The Visual Computer
https://doi.org/10.1007/s00371-021-02123-4

ORIGINAL ARTICLE

A robust framework for spoofing detection in faces using deep learning

Shefali Arora[1] · **M. P. S. Bhatia**[1] · **Vipul Mittal**[1]

Accepted: 25 March 2021

Abstract
Face recognition is used in biometric systems to verify and authenticate an individual. However, most face authentication systems are prone to spoofing attacks such as replay attacks, attacks using 3D masks etc. Thus, the importance of face anti-spoofing algorithms is becoming essential in these systems. Recently, deep learning has emerged and achieved excellent results in challenging tasks related to computer vision. The proposed framework relies on the extraction of features from the faces of individuals. The approach relies on dimensionality reduction and feature extraction of input frames using pre-trained weights of convolutional autoencoders, followed by classification using softmax classifier. Experimental analysis on three benchmarks, Idiap Replay Attack, CASIA- FASD and 3DMAD, shows that the proposed framework can attain results comparable to state-of-the-art methods in both cross-database and intra-database testing.

Keywords Spoofing · Autoencoders · Deep learning · CNN · Security · Face · Biometric systems

1 Introduction

These days, various knowledge-based and ownership-based methods are being used to secure the identities of individuals. This would control the authentication of user data. While knowledge-based methods like passwords and PINs are commonly used, they are vulnerable to several attacks such as man-in-the-middle attack, Replay Attack, and stolen-verifier attacks. On the contrary, ownership-based methods make use of smart codes or tokens to authenticate a user. These means of authentication can also be stolen, misused and forged to get unauthorized access[1].

Face authentication in biometric systems is a popular area for research. There are various challenges like a variety of viewpoints, occlusions and age of the individual, along with factors like lighting conditions. Although there is a significant number of works addressing these issues, the problems of spoofing attacks are often ignored in biometric systems using face authentication. For instance, there are built-in webcams in desktops and laptops with operating systems like Windows XP and Vista. These have biometric systems embedded in them to authenticate users. However, the Security and Vulnerability Research Team of the University of Hanoi stated that it is easy to bypass these systems with the help of spoofed faces. Thus, fake facial images can be used in place of authentic images to get access to systems. Also, this threat has been listed now in the National Vulnerability Database of the National Institute of Standards and Technology (NIST) in the USA. There are various such examples which show the vulnerabilities of these systems using face recognition. Thus, there is an urgent need to address these attacks to enhance the security of these systems and bring them into practical use. A spoofing attack occurs when an intruder tries to impersonate another user and falsifies data, thus gaining unauthorized access to the system using the authentic user's credentials. For example, a user's identity can be forged by presenting a photograph, video or 3D mask of an authentic user to the sensor. One can also make use of make-up or undergo plastic surgery for spoofing an individual's identity. However, the most common attacks make use of photographs and 3D masks. This is because it is easy to download photographs of facial images or videos. The research in this domain is growing and lately

✉ Shefali Arora
arorashef@gmail.com

M. P. S. Bhatia
bhatia.mps@gmail.com

Vipul Mittal
vipulmittal@gmail.com

[1] Division of Computer Engineering, Netaji Subhas Institute of Technology, Delhi, India

Published online: 13 April 2021

many conferences are being organized to give a boost to the progress achieved in this field. The vulnerability of spoofing has received a lot of attention, and various countermeasures are being developed to deal with 2D and 3D spoofing attacks. Therefore, robust face authentication systems are required to distinguish between authentic faces and spoofed faces, thus ensuring the security of such systems. In case of replay attacks, an intruder intercepts the video of a user and misuses the signal for his benefit. The success of an anti-spoofing framework designed for a system is usually connected to the specific trait[2]. Such designs rely on the extraction of features which would be able to capture details related to a particular kind of attack. The need for custom-made solutions might impose certain constraints since any small changes in the attack could lead us to redesign the entire framework.

Based on various approaches, many algorithms have been designed for the detection of spoofing[3].Various databases have been made available in the public domain to validate the strengths of these frameworks[4]. Many competitions have been organized so as to give appropriate countermeasures to facial spoofing attacks, which would give a great boost in this domain.Face image quality, facial motions, and scenic motions provide different features for face anti-spoofing. Other techniques using hand-engineered features to detect spoofing in faces are based on the aspects of image quality as well as motion[5].

Researchers are using deep learning architectures to achieve state-of-the-art results in the domain of computer vision. Among various such architectures, Convolutional Neural Networks(CNN) have shown superior performance in dealing with complex computer vision problems and problems related to images. . In this context, we propose a robust framework which matches state-of-the-art performance in detection of such attacks.

In this paper, we leverage the use of deep learning along with dimensionality reduction to build a robust approach that deals with the detection of spoofing in faces. The framework is used to detect replay attacks, 3D mask attacks and print attacks. The two aspects of the proposed framework are:

- Dimensionality reduction using deep convolutional autoencoders. The pre-training step of dimensionality reduction on local and fixed regions of the faces improves the performance of the framework as well as convergence speed.
- Fine tuning in which weights are learned from the autoencoders,and the pre-trained encoder weights are further used for the task of classification into real and spoofed faces.

Results obtained on three popular benchmarks i.e Idiap Replay Attack [6] ,CASIA FASD [6] and 3DMAD[7] are comparable to many state-of-the-art methods used for detection of spoofed faces. The rest of the paper is organized as follows : Section 2 comprises of the related works on the detection of spoofing in faces. Section 3 presents the methodology used with results achieved using the framework in Sect. 4.Finally, Sect. 5 concludes the paper. Our main contribution in the paper is as follows:

- The paper explores the use of convolutional autoencoders to reduce the dimensionality of input images.Dimensionality reduction helps us to get rid of redundant features, thus improving the generalization capability of the framework.
- We propose the use of pre-trained encoder weights for feature extraction and classification of facial images reconstructed using the convolutional autoencoder. This helps to improve the performance of the framework.
- Liveness features are automatically learned by the layers in the convolutional neural network architecture. The classification of images into real and spoofed is performed with the help of a softmax classifier.
- Further, we compare the effect of dimensionality reduction using convolutional autoencoders with that achieved using Principal Component Analysis(PCA) on the input facial images. It is observed that the proposed approach achieves a superior classification accuracy which is comparable to state-of-the-art methods.

2 Related work

Extensive literature survey has been done in the domain of face detection. According to authors[8], biometric systems can be attacked directly as well as indirectly. In the direct attacks (spoofing attacks), they make use of photographs, gelatin fingers, contact lenses etc. to generate synthetic samples of biometric traits. These can gain access to the authentication systems.In the case of indirect attacks, they get details of the system's internal functionality. They further modify the algorithms used to protect biometric templates.Face is a promising trait for biometric authentication because of low acquisition cost and universality.Yet face recognition systems are can be fooled by using various methods[9].

Figure 1 shows the main points of attack of a biometric system.There are eight points of attacks in a biometric system,divided into four categories:i) attack at the user interface ii) attack between interface and modules iii) attack on modules iv) attacks on database[5].Out of these,the first kind of attacks i.e. those done at the input level have proved to be successful, as they only require a fake biometric spoof.Attacks involving dummy fingers have been found to be successful in Apple IPhone 5S and Samsung Galaxy S5. Other attacks

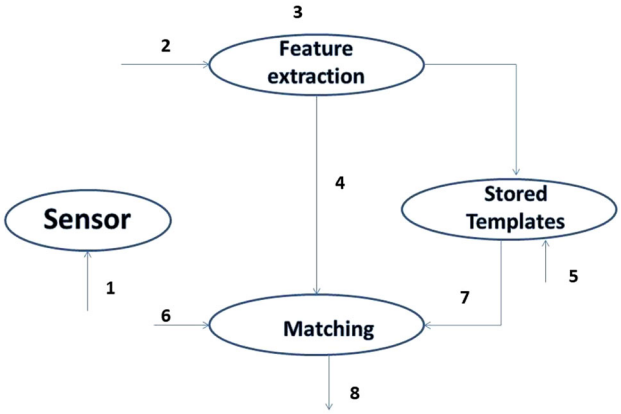

Fig. 1 Points of attack in biometric system

can be carried out on the sensor, on the features extracted and stored as templates,as well as on the matcher.

Face spoofing attack detection has become an important issue in face recognition since biometric sensors are being used in various authentication systems and mobile devices. Based on different kinds of forged faces, there are four types of presentation attacks: printed photo attack, displayed image attack, replayed video attack and 3D mask attack. Except for 3D mask attack, all other attacks are carried out in a 2D medium (paper, screen etc.).Attackers use a 3D face mask to carry out a 3D mask attack. In this section, we discuss some of the relevant literature in the field of spoofing detection.

2.1 Conventional methods to detect spoofed faces

A lot of research is going on in the field of spoofing detection in faces using fixed features.These features could be based on motion, texture,reflectance properties etc.These make use of handcrafted features to distinguish between real and spoofed faces.The features are calculated before training the data using any algorithm. Methods such as Haar-like features and Linear Discriminant Analysis(LDA) can be used to analyze faces in images and detect if any spoofing attack has been carried out[7].The work in [9] proposed the co-occurrence of adjacent local binary pattern (CoALBP) algorithm to extract facial texture features in order to detect spoofed faces. Authors[10] worked on texture differences to detect spoofing by analysing Fourier spectra of 2D and 3D images. They found that there was a difference in the frequency distributions of these images due to the property of surface reflection.Authors[11] extracted hidden face texture features using Difference-of-Gaussian (DoG) filters to distinguish between real and spoofed faces. Authors[12] propose a technique which extracts time-spectral descriptors from a video, thus capturing spatial as well as temporal information from a biometric sample. These descriptors are found to be robust in detecting several kinds of attacks in various scenarios.

Various motion-based methods have been proposed by researchers in the domain of spoofing detection. This involves the detection of movements such as eye blinking [13]. Other methods include the assessment of image quality and property of reflection to differentiate between real and fake faces. Authors make use of reflection, color diversity and blurriness between real and fake faces in order to detect spoofing.

2.2 Deep learning based techniques

These days, the most widely used approach to deal with spoofing is the use of CNN. These architectures used for face recognition find relationships between facial regions by making use of spatial information. However, the research is in progress, and little literature is available that studies detection of face spoofing using deep learning. These fall in the category of automatic learnable feature-based face anti-spoofing algorithms. A CNN network extracts features from real and spoofed faces at the time of training. This is done by mapping input pixels of an image to a final probability score, after feature extraction using several hidden layers. Most of the studies work on faces as a whole, or small patches taken from faces. Thus they do not learn local cues from each facial region. Authors used transfer learning with the help of pre-trained VGG16 for feature extraction from faces [14]. Spoofing detection was done using transfer learning in another approach[15]. Other works in the literature make use of global features by using similar pre-trained architectures. Authors explored random patches for face spoofing detection in their proposed work [16]. Despite giving good results, this approach can distract the network from learning spoofing cues. It would only learn the structural information available from facial regions. The backpropagation algorithm can be influenced by structural aspects of faces (such as the shape of eyes etc.), rather than learning the spoofed features as a whole. There are various datasets and systems which help to aid the development of robust face anti-spoofing techniques.

Authors used a CNN network to detect spoofing for all traits like face, iris and fingerprint [2]. Another approach combined the 2 layers of CNN with Long Short Term Memory Networks(LSTM) to detect spoofing in faces [5].LSTM is used to capture temporal information from frames. The model was robust but could not handle multiple attacks. Authors proposed a shallow CNN network to distinguish between authentic and spoofed faces [1]. They used non-linear diffusion with an additive operator, and the diffused image was given as an input to CNN to classify the image. As the network was shallow, it had lesser generalization capability. Authors proposed a CNN network to detect spoofing attacks on individuals, using data from two popular bench-

marks: CASIA-FASD database and Replay-Attack database. They localized the face regions before training the CNN network. In another work, authors tackled the 3D mask attacks by detecting the pulse from videos, which may be sensitive to camera settings and light conditions [17]. Besides, it becomes easier to obtain 3D masks by attackers with the development of 3D printing. Thus, it is necessary to develop new methods for detecting 3D face mask. Authors[18] proposed a novel fusion technique based on attention, which simulated the visual system of an individual to analyze a particular region of interest. This also improved performance as compared to existing methods. The framework used in [19] made use of a two-stream convolutional neural network which worked on two spaces: RGB colour space and multi-scale retinex space. The RGB space gave detailed information about the texture of faces. The latter captured high-frequency information about faces to discriminate between them. The experiments were conducted on various popular benchmarks and showed effective results.Authors explore the role of Conditional Random Fields (CRF) for the task of predicting blinking movements[20]. Another approach[21] combined the use of Local Binary patterns and Simplified Weber Local Descriptor encoded CNN models. The features extracted were combined to preserve the intensity of information and orientation of edges. Further, a Support Vector Machine (SVM) classifier was used to determine if the image was spoofed or not. Authors[22] demonstrated the effectiveness of CNN in the detection of spoofing by implementing their framework on a custom dataset. They achieved an appreciable accuracy of 96%.Authors explored the perspective of data instead of the network architecture in [23]. They analysed the role of hair geometry and nonlinear adjustment to find a contrast between real and spoofed faces. Further, a simple CNN network was used in different attack scenarios to detect spoofing. Authors[24] use depth maps to train their fully convolutional network (FCN), which gives better results as compared to some CNN-based approaches. Another framework combines the use FCN with a domain adaptation layer and lossless size adaptation. The DA layer improves generalization while LSA preserves the frequent spoofing clues obtained during recapture of faces[35].

2.3 Research gaps and motivation

- The use of autoencoders based on CNN architectures can help to compress data, reduce storage and improve the computation time. A convolutional autoencoder outperforms a simple autoencoder. We further evaluate how autoencoders give a superior performance as compared to conventional dimensionality reduction techniques such as Principal Component Analysis(PCA).
- As compared to the existing studies[1,2], the use of pre-trained encoder weights leads to an improvement in the rate of detection of spoofed faces. This is because the encoder part has already been trained during the dimensionality reduction phase and does not need to be re-trained.Thus, only the fully connected part of the CNN is trained for the classification and detection of spoofed faces.
- In various studies, hyperparameter tuning in the CNN architecture helped to achieve good results,however, there is a scope for improvement in the associated performance metrics obtained on popular benchmarks[5].

3 Methodology

3.1 Autoencoders vs PCA

An autoencoder is a neural network which encodes an image using encoder $g_e(\Theta)$ and reconstructs an image X using a decoder. This representation can be given as:

$$\hat{X} = Q(X) \tag{1}$$

Autoencoding is the an algorithm used for compression of data. The functions used for compression and decompression are:

- data specific i.e. they can compress data similar to what they were trained on. For example, an autoencoder trained on faces would perform well on compression of faces and not on other objects
- lossy i.e. the decompressed outputs would be a little degraded as compared to original inputs, but the reconstruction loss can be minimized.
- automatically learned from training instances i.e. no human-engineered features are required for this purpose.

In addition, neural networks are used to perform compression and decompression in an autoencoder. To build an autoencoder, we need an encoder, a decoder and a distance function to map the loss between the original input image as well as the compressed representation of the image. Autoencoders can be sparse or stacked. A sparse autoencoder consists of a single neuron connected with the input vector matrix, playing the role of an encoder. The hidden layer outputs a reconstructed output vector matrix which plays the role of a decoder. A stacked autoencoder, on the other hand, is a neural network with many layers of sparse autoencoders. The output of each hidden layer is connected to the input of the next hidden layers. Autoencoders are better as compared to various other methods used to reduce the dimensionality of images. Some of them are:

- Principal component analysis(PCA) is a commonly used approach to perform dimensionality reduction. It finds a linear subspace of dimension d which is lower than the dimension of the input, thus maintaining the variability in the data.
- Linear discriminant analysis(LDA) works on the concept of finding maximum -between-cluster-distance and minimum-in-cluster-distance in the new subspace.
- Locally linear embedding(LLE) is a nonlinear approach used for dimensionality reduction while also maintaining the embeddings of the high dimensional data. A dataset of n dimensions is assumed to lie near a nonlinear manifold of a lower dimensionality, and subsequently mapped to a single global coordinate system of lesser dimension.
- Isomap : It is a nonlinear generalization of dimensionality scaling. However, this kind of reduction is done in the geodesic space of the nonlinear data as compared to the input space.

In this paper, we make use of convolutional autoencoders for the process of dimensionality reduction. Autoencoders, in their traditional form, do not observe an input signal as a sum of other signals. However, the use of convolutions helps to encode a set of input signals and reconstruct the output by learning from them. This gives it an advantage over other variations of autoencoders. The network consists of several convolutional layers. Given image matrix $n = h \times w \times n_c, X \epsilon R_n$. h refers to the image height, w is the image width and n_c is the number of channels. The encoder g_e can be denoted as:

$$min_{\Theta,\alpha_e,\alpha_d,\sigma}(E[||X - g_d(Q(g_e(X;\Theta,\alpha_e));\alpha_d,\phi||_f)^2 + \gamma \Sigma_{i=1}^{n_c} H_i \quad (2)$$

Here, α_e and α_d are the encoder and decoder parameters, whereas gradient methods are used to find Θ and Φ. γ is the regularization term, whereas H_i denotes the output after application of an activation function on input i in a layer. In this paper, we also compare the results achieved using autoencoders and PCA. Both the techniques help to scale down the dimensionality of images, by reducing the number of redundant features and the amount of noise. As compared to autoencoders, PCA works by learning linear transformation and projecting data into another space,where variance defines the vectors of projections. Dimensionality reduction is achieved by restricting the number of components of the data, that account for most of the variance. Autoencoders work by reducing data to a lower-dimensional space by stacking multiple nonlinear transformations. The encoder part of the autoencoder maps image to a latent space and the decoder part reconstructs the image. With the help of these low dimensional latent variables, the output image can be reconstructed as these hold important features from the input.There are various reasons why autoencoders achieve better results as compared to PCA. Some of them are as follows:

- PCA features are linearly uncorrelated whereas autoencoded features might have correlations that help in accurate image reconstruction.
- PCA is used for linear transformation whereas autoencoders are used to model complex nonlinear functions.
- PCA can be computationally cheaper but autoencoders provide a better image reconstruction than PCA when there is a nonlinear relationship between features, i.e both in 2D and 3D space.

3.2 Proposed framework

In this work, we propose a novel deep learning framework for detection of spoofing in faces during biometric authentication. This would ensure more effective learning of spoofed features, on the basis of two steps:

- Dimensionality reduction phase in which the input images are reconstructed using a convolutional autoencoder. The hyperparameters of this model are chosen by experimentation and modification of parameters used in the pre-trained VGG16 model. These are finalized on the basis of lowest value of loss achieved during reconstruction of images.
- The encoder weights from this autoencoder are loaded into another architecture comprising of a Flatten layer and fully connected layer comprising of 1024 neurons. The images obtained after dimensionality reduction are fed to this model.
- A softmax classifier is further used to classify the image as real or spoofed.

The proposed network has been shown in Fig. 2.

The encoder and decoder part of the convolutional autoencoder consist of convolutional layers followed by max pooling.In case of the encoder, max pooling is used for thebibliography purpose of downsampling while it upsamples the image at the decoder end. The network of this convolutional autoencoder is tabulated in Table 1.

The next step constitutes of a forward-pass through a CNN network which makes use of the pre-trained encoder weights in a Flatten layer. The reconstructed input facial images to be trained are labeled as real or spoofed. With the addition of more fully connected layers, the entire network is setup for training these images. This is tabulated in Table 2. As the model uses pre-trained encoder weights, therefore these layers would not be re-trained. The images obtained after dimensionality reduction are fed to the fully connected layers of CNN added to the framework, along with a softmax

Fig. 2 Proposed framework

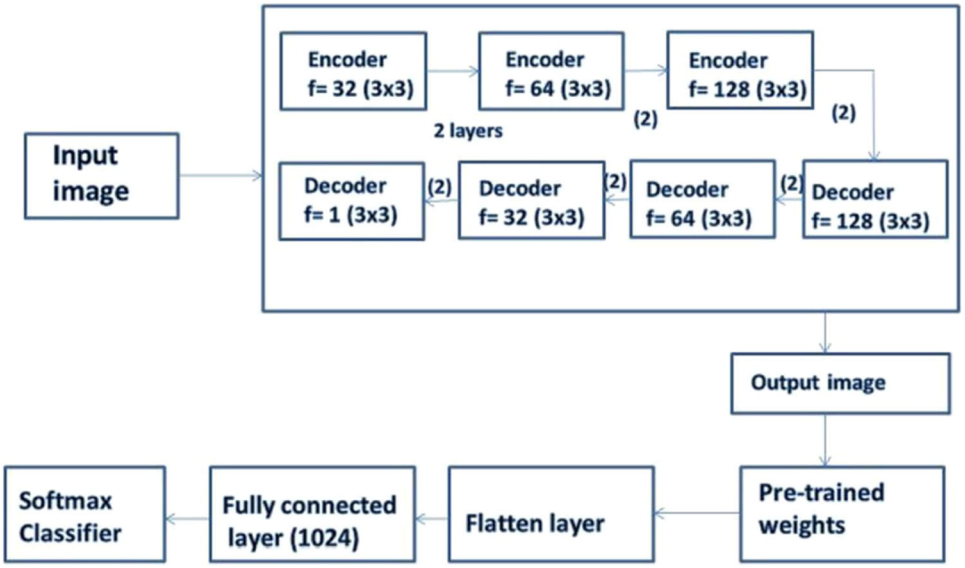

Table 1 Encoder and Decoder architecture

Block	Layer	Number of filters	Size	Activation
Encoder	1	32	3	Relu
Encoder	2	32	3	Relu
Encoder	Downsample1	-	2	-
Encoder	3	64	3	Relu
Encoder	4	64	3	Relu
Encoder	Downsample2	-	2	-
Encoder	5	128	3	Relu
Encoder	6	128	3	Relu
Encoder	Downsample3	-	2	-
Decoder	7	128	3	Relu
Decoder	8	128	3	Relu
Decoder	Upsample1	-	2	-
Decoder	9	64	3	Relu
Decoder	10	64	3	Relu
Decoder	Upsample2	-	2	-
Decoder	9	32	3	Relu
Decoder	10	32	3	Relu
Decoder	Upsample3	-	2	-
Decoder	11	1	3	Sigmoid

classifier which predicts whether the input facial image is real or spoofed.

For each forward pass, the network deals with a randomly sampled mini-batch.Our framework is trained for 100 epochs. The advantages of training the framework in this manner are::First, the network is trained on multiple frames and features can be efficiently extracted i.e more discriminative features can be learnt in this manner. Second, in the absence of dimensionality reduction there are high

Table 2 CNN architecture using encoder weights

Layer	Number	Activation
Flatten	-	-
Fully connected layer	1024	Relu
Fully connected layer	2	Softmax

chances of overfitting of data. Thus,reconstruction of images by downsampling the number of features would get rid of this problem. Also, the main reason for selecting encoder weights for classification is that randomly initialized weights or those trained using backpropagation would not be able to learn discriminative features from the input images very efficiently.The learning rate has been set to 0.01 for the specified number of epochs.

4 Experimental setup

4.1 Datasets used

We evaluate the proposed architecture on three popular benchmarks: CASIA-FASD, 3DMAD and Replay-Attack IDIAP database. The

– Replay-Attack Database: This database provided by the Idiap Research Institute and consists of live and spoofed videos of 50 subjects. There are 1300 video clips and each .mov clip is of around 9 seconds in length. All videos have been generated by clients by using a built-in webcam or using a video recording of themselves. To perform the attacks, high resolution videos are recorded in the same session using Canon PowerShot SX150 IS camera.

Table 3 Datasets used

Database	Type of data	Type of attack
3DMAD	Video clips	Mask creation according to face image
Replay-attack	Video clips	Recorded attack videos in different conditions
CASIA-FASD	Images and video clips	Warped photo attack, cut attack and video attack

Fig. 3 Frames of real and spoofed images(top to bottom) from **a** CASIA-FASD **b**3DMAD **c**Replay-Attack (left to right)

20 attack videos are recorded for each subject, in two different modes: controlled and adverse (10 for each).

- 3DMAD database: It consists of videos of real and spoofed faces, recorded for 17 clients using Kinect. Each recording consists of frames an RGB image along with a depth image and the labeled eye positions. 3D attacks are captured by an operator in the session. For this, real-sized face masks are used to carry out the attacks.
- CASIA-FASD database: This database consists of three types of attacks : warped printed photographs, printed photographs and cut attacks. To perform warped photo attack, both images and videos are recorded using a camera.The attacker deliberately tries to warp an intact photo and tries to simulate facial motion. In this attack, there is no cut-off region for the face unlike in cut photo attacks, in which eye regions are cut off and attacker hides behind the holes. Video attacks are performed on videos recorded through the camera.

Table 3 shows the type of data available in the three benchmarks.

4.2 Pre-processing of videos to frames

In the pre-processing stage, the videos from all the three datasets are converted into frames with the help of ffmpeg. It is an open source video processing framework which extracts frames from videos. Frames are extracted at a rate of 10 frames per second and a bit rate of 2318 kb/s. The images are

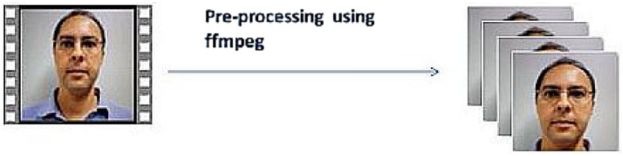

Fig. 4 Pre-processing of a video clip from the Replay Attack database

further input to the proposed framework for dimensionality reduction and feature extraction. Figures 3 shows examples of authentic images and spoofed images in all the three databases: CASIA-FASD,3DMAD and Replay-Attack (left to right). The frames on the top are genuine attempts, whereas those at the bottom are spoofed attempts. Figure 4 depicts the process of frame extraction from a video clip in the Replay-Attack database.

4.3 Dimensionality reduction and training

After pre-processing of videos using ffmpeg, the experiments are conducted on facial images from the three datasets. The images are reconstructed with the help of convolutional autoencoders.The output images are further subjected to feature extraction and classification of images into real or spoofed. Figure 5 shows the accuracy obtained during feature extraction of images input from the Replay-Attack database for the first ten epochs(out of 100 epochs). Figure 6 shows the loss curve during reconstruction of images using the convolutional autoencoder architecture, on images input from the

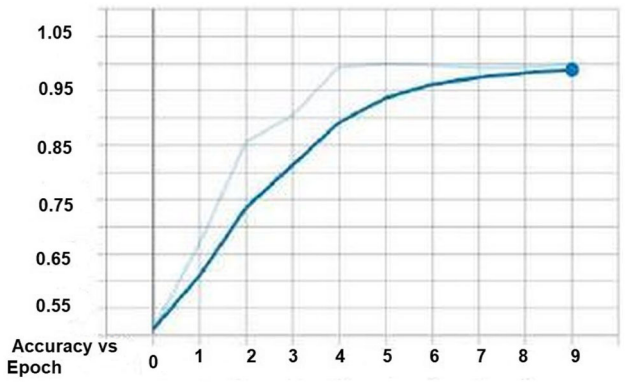

Fig. 5 Training loss for 10 epochs on Idiap-Replay Attack database

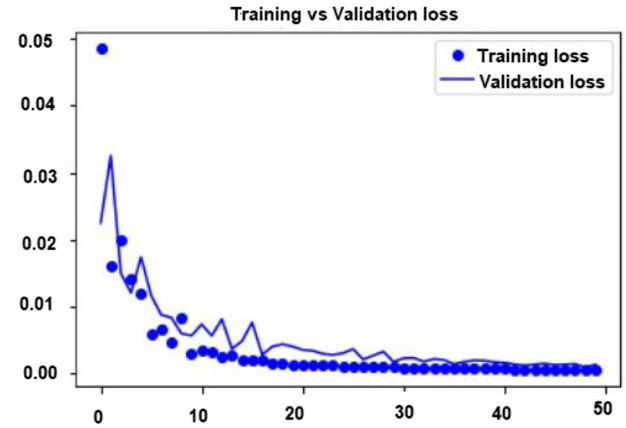

Fig. 6 Loss during reconstruction of images in 3DMAD dataset

3DMAD dataset. It can be observed that the loss has been minimized during the reconstruction of images using autoencoders.

5 Results

We compare the results achieved using convolutional autoencoders with those obtained using Principal Component Analysis on the three benchmarks. Tables 4 and 5 show this comparison on the basis of performance metrics like Accuracy,False Accept Rate(FAR), False Reject Rate(FRR) and Half Total Error Rate(HTER). The strength of the framework is further validated using performance metrics such as Precision, Recall and F1-score. This has been tabulated in Table 6 and 7 for the Replay-Attack database and 3DMAD database. Table 5 summarizes the results of spoofed face detection in which dimensionality reduction using convolutional autoencoders is followed by feature extraction using fully connected layers with pre-trained encoder weights. According to this table, our method outperforms various state-of-the-art techniques in terms of the mentioned metrics. The use of pre-trained weights improves the overall performance of the framework.

Table 5 Results achieved using proposed framework

Training dataset	Accuracy	FAR(%)	FRR(%)	HTER(%)
Replay Attack	99.03	0.045	0.032	3.85
CASIA	99.17	0.0176	0.0176	1.76
3DMAD	100	0	0	0

Table 6 Results achieved using proposed framework on the Replay-Attack database

Dataset	Precision	Recall	F1-score
Spoofed	1.00	0.97	0.98
Real	0.97	1.00	0.98

Table 7 Results achieved using proposed framework on the 3DMAD database

Class	Precision	Recall	F1-score
Spoofed	1.00	1.00	1.00
Real	1.00	1.00	1.00

5.1 Comparison with existing literature

In Table 8, the comparison has been performed with various studies which involve application of pulse based [25], motion-based techniques[26] and RGB color-based methods on the 3DMAD benchmark to detect spoofed faces. The pulse-based method used in [25] is capable of detecting unseen types of masks as well. Our proposed framework gives a better performance even when tested on other databases with unseen types of attacks, as observed in cross database testing. The performance achieved is better in terms of HTER as compared to the Local Binary Patterns technique Using Three Orthogonal Planes (LBP-TOP) in [26], which

Table 4 Results achieved using PCA

Training dataset	Accuracy(%)	FAR(%)	FRR(%)	HTER(%)
Replay Attack	76.60	0.175	0.0	8.79
CASIA	62.44	0.227	0.0	11.36
3DMAD	75.34	0.30	0.0	15.15

Table 8 Comparison with studies on 3DMAD

Method	HTER(in %)
CNN[20]	3.2
RGB color space [25]	8.88
CNN[25]	0
LBP [26]	12.29
rPPGS[29]	8.59
Motion based [30]	11.70
LBPTOP [30]	5.41
Image distortion analysis[31]	13.88
Pulse based[38]	7.9

Table 9 Comparison with studies on Replay Attack Database

Method	HTER(in %)
Component dependent descriptor [27]	5.93
LBP-TOP[30]	8.51
Partial CNN[36]	6.1
Partial CNN[36]	11.2
Partial CNN[36] (2016a)	9.8
Micro-texture analysis[37]	13.8
Linear Discriminant Analysis[39]	15.2

showed a drastic improvement as compared to other methods. Another technique used in [25] obtains an HTER of 0 in case of 3DMAD database. Our proposed framework attains this HTER, which is the lowest for any database. The technique used in [28] attains an HTER of 3.2% using a fully connected CNN.

In Table 9, we compare our results with studies implemented on the Replay Attack database.As compared to the deep learning techniques applied by authors in [27],the reduction observed in the HTER is more than 30 percent. The results achieved are better as compared to the technique used in [36] which extract features using partial CNN to detect spoofed images.Also, use of conventional techniques such as LBP-TOP [26] and Local Discriminant Analysis (LDA) [4], the values of HTER achieved are 8.51% and 15.2%, respectively.A lower HTER signifies that there is lesser false acceptance and false rejection of input facial images.

Our performance achieved is better as compared to other approaches which make use of CNN to detect spoofing in the CASIA-FASD benchmark [29].The comparison can be observed in Table 10. As compared to the existing approaches, there is a reduction in HTER by around 50% [31]. The technique used by [32] works on the use of continuous data randomization so as to work effectively on limited amount of data. The HTER achieved is 4.59%. An HTER of 10.65% is achieved by authors, with the use of 3D CNN on short video frame levels.

Table 10 Comparison with studies on CASIA-FASD

Method	HTER (in %)
Patch based CNN[29]	2.27
Fusion method[29]	3.88
Concatenated approach on print attacks[31]	3.88
Livenet [32]	4.59
3D CNN [33]	10.65

5.2 Cross-database testing

We also perform inter-database testing using models trained on other benchmarks to validate strength of our framework. Table 11 shows the results of cross-database testing which shows promising results as compared to existing works. To evaluate the performance across dataset A and dataset B, an inter-test is executed by using training set from one benchmark and testing set from the other. Figure 7 visualizes the performance metrics reported for these tests using the proposed method and by comparing with results achieved using existing techniques.These involve the application of CNN [27],motion-based techniques [30] and the use of adversarial learning along with machine learning to detect spoofed faces [34], It is observed that our proposed approach can obtain good performance, even with a reduced accuracy of 39.89% on the CASIA-FASD dataset and that of 41.38% on the Replay Attack database.

As it is observed, both inter-test and intra-test on the three benchmarks achieve very good results using convolutional autoencoders. While the accuracy is nearly 100% for the 3DMAD dataset,an accuracy of more than 99% is achieved on both Replay Attack database and CASIA FASD database. The main reason for such a performance is the reduction in overfitting and efficient feature extraction using pre-trained encoder weights. The accuracy achieved for inter-database testing is more as compared to that achieved in intra-database testing. This is because there is a significant degradation in the performance if testing is done on a different dataset. Different from the traditional CNN architectures,the proposed method reduces overfitting and improves detection performance using pre-trained weights,thus bridging the gap between the feature distributions and achieving good performance, only with the presence of limited amount of training data.

From the figure, it can be observed that the results of cross-database testing achieved by application of a different trained model on the CASIA-FASD dataset are comparable to those achieved using state-of-the-art methods. Figure 8 shows the detection of a spoofed face in the CASIA-FASD database using a model trained on the same dataset.

Table 11 Results achieved in cross-database testing

Training dataset	Testing dataset	Accuracy (%)	FAR (%)	FRR (%)	HTER (%)
3DMAD	CASIA	60.76	0.406	0.3914	39.89
CASIA	Replay Attack	59.15	0.422	0.405	41.38

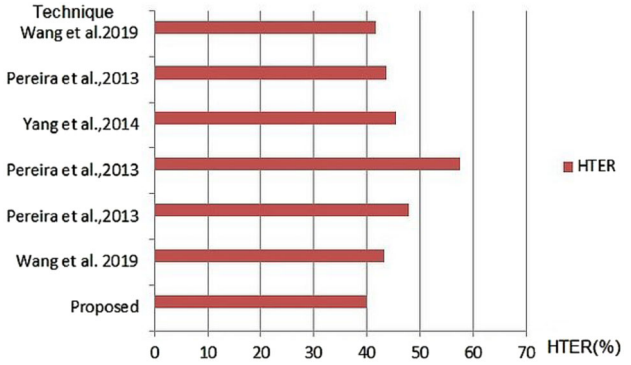

Fig. 7 Comparison of studies involving testing cross-database on CASIA-FASD

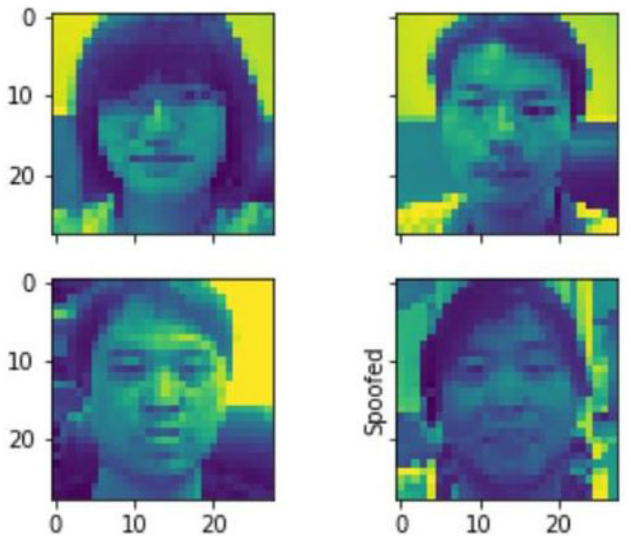

Fig. 8 Prediction of spoofed faces in CASIA-FASD database

6 Conclusion and future work

The proposed study is useful in the detection of various kinds of spoofing attacks in biometric systems that use faces for authentication. These attacks are replay attacks, 3D mask attacks and photo attacks. Our framework uses convolutional autoencoders to reduce the dimensionality of images. This is followed by feature extraction and classification using pre-trained encoder weights and a softmax classifier. Our proposed framework is robust and efficient, which is validated on three benchmarks using intra-database and cross-database testing. The framework gives a solution for problems like overfitting, and performance improves due to pre-trained encoder weights. Our results are comparable to state-of-the-art methods in terms of performance metrics, with an accuracy of more than 99% in case of intra-database testing. Further, we perform cross-database testing and compare the results achieved on CASIA-FASD with other existing works in this domain. It is observed that even with a decline in the performance, the results are on par with the compared approaches. Thus, the framework can detect spoofed faces under various scenarios during authentication in a biometric system. The limitation of this framework is that there would be some loss in the reconstruction of images. Although our framework does not radically degrade the image quality, we can further work on the optimization of hyperparameters to further reduce the loss and network complexity. This would be a part of our future work. We would work on even more challenging datasets and improve the performance achieved in cross-database experiments. We would also apply the framework to detect spoofing attacks on other biometric traits and compare our results with already existing approaches. Further, we will explore the domain of behavioural biometrics for vulnerabilities like spoofing attacks.

Declarations

Funding The authors have not received any funding for this research

Conflict of interest The authors Shefali Arora, M.P.S Bhatia and Vipul Mittal declare that they have no conflict of interest.

References

1. Alotaibi, A., Mahmood, A.: Deep face liveness detection based on nonlinear diffusion using convolution neural network. SIViP **11**(4), 713–720 (2017)
2. Menotti, D., Chiachia, G., Pinto, A., Schwartz, W., Pedrini, H., Falco, A., Rocha, A.: Deep representations for iris, face, and fingerprint spoofing detection. IEEE Trans. Inf. Forensics Secur. **10**(4), 864–879 (2015)
3. Tirunagari, S., Poh, N., Windridge, D., Iorliam, A., Suki, N., Ho, A.: Deep representations for iris, face, and fingerprint spoofing detection. IEEE Trans. Inf. Forensics Secur. **10**(4), 762–777 (2015)
4. Galbally, J., Marcel, S., Fierrez, J.: Biometric antispoofing methods: a survey in face recognition. IEEE Access **2**(1), 1530–1552 (2014)
5. Feng, L., Po, L., Li, Y., Xu, X., Yuan, F., Cheung, T., Cheung, K.: Integration of image quality and motion cues for face anti-spoofing: a neural network approach. J. Vis. Commun. Image Represent. **38**(C), 451–460 (2016)
6. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.: A face antispoofing database with diverse attacks. International Conference on Biometrics 26–31, (2012)

7. Erdogmus N., Marcel S.: Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect. 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, 1–6 (2013)
8. Li, L., Correia, P., Hadid, A.: Face recognition under spoofing attacks: countermeasures and research directions. IET Biom **7**(1), 3–14 (2017)
9. Gragnaniello, D., Poggi, G., Sansone, C., Verdoliva, L.: An investigation of local descriptors for biometric spoofing detection. IEEE Trans. Inf. Forensics Secur. **10**(4), 849–863 (2015)
10. Li, J., Wang, Y., Tan, T., Jain, A.K.: Live face detection based on the analysis of fourier spectra. Biometric technology for human identification. SPIE **5404**(1), 296–304 (2004)
11. Mahitha, M.: Face spoof detection using machine learning with colour features. Int. Res. J. Eng. Technol. **9**(5), 1–4 (2018)
12. Tan, X., Li, Y., Liu, J., Jiang, L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) Computer Vision - ECCV, : Lecture notes in computer science 6316, pp. 504–517. Springer, Heidelberg (2010)
13. Bharadwaj, S., Dhamecha, I., Vatsa, M., Singh, R.: Face anti-spoofing via motion magnification and multifeature videolet aggregation. Indraprastha Inst. Inf. Technol., New Delhi, India IIITD-TR2014-002(002) 1-14 (2015)
14. Tammina, S.: Transfer learning using vgg-16 with deep convolutional neural network for classifying images. Int. J. Sci. Res. Pub. (IJSRP) **9**(10), 9420 (2019)
15. Tu, X., Fang, Y.: Ultra-deep neural network for face anti-spoofing. Neural Inf. Process. **10635**(1), 686–695 (2017)
16. Zinelabidine, B., Akhtar, Z., Feng, X., Hadid, A.: Analysis of textural features for face biometric anti-spoofing. Face Anti-spoofing in Biometric Systems 1–5, (2017)
17. Haoliang, L., Peisong, H., Wang, S., Rocha, A., Jiang, X., Kot, A.: Learning generalized deep feature representation for face anti-spoofing. IEEE Trans. Inf. Forensics Secur. **13**(4), 2639–2652 (2016)
18. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)
19. Chen, H., Hu, G., Lei, Z., Chen, Y., Robertson, N., Li, S.: Attention-based two-stream convolutional networks for face spoofing detection. IEEE Trans. Inf. Forensics Secur. **15**(1), 578–593 (2019)
20. Sun, L., Pan, G., Wu, Z., Lao, S.: Blinking-based live face detection using conditional random fields. Lect. Notes Comput. Sci. **4642**(1), 252–260 (2007)
21. Khammari, M.: Robust face anti-spoofing using CNN with LBP and WLD. IET Image Proc. **13**(11), 1880–1884 (2019)
22. Alrikabi, W., Alnasrallah, A., Alrikabi, A.: System designing for face spoofing detection using convolution neural network (cnn). Solid State Technol. **63**(6), 14646–14655 (2020)
23. Sun, Y., Xiong, H., Yiu S.: Understanding deep face anti-spoofing: from the perspective of data. Vis Comput, 1–15 (2020)
24. Atoum Y., Liu Y., Jourabloo A., Liu X.: Face anti-spoofing using patch and depth-based cnns. Proc. IEEE Int. Joint Conf. Biometrics(IJCB), 319–328 (2017)
25. Li, L., Xia, Z.: 3d face mask presentation attack detection based on intrinsic image analysis. arXiv:1903.11303v1 [cs.CV], 1–24 (2019)
26. Pereira, T., Komulainen, J., Anjos, A., Martino, J., Hadid, A., Pietikainen, M., Marcel, S.: Face liveness detection using dynamic texture. EURASIP J. Image Video Process. **2014**(2), 1–15 (2014)
27. Yang, J., Lei, Z., Liao, S., Li, S.: Face liveness detection with component dependent descriptor. International Conference on Biometrics 85–100, (2013)
28. Sun, L., Pan, G., Wu, Z., Lao, S.: Blinking-based live face detection using conditional random fields. Adv. Biom. **4642**, 252–260 (2007)
29. Liu, S., Yuen, P., Zhao, S.: 3d mask face anti-spoofing with remote photoplethysmography. Proc. ECCV **9911**(1), 85–100 (2016)
30. Pereira, T., Anjos, A., Marcel, S.: Lbptop based counter-measure against face spoofing attacks. ACCV 2012: Comput. Vis. - ACCV 2012 Workshop **7728**, 121–132 (2012)
31. Wen, D., Han, H., Jain, A.: Face spoof detection with image distortion analysis. IEEE Trans. Inf. Forensics Secur. **10**(4), 746–761 (2015)
32. Rehman, Y., Po, L., Liu, M.: Livenet: improving features generalization for face liveness detection using convolution neural networks. Expert Syst. Appl. **108**, 159–169 (2018)
33. Gan, J., Li, S., Zhai, Y.,Liu, C.: 3d convolutional neural network based on face anti-spoofing. 2nd International Conference on Multimedia and Image Processing (ICMIP), 252–260 (2017)
34. Wang, G., Han, H., Shan, S., Chen, X.: Improving cross-database face presentation attack detection via adversarial domain adaptation. International Conference on Biometrics 1–8, (2019)
35. Sun, W., Song, Y., Zhao, H., Jin, Z.: A face spoofing detection method based on domain adaptation and lossless size adaptation. IEEE Access **8**, 66553–66563 (2020)
36. Li, L., Feng, X., Boulkenafet, Z., Xia, Z., Li, M., Hadid, A.: An original face anti-spoofing approach using partial convolutional neural network.2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), 1-6 (2016)
37. Matta, J., Hadid, A., Pietikainen, M.: Face spoofing detection from single images using micro-texture analysis. International Joint Conference on Biometrics (IJCB) 1–7, (2011)
38. Li, X., Komulainen, J.: Generalized face anti-spoofing by detecting pulse from face videos. International Conference on Pattern Recognition 4244–4249 (2017)
39. Galbally, Marcel S., Fierrez, J.: Image quality assessment for fake biometric detection: application to iris, fingerprint and face recognition. IEEE Trans. Image Process. **23**(2), 710–724 (2014)
40. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.: A face anti-spoofing database with diverse attacks. International Conference on Biometrics 26–31 (2012)

Shefali Arora is currently pursuing a Ph.D. in from the Department of Computer science and Engineering, Netaji Subhas Institute of Technology, New Delhi (University of Delhi). Her areas of interest include machine learning, Convolutional Neural Networks, data mining and cyber security.

M.P.S. Bhatia is a Professor in the Department of Computer Science at Netaji Subhas University of Technology, Delhi. He has been working with the University for more than 25 years. He is the Head of Training and Placement at the university. With an expansive teaching and research profile, Prof. Bhatia has guided many masters & doctorate students. He has presented several papers in international conferences and published work in peer-reviewed and science cited journals. His research interests include but not limited to Data Mining, Cyber Analytics, Soft Computing, Social Media Analytics and Software Engineering.

Vipul Mittal is an engineering student with a deep interest in machine learning and computer vision. He is pursuing his Btech in Computer Science from Netaji Subhas Institute of Technology.