoofing countermeasures work in a real world scenario?" in *Proc. Int. Conf. Biometrics (ICB)*, 2013, pp. 1–8.

[16] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1818–1830, Aug. 2016.

[17] J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection with component dependent descriptor," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–6.

[18] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, "An investigation of local descriptors for biometric spoofing detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 849–863, Apr. 2015.

[19] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face antispoofing using speeded-up robust features and Fisher vector encoding," *IEEE Signal Process. Lett.*, vol. 24, no. 2, pp. 141–145, Feb. 2017.

[20] L. Li, Z. Xia, J. Wu, L. Yang, and H. Han, "Face presentation attack detection based on optical flow and texture analysis," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1455–1467, Apr. 2022.

[21] L. Li, Z. Xia, X. Jiang, Y. Ma, F. Roli, and X. Feng, "3D face mask presentation attack detection based on intrinsic image analysis," *IET Biometrics*, vol. 9, no. 3, pp. 100–108, May 2020.

[22] W. Sun, Y. Song, C. Chen, J. Huang, and A. C. Kot, "Face spoofing detection based on local ternary label supervision in fully convolutional networks," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3181–3196, 2020.

[23] A. Pinto, S. Goldenstein, A. Ferreira, T. Carvalho, H. Pedrini, and A. Rocha, "Leveraging shape, reflectance and albedo from shading for face presentation attack detection," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3347–3358, 2020.

[24] Y. A. U. Rehman, L.-M. Po, M. Liu, Z. Zou, W. Ou, and Y. Zhao, "Face liveness detection using convolutional-features fusion of real and deep network generated face images," *J. Vis. Commun. Image Represent.*, vol. 59, pp. 574–582, Feb. 2019.

[25] Y. A. U. Rehman, L.-M. Po, and M. Liu, "SLNet: Stereo face liveness detection via dynamic disparity-maps and convolutional neural network," *Exp. Syst. Appl.*, vol. 142, Mar. 2020, Art. no. 113002.

[26] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel, "Biometric face presentation attack detection with multi-channel convolutional neural network," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 42–55, 2020.

[27] H. Chen, Y. Chen, X. Tian, and R. Jiang, "A Cascade face spoofing detector based on face anti-spoofing R-CNN and improved Retinex LBP," *IEEE Access*, vol. 7, pp. 170116–170133, 2019.

[28] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[29] H. Chen, G. Hu, Z. Lei, Y. Chen, N. M. Robertson, and S. Z. Li, "Attention-based two-stream convolutional networks for face spoofing detection," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 578–593, 2020.

[30] G. Heusch, A. George, D. Geissbühler, Z. Mostaani, and S. Marcel, "Deep models and shortwave infrared information to detect face presentation attacks," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 4, pp. 399–409, Oct. 2020.

[31] R. Koshy and A. Mahmood, "Enhanced deep learning architectures for face liveness detection for static and video sequences," *Entropy*, vol. 22, no. 10, p. 1186, Oct. 2020.

[32] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[33] E. Erdem, *Linear Diffusion*. Ankara, Turkey: Hacettepe Univ., Feb. 2012.

[34] P. Perona, T. Shiota, and J. Malik, "Anisotropic diffusion," in *Geometry-Driven Diffusion in Computer Vision*. Cham, Switzerland: Springer, 1994, pp. 73–92.

[35] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data," *IEEE Trans. Med. Imag.*, vol. 11, no. 2, pp. 221–232, Jun. 1992.

[36] J. Weickert, *Anisotropic Diffusion in Image Processing*, vol. 1. Stuttgart, Germany: Teubner Stuttgart, 1998.

[37] D. Ziou and A. Horé, "Reducing aliasing in images: A PDE-based diffusion revisited," *Pattern Recognit.*, vol. 45, no. 3, pp. 1180–1194, Mar. 2012.

[38] Y. Q. Wang, J. Guo, W. Chen, and W. Zhang, "Image denoising using modified Perona–Malik model based on directional Laplacian," *Signal Process.*, vol. 93, no. 9, pp. 2548–2558, Sep. 2013.

[39] N. Wang, Y. Shang, Y. Chen, M. Yang, Q. Zhang, Y. Liu, and Z. Gui, "A hybrid model for image denoising combining modified isotropic diffusion model and modified Perona–Malik model," *IEEE Access*, vol. 6, pp. 33568–33582, 2018.

[40] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar. 2012, pp. 26–31.

[41] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" 2014, *arXiv:1411.1792*.

[42] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 512–519.

[43] E. Cetinic, T. Lipic, and S. Grgic, "Fine-tuning convolutional neural networks for fine art classification," *Exp. Syst. Appl.*, vol. 114, pp. 107–118, Dec. 2018.

[44] L. Sifre, "Rigid-motion scattering for image classification," Ph.D. thesis, Ecole Polytechnique, CMAP, Lausanne, Switzerland, 2014.
[45] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2012, pp. 1–7.
[46] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, "The replay-mobile face presentation-attack database," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2016, pp. 1–7.
[47] H. Li, W. Li, H. Cao, S. Wang, F. Huang, and A. C. Kot, "Unsupervised domain adaptation for face anti-spoofing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 7, pp. 1794–1809, Jul. 2018.
[48] Z. Li, R. Cai, H. Li, K. Lam, Y. Hu, and A. C. Kot, "One-class knowledge distillation for face presentation attack detection," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2137–2150, 2022.
[49] W. Kim, S. Suh, and J. Han, "Face liveness detection from a single image via diffusion speed model," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2456–2465, Aug. 2015.
[50] O. Lucena, A. Junior, V. Moia, R. Souza, E. Valle, and R. Lotufo, "Transfer learning using convolutional neural networks for face anti-spoofing," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, 2017, pp. 27–34.
[51] A. Alotaibi and A. Mahmood, "Deep face liveness detection based on nonlinear diffusion using convolution neural network," *Signal, Image Video Process.*, vol. 11, no. 4, pp. 713–720, May 2017.
[52] G. Wang, H. Han, S. Shan, and X. Chen, "Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 56–69, 2021.
[53] L. Li, Z. Xia, X. Jiang, F. Roli, and X. Feng, "CompactNet: Learning a compact space for face presentation attack detection," *Neurocomputing*, vol. 409, pp. 191–207, Oct. 2020.
[54] S. Fatemifar, S. R. Arashloo, M. Awais, and J. Kittler, "Client-specific anomaly detection for face presentation attack detection," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107696.
[55] S. Fatemifar, M. Awais, A. Akbari, and J. Kittler, "Particle swarm and pattern search optimisation of an ensemble of face anomaly detectors," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 3622–3626.
[56] S. Fatemifar, S. Asadi, M. Awais, A. Akbari, and J. Kittler, "Face spoofing detection ensemble via multistage optimisation and pruning," *Pattern Recognit. Lett.*, vol. 158, pp. 1–8, Jun. 2022.
[57] S. R. Arashloo, "Matrix-regularized one-class multiple kernel learning for unseen face presentation attack detection," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 4635–4647, 2021.
[58] W. Li, L. Chen, D. Xu, and L. Van Gool, "Visual recognition in RGB images and videos by learning from RGB-D data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 2030–2036, Aug. 2018.
[59] S. Fatemifar, M. Awais, S. R. Arashloo, and J. Kittler, "Combining multiple one-class classifiers for anomaly based face spoofing attack detection," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–7.
[60] N. Bousnina, L. Zheng, M. Mikram, S. Ghouzali, and K. Minaoui, "Unraveling robustness of deep face anti-spoofing models against pixel attacks," *Multimedia Tools Appl.*, vol. 80, no. 5, pp. 7229–7246, Feb. 2021.
[61] S. Fatemifar, M. Awais, A. Akbari, and J. Kittler, "Developing a generic framework for anomaly detection," *Pattern Recognit.*, vol. 124, Apr. 2022, Art. no. 108500.
[62] D. Pérez-Cabo, D. Jiménez-Cabello, A. Costa-Pazo, and R. J. López-Sastre, "Deep anomaly detection for generalized face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1591–1600.
[63] Y. Ma, Y. Xu, and F. Liu, "Multi-perspective dynamic features for cross-database face presentation attack detection," *IEEE Access*, vol. 8, pp. 26505–26516, 2020.
[64] X. Tu, Z. Ma, J. Zhao, G. Du, M. Xie, and J. Feng, "Learning generalizable and identity-discriminative representations for face anti-spoofing," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 5, pp. 1–19, Oct. 2020.
[65] R. Quan, Y. Wu, X. Yu, and Y. Yang, "Progressive transfer learning for face anti-spoofing," *IEEE Trans. Image Process.*, vol. 30, pp. 3946–3955, 2021.
[66] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2962–2971.
[67] G. Wang, H. Han, S. Shan, and X. Chen, "Improving cross-database face presentation attack detection via adversarial domain adaptation," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.

MADINI O. ALASSAFI received the B.S. degree in computer science from King Abdulaziz University, Jeddah, Saudi Arabia, in 2006, the M.S. degree in computer science from California Lutheran University, USA, in 2013, and the Ph.D. degree in security cloud computing from the University of Southampton, U.K., in February 2018. He is currently an Associate Professor with the Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, where he is also the Vice-Dean of the Faculty of Computing and Information Technology. His research interests include cloud computing and security, distributed systems, the Internet of Things (IoT) security issues, cloud security adoption, risks, cloud migration project management, and the cloud of things and security threats.

MUHAMMAD SOHAIL IBRAHIM received the B.E. degree in electronic engineering from Iqra University, Karachi, Pakistan, in 2012, and the M.E. degree in telecommunications from the N. E. D. University of Engineering and Technology, Pakistan, in 2016.

From 2013 to 2019, he was a Lecturer with the Faculty of Engineering, Science, and Technology, Iqra University. From 2013 to 2014, he was also a Research Assistant with the Embedded Systems Research Group, Karachi Institute of Economics and Technology, Pakistan. He is currently with the Smart Energy Systems Laboratory, College of Electrical Engineering, Zhejiang University, China. His research interests include deep learning, computer vision, and deep learning applications in energy systems. He was a recipient of 2020 Highly Cited Review Paper Award from *Applied Energy* (Elsevier) for his review paper titled "Machine Learning Driven Smart Electric Systems: Current Trends and New Perspectives."

IMRAN NASEEM received the B.E. degree in electrical engineering from the NED University of Engineering and Technology, Pakistan, in 2002, the M.S. degree in electrical engineering from the King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia, in 2005, and the Ph.D. degree from The University of Western Australia, in 2010. He did his post doctorate with the Institute for Multi-sensor Processing and Content Analysis, Curtin University of Technology, Australia. He joined the College of Engineering, KIET, Pakistan, in 2011 where he is currently a Professor. He is also an Adjunct Research Fellow with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia. His research interests include pattern classification and machine learning with a special emphasis on biometrics and bioinformatics applications. He has authored several publications in top journals and conferences, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and IEEE International Conference on Image Processing. His benchmark work on face recognition has received more than 180 citations in less than four years. He is also a reviewer of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING and IEEE SIGNAL PROCESSING LETTERS.

RAYED ALGHAMDI received the bachelor's degree in computer science, and the master's and Ph.D. degrees in communication and information technology. He is currently involved in designing a full interactive e-learning course to target enhancing soft skills for computing students. His current research interests include e-learning applications and computing student's readiness to confidently join industry manpower.

REEM ALOTAIBI received the Ph.D. degree in computer science from the University of Bristol, Bristol, U.K., in 2017. She is currently an Associate Professor with the Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. She is also the Supervisor of the Information Technology Department. From 2017 to 2018, she was a Visiting Lecturer with the Intelligent Systems Laboratory, University of Bristol. Her research interests include artificial intelligence, machine earning, data mining, and crowd management. Her research has been funded by several sources in Saudi Arabia, including Deputyship for Research and Innovation, Ministry of Education, King Abdulaziz City for Science and Technology (KACST), and the Deanship of Scientific Research (DSR), King Abdulaziz University.

FARIS A. KATEB received the Ph.D. degree in computer science from the University of Colorado, USA. Currently, he is an Assistant Professor and the Head of the IT Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include computer vision and image processing applications, such as object detection, face recognition, and adversarial examples. He is also working on the natural language process for Arabic and English. He participates as a speaker or a presenter in conferences and artificial intelligence areas and a member of the advisory board in other departments.

HADI MOHSEN OQAIBI received the B.S., M.S., and Ph.D. degrees in computer science from King Abdulaziz University, Saudi Arabia. He is currently an Assistant Professor with the Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia. He has published several peer reviewed journal articles and conference papers. His research interests include machine learning, deep learning, pattern recognition, and image processing.

ABDULRAHMAN A. ALSHDADI received the Ph.D. degree in cloud computing from the University of Southampton, Southampton, U.K., in February 2018. He is currently an Assistant Professor in computer science with the Faculty of Computing and Information Technology, University of Jeddah, Jeddah, Saudi Arabia. He is also the Head of the Computer Science and Artificial Intelligent Department (CSAI) and the Vice Dean of the College of Computer Science and Engineering (CCSE), University of Jeddah. He has published numerous conference papers, journal articles, and one book chapter. His research interests include industry 4.0 prestaining issues of cloud computing and fog computing security, the Internet of Things (IoT), smart cities, intelligent systems, deep learning, data science analytics, and modeling.

SYED ADNAN YUSUF is currently the Director of a U.K.-Based Research and Development Firm specializing in advanced computer vision algorithms with a focus on deep-learning technologies. His career originates from a computer systems engineering background with an interest in advanced biometrics, intelligent transport systems, long-range object detection, tracking, and identification. As a research scientist, he has lead various teams working in the domains of document verification, facial identity analysis, and traffic violation. In his previous role with Hitachi Europe, he led a team of scientists and engineers developing autonomous perception and motion planning systems for the Nissan Leaf Electric. The work led to a fully autonomous 200+-mile journey on a variety of U.K., roads as part of the Human Drive Project. In the research domain, his focus is on CNN/RNN algorithms with a focus on driverless autonomous control and video analytics systems and deep residual networks for the face recognition domain. Having a background in computer vision and deep learning domains, he has contributed in projects, including firefighter safety, maritime condition monitoring, financial technologies, and intelligent transport systems.