International Journal of Engineering

Journal Homepage: www.ije.ir

Enhanced Face Presentation Attack Prevention Employing Feature Fusion of Pre-trained Deep Convolutional Neural Network Model and Thepade's Sorted Block Truncation Coding

S. D. Thepade*, M. R. Dindorkar, P. R. Chaudhari, S. V. Bang

Computer Engineering Department, Pimpri Chinchwad College of Engineering, Pune, India

PAPER INFO

Paper history:
Received 23 October 2022
Received in revised form 24 November 2022
Accepted 24 November 2022

Keywords:
Face Presentation Attack
Transfer Learning
DensNet121
VGG16
Xception
VGG19
Pretrained Deep CNN
InceptionV3
MobileNet
Thepade's SBTC

ABSTRACT

The evolution and improvements of deep learning are being used to tackle any research obstacles that could be converted into classification problems in all spheres of life. Each Deep convolutional neural network (DCNN) design's output is determined by the depth and value of the hyperparameters, which explains why so many of them have been proposed. These DCNN architectures must be created entirely from scratch, and they can only be used for the applications for which they were intended. Transfer learning may be used to modify these pre-trained networks so they are more appropriate for particular purposes. This article aims to evaluate the empirical performance of the applicability of pre-trained DCNN models to identify human face presentation threats (FPAD). Human FPAD is one of the most significant and crucial areas of research right now because of the introduction of ambient computing, which necessitates contact-free identification of persons with the help of their biometric traits. Six pre-trained DCNN models are taken into account for an experimental evaluation in human FPAD alias VGG19, VGG16, DensNet121, MobileNet, Xception, and InceptionV3. The investigation makes use of the NUAA and Replay-Attack benchmark FPAD datasets. Thepade's sorted block truncation coding (SBTC) 10-ary features are merged with deep learning features produced from the finest performing finetuned DCNNs to enhance the FPAD capabilities of analyzed machine learning (ML) classifiers. The integration of features of Thepade's SBTC 10-ary and DCNN has considerably increased the FPAD accuracy of ML classifiers with slightly more computations of feature extraction.

doi: *10.5829/ije.2023.36.04a.17*

1. INTRODUCTION

The security and access management of assets in the contemporary era of pervasive computing is largely dependent on the biometric characteristics of a person, such as a face [1-4], iris [5-7], fingerprints [8, 9], etc. Due to the numerous traits shown by the face and the simplicity of using face recognition algorithms, face recognition has emerged as an important physiological biometric technique utilized in information security [10]. Attackers trick these systems by using a variety of spoofing techniques, including photo, video, cut-photo, mask attacks, etc. Existing facial recognition systems should be modified to counteract these threats. Current antispoofing techniques use features based on shape, motion, depth, color, texture, and deep learning for spoof identification. Convolutional neural networks (CNN) are being researched in several fields of image processing [11-13]. For image classification tasks, there are numerous accessible pre-trained deep CNN models similar to Inception V3, VGG19, VGG16 etc.

A very precise model must have huge annotated datasets as a minimum requirement. It is difficult to obtain such massive datasets for any domain. A method called transfer learning was developed to reduce this requirement. Transfer learning, in general, is a method that adapts a model which has already been trained to execute one job to perform another. The transfer learning

*Corresponding Author Institutional Email: sudeepthepade@gmail.com (S. D. Thepade)

Please cite this article as: S. D. Thepade, M. R. Dindorkar, P. R. Chaudhari, S. V. Bang, Enhanced Face Presentation Attack Prevention Employing Feature Fusion of Pre-trained Deep Convolutional Neural Network Model and Thepade's Sorted Block Truncation Coding, ***International Journal of Engineering, Transactions A: Basics***, Vol. 36, No. 04, (2023), 807-816

method accelerates the development of new models while also enhancing their functionality. Several transfer learning strategies are employed depending on the study domain and the available data. However, manually created features use less computational power and let us extract global texture data from various parts of the image. Handcrafted features occasionally give CNN more data in a variety of classification tasks. As a result, the current work suggests an FPAD technique by combining deep learning features and manually creating texture characteristics from Thepade's SBTC [14-18].

The following are the key contributions of the current paper:

- Performance assessment of 6 different pre-trained DCNNs for FPAD.
- Capability evaluation of Thepade's SBTC features of 10-ary for different Machine Learning (ML) classifiers in FPAD.
- Improving the FPAD capability of a classifier by fusing Thepade's SBTC 10-ary texture features and auto-extracted DCNN features.

The sections in the paper are placed as given here. Section 1 has an introduction; section 2 surveys pertinent literature; section 3 elaborates on current deep CNN models; section 4 contains the proposed technique; section 5 illustrates the environment set up for experimentation; section 6 converses the observed results, and section 7 has concluding remarks.

2. LITERATURE SURVEY

Recent attempts to use pre-trained CNN models for FPAD have had some notable success. Here is a quick summary of their endeavours.

A CNN architecture called FASNet, which is a modified version of VGG16, is proposed by Lucena et al. [19]. The authors used back-propagation to adjust the weights to the uppermost layers starting from the fourth block. The findings were produced on the datasets 3DMAD (3D Mask Attack Dataset) and Replay-Attack to appraise the significance of the suggested strategy.

The method using the Rotation Invariant Local Binary Pattern (RILBP) and the ResNet-18 pre-trained CNN model was proposed by Chen et al. [20]. Support Vector Machine (SVM) is trained with the fused features from ResNet-18 and RILBP for binary classification. In intra- and cross-database testing, better results were obtained using a pre-trained CNN model.Tang et al. [21] suggest a method for obtaining class probabilities that involve feeding colour, temporal, and patch-based characteristics to a pre-trained CNN architecture called ResNet-18. To create a class-probability vector, these class probabilities are further concatenated. Therefore, SVM receives a vector of class probability for binary categorization.

The suggested approach by Tu et al. [22] locates pertinent hidden features from an input facial image using a pre-trained CNN model called ResNet-50. ResNet-50, which learns temporal properties from frame sequence, is put on top of LSTM with the exception of the top layers. The face's authenticity can be determined using these learning attributes.

Das et al. [23] suggest a novel human face antispoofing method that combines hand-crafted features extracted from an input face image using an LBP descriptor with deep features derived via VGG16.

Two Presentation Attack Detection (PAD) techniques are developed by Elloumi et al. [24] based on deep learning with the quality evaluation of the image. The first strategy uses the LBP histogram computation and VGG16 finetuning, whereas the other strategy relies on Image Quality Measures.

The work of Song and Hongbin [25] makes use of the feature extractors like ULTP (Uniform Local Ternary Pattern) and ULBP (Uniform Local Binary Pattern) with G-R color (color-INF and color-MMT). The SVM classifier uses these extracted features to further combine them and decide if the given face image is real or fake. On three separate face antispoofing datasets, the approach has produced notable results.

These nine different types of features as MeanRBG, SSIM, Energy, MeanYCbCr, Entropy, SBTC, BTC, LBP, and Luminance, are extracted and explored by Jagdale and Thepade [26], Thepade et al. [27]. To determine if a face is alive, the further fusion of these features is given to SVM. Although Thepade's SBTC produces notable results, the method is only tested on the NUAA dataset.

Thepade et al. [28] studied two ML classifiers, three ensembles, and three colour spaces: YCrCb, Kekre-LUV, and CIE-LUV. By computing a histogram over the colorspaces under consideration, features are derived. These features are used to train the classifiers and ensembles indicated above. The combination of CIE-LUV, YCrCb, and RandomForest produces impressive results for FPAD. The method's performance is evaluated on two datasets, Replay-attack and NUAA.

The fusion of features formed by using pre-trained DCNN models and conventional content features is presented by Abdullakutty et al. [29]. Colour LBP was extracted in YCbCr and HSV colour spaces for content features. The ResNet-50 with CLBP has given better performance over the three datasets used.

For 2D face PAD, a video preprocessing method called Temporal Sequence Sampling (TSS) by removing the predicted inter-frame 2D affine motion with the features of a Convolutional Neural Network (CNN) through a self-supervised representation learning scheme is proposed by Usman et al. [30].

Face presentation attack detection using various openly available datasets and their combinations are presented by Abdullakutty et al. [31], where the binary classification using transfer learning is explored for attack detection.

3. EXISTING DCNN MODELS

The six well-known DCNN models, MobileNet, VGG16, InceptionV3, Xception, Densenet121, and VGG19, are empirically evaluated here. With the aid of the transfer learning method for FPAD, all of these well-known DCNN models are explored.

Simonyan et al. [32] trained a CNN model known as VGG16 having a total of 16 layers (3 fully connected preceded by 13 convolutional) with the ImageNet data, having about 1000 classes and 14 million images. A total of 19 layers (3 fully connected preceded by 16 convolutional) make up VGG19 [25]. The VGG19 was significantly improved by enhancing the layer count from 16 to 19.

The 42 layers-deep InceptionV3 [33] network took the first runner-up in the ILSVRC-2015 because of its low error rate achievement (ImageNet Large Scale Visual Recognition Competition). Despite having a 2.5-fold higher computational cost than InceptionV1 (GoogleNet), InceptionV3 is more effective than VGGNet.

DenseNet121 is available in four different versions: DenseNet201, DenseNet169, DenseNet161, and DenseNet121. Each layer of DenseNet is connected as feed-forward [34]. The main benefits of these networks are 'encouragement to feature reuse' and 'reduction in concerns of vanishing gradient.' The DenseNet121 architecture has 121 layers.

The architecture of Xception [35] is 36 layers thick. The architecture of Xception was inspired by that of Inception, where depth-wise removable convolutions were used in place of the inception modules. There are exactly the same amount of parameters in Xception and InceptionV3.

The MobileNet [36] architecture comprises 28 layers when the pointwise and depthwise convolutions are considered separate layers. The pointwise and depthwise convolutions created in the form of piles are used to organise the separable convolution modules found in the MobileNet.

4. PROPOSED APPROACH

A stored network that has been trained on a sizable dataset to do general image categorization is the pre-trained model. The machine learning (ML) algorithm learns from the specific dataset and applies that information to others during the transfer learning process. Two methods can be used to implement transfer learning, (i) The pre-trained model is taken as an automatic feature extractor in a particular image categorization task. (ii) Fine-tuning the pre-trained model, which involves either retraining the entire model or only a section of it using fresh data. Transfer learning is frequently used when there is little data to prevent overfitting. Transfer learning provides the advantage of reducing ‘time’ and ‘computation resources’ needed in the training stage.

The gray version of the input face image is taken in work proposed here (Figure 1), which is followed by passing it through the Haar-cascade classifier to determine the area of interest (ROI), which is shrunk to fit within a '224 x 224' pixel window. The suggested methodology leverages the finetuning strategy on various pre-trained DCNN models for FPAD tasks after preprocessing. For each pre-trained model, a classification head having two fully connected layers with sizes "1" and "256" correspondingly replaces the final fully connected layers (FC). Between these FC layers is a dropout layer with a 0.2% dropout rate in order to prevent overfitting. The adam optimizer is used with a learning rate of 10^{-4} and weight decay at a rate of 10^{-6} (refer to Table 1). Additionally, sigmoid is employed (rather than softmax) decision function, which is suitable in binary categorization.

Finetuning the model using a randomly initialized classification head (depicted by pink color in Figures 2 to 7) may make the pre-trained base model disremember what it learned due to massive gradient updates. Hence here, the base pre-trained model is set to be non-trainable, and the classification head is trainable for the first 5

TABLE 1. Optimizer parameter set for DCNN Models in experimentation for person face liveness detection

Optimizer Parameters	Value
Optimizer	Adam
Beta-1	0.9
Learning rate	0.0001
Beta-2	0.999
Batch size	10^2
Epsilon	1×10^{-8}

TABLE 2. Parameters used to enhance the data during the DCNN model's training

Parameter for Data Augmentation	Value
Rescale	1/0.255
Fill mode	nearest
Rotation range	20°
Range of Width shift/ Shear/ Height Shift/ Zoom	0.2

epochs of training. Then the last two blocks of the base pre-trained model (shown in the red-marked rectangle) are set to be trainable (unfrozen). Further, the classification head and the base model are jointly trained using Keras's[1] early stopping criteria with parameter 'patience' equal to 10, indicating the training process will be automatically terminated if enhancement is not detected in test accuracy over 10 epochs. Further, to mitigate overfitting in training, data augmentation is carried out on the training dataset using the parameters listed in Table 2. The last 2 convolution blocks enclosed with a red border are finetuned in the case of VGG16, VGG19 and DenseNet121, as specified in Figures 2, 3, and 4, respectively. As depicted in Figure 5, the lattermost 2 inception blocks are chosen for finetuning in the case of InceptionV3. The bottommost 2 separable convolution blocks are finetuned for Xception (see Figure 6). Similarly, as per Figure 7, the last 2 depthwise separable convolution blocks are finetuned for MobileNet.

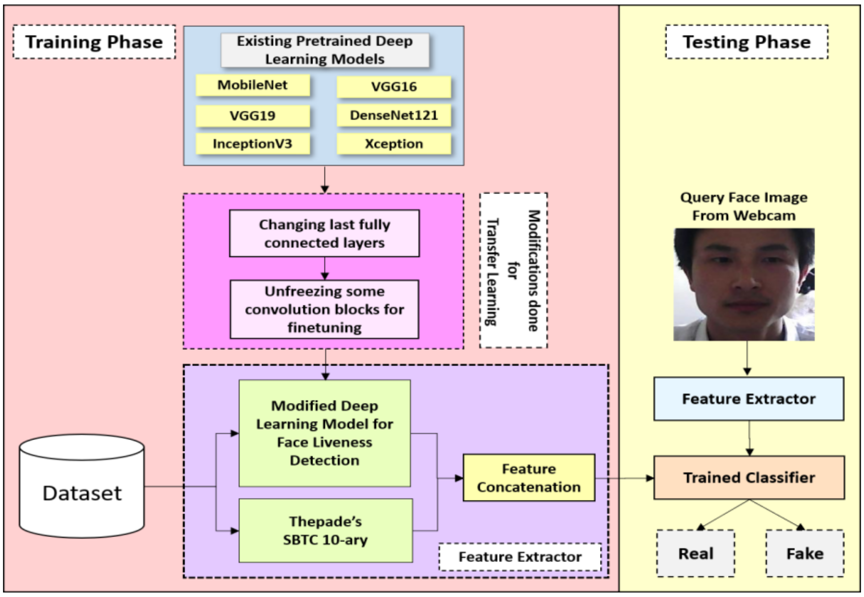

Figure 1. Flowchart illustrating the proposed method of human face liveness detection

Figure 2. Custom-made VGG16 model indicating the finetuned convolution blocks with red borderline

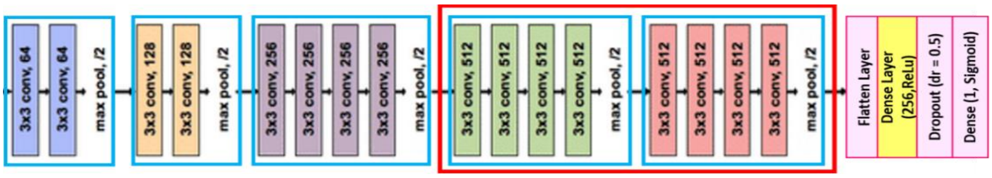

Figure 3. Custom-made VGG19 model depicting the finetuned convolution blocks with the red borderline (Output of dense layer indicated by a yellow color, is utilized to extract deep learning features that are used for feature fusion with Thepade's SBTC 10-ary)

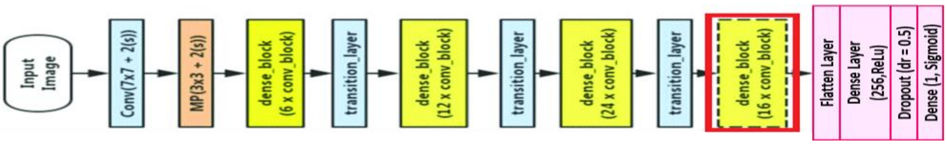

Figure 4. Custom-made DenseNet121 model indicating the finetuned convolution block with red borderline

[1] www.keras.io

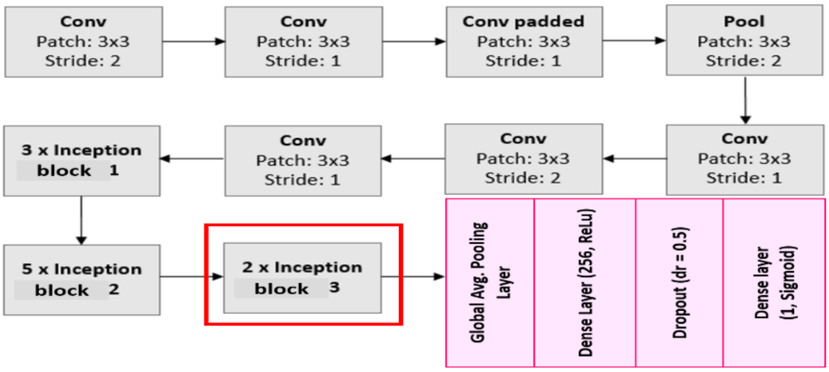

Figure 5. Custom-made InceptionV3 model indicating the finetuned inception blocks with red borderline

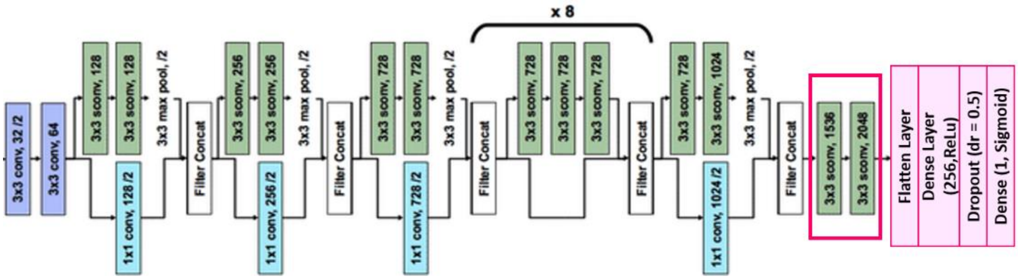

Figure 6. Custom-made Xception model indicating the finetuned separable convolution blocks by red borderline

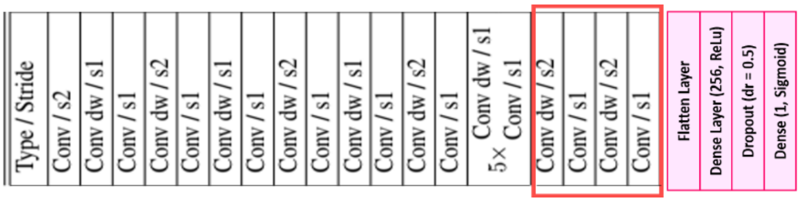

Figure 7. Custom-made MobileNet model indicating the finetuned convolution blocks with red borderline

4. 1. Thepade's SBTC 10-Ary Based Handcrafted Feature Generation of Input Face Image Let the input face image captured by the camera be described as FI (x, y) of dimension 'r * c * 3', where 'r' denotes the count of pixels across the X-axis, 'c' represents the count of pixels across Y-axis and 3 corresponds to Red, Blue, and Green color planes. The feature vector formed using Thepade's SBTC N-ary is assumed as [TR_1, TR_2, TR_3,.., TR_n, TG_1, TG_2, TG_3, ..., TG_n, TB_1, TB_2, TB_3,.., TB_n]. Where the TR_i, TG_i, and TB_i represent the respective color plane centroids for cluster' i' calculated with the help of Thepade's SBTC N-ary methodology.

Thepade's SBTC [37] 2-ary may get computed through equations 1 to 6. Let sortRed, sortGreen, and sortBlue be the sorted version of 1-dimensional arrays corresponding to relevant color planes R, G, and B of the input face image.

$$TR_1 = \frac{2}{x*y}\sum_{p=1}^{\frac{x*y}{2}} sortRed(p) \tag{1}$$

$$TR_2 = \frac{2}{x*y}\sum_{p=1+\frac{x*y}{2}}^{x*y} sortRed(p) \tag{2}$$

$$TG_1 = \frac{2}{x*y}\sum_{p=1}^{\frac{x*y}{2}} sortGreen(p) \tag{3}$$

$$TG_2 = \frac{2}{x*y}\sum_{p=1+\frac{x*y}{2}}^{x*y} sortGreen(p) \tag{4}$$

$$TB_1 = \frac{2}{x*y}\sum_{p=1}^{\frac{x*y}{2}} sortBlue(p) \tag{5}$$

$$TB_2 = \frac{2}{x*y}\sum_{p=1+\frac{x*y}{2}}^{x*y} sortBlue(p) \tag{6}$$

In Thepade's SBTC 2-ary, the sortRed, sortGreen and sortBlue arrays got divided into 2 clusters to get feature vector elements. Similarly, a global feature vector is computed using Thepade's SBTC 10-ary method, where the sortRed, sortGreen and sortBlue arrays are divided into 10 clusters. Per cluster, the centroid as the average of that cluster is computed, and the feature vector is obtained by fusing [TR_1, TR_2, TR_3,…, TR_{10}], [TG_1, TG_2, TG_3,…TG_{10}], and [TB_1, TB_2, TB_3,…TB_{10}] vectors. The feature vector size computed using Thepade's SBTC 10-ary is (10 * 3) 30 units.

5. EXPERIMENTATION ENVIRONMENT

Here, experimental evaluation is conducted using a platform made available by Kaggle[1], an online community for machine learning professionals and data scientists. The Python code written in Kaggle notebooks is run using the GPU as an accelerator for experimentation here.

Here, two widely used datasets for human FPAD, dubbed "Replay-Attack" and "NUAA," are investigated to validate the performance of current DCNN models using transfer learning. For each improved DCNN model, the testing accuracy is employed as a performance metric.

The Replay-Attack dataset [38] was generated by the IDIAP Research Institute and contained 1300 movies of 50 people under two different brightness settings as controlled and unfavourable during the acquisition (shown in Figure 8). The collection contains only videos that were shot at a 25 Hertz frame rate.

The NUAA dataset was generated by the Nanjing University of Aeronautics and Astronautics [39]. A web camera is used to capture 15 people's real and faked face photos. 5105 genuine and 7509 artificial face photos make up the data split over training and testing. Each facial image is 640 x 480 pixels in size. Only a sort of spoofing assault, 'a photo attack', is listed in the database, which includes changes in appearance, such as gender, light, and whether or not glasses are worn. The NUAA dataset's few candidate faces are given in Figure 9. Table 3 lists the count of training and testing sets collected from real and fake face photos for the datasets "Replay-Attack" and "NUAA." False Rejection Rate (FRR), Half Total Error Rate (HTER), False Acceptance Rate (FAR), and accuracy are each represented by Equations (7), (8), (9), and (10), correspondingly.

$$\text{False Rejection Rate } (FRR) = \frac{FN}{TP+FN} \quad (7)$$

$$\text{Half Total Error Rate } (HTER) = \left(\frac{FAR+FRR}{2}\right) * 100 \quad (8)$$

(a) (b) (c) (d)

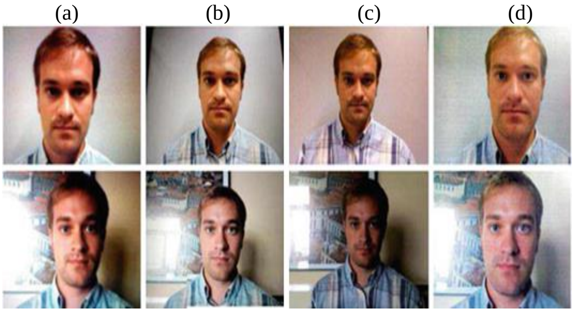

Figure 8. Candidate images taken from the Replay attack dataset [40] Top row represents the Controlled Scenario while the Last Row depicts Adverse Scenario. (a) LCD Photo Attack and (b) HD Photo Attack, (c) Real face samples, (d) Print Photo Attack

[1] https://www.kaggle.com

Figure 9. Candidate images taken from the NUAA dataset [41] Row one depicts real access face images, and row two gives fake access face images from the NUAA dataset

$$\text{False Acceptance Rate } (FAR) = \frac{FP}{TN+FP} \quad (9)$$

$$Accuracy = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) * 100 \quad (10)$$

where, TP => True Positives (Quantity of Live human faces predicted as Live), TN => True Negatives (Quantity of Spoofed human faces predicted as Spoofed). FP => False Positives (Quantity of Spoofed human faces predicted as Live). FN => False Negatives (Quantity of Live human faces predicted as Spoofed)

6. RESULTS AND DISCUSSION

Six of the current pre-trained deep learning CNN models, referred to here as VGG19, VGG16, DensNet121, Xception, MobileNet, and InceptionV3, are taken into consideration for performance evaluation and to ensure appropriateness in human FPAD. These models are changed for the human FPAD to achieve transfer learning. The investigated transfer learning performance is validated using the "NUAA" and "Replay-Attack" face spoofing detection datasets.

It is clear from Tables 4 and 5 that tuned VGG19 outperforms other DCNNs in terms of FPAD test accuracy. Deep learning characteristics are extracted from the second-to-last FC layer, which consists of 256 neurons, to improve the FPAD capability (highlighted in yellow in Figure 3).

Thepade's SBTC 10-ary features are attached with these automatically derived 256-dimension deep learning

TABLE 3. The count of human face images used from the datasets for evaluation of the proposed FPAD method

Face Class	Replay-Attack		NUAA	
	Training	Testing	Training	Testing
Live Face	900	1200	1743	3362
Spoofed Face	900	1200	1748	5761
	Spoofed by four sorts		Spoofed by only a sort	

TABLE 4. Percentage test accuracy of considered finetuned DCNN models for face liveness detection over the NUAA dataset

Finetuned DCNN model	Test Accuracy
VGG16	90.07
VGG19	**94.31**
MobileNet	78.18
Xception	71.6
InceptionV3	68
DenseNet121	70.9

features. The weighting of handcrafted and automatically derived features is displayed in Table 6 to provide a thorough knowledge of the feature combinations used in the current experiment. Different ML classifiers, including RandomForest, ExtraTree, SVM (kernel = "linear"), and RandomForest + ExtraTree + SVM ensemble, are trained for the FPAD task using these feature combinations.

On the NUAA dataset and using various types of features to train them. Table 7 compares the performance of several ML classifiers taken into account for FPAD. With a test accuracy of 94.52% for the NUAA dataset, it is noted that the RandomForest classifier trained with VGG19 + VGG19 + Thepade's SBTC 10-ary fused features outperforms all other investigated classifier and feature combinations.

According to Table 8, where the percentage test accuracy of considered ML classifiers trained over different features for face liveness detection on the Replay-Attack dataset is compiled, the SVM trained by VGG19 + VGG19 + Thepade's SBTC 10-ary fused features beats other classifiers for the Replay-Attack dataset by reaching 98.67% test accuracy. Tables 7 and 8 show that utilising VGG19 + VGG19 + Thepade's SBTC 10-ary fused features gives classifiers more FPAD capability compared to simply using deep learning features from VGG19.

Table 9 compares a few face presentation attack detection methods that have been suggested in the literature, with the best results found in the current study for two datasets known as NUAA and Replay-Attack. It is difficult to compare performance because each existing approach uses a different testing environment, performance measurements, and datasets.

TABLE 5. Percentage test accuracy of considered finetuned DCNN models for face liveness detection over Replay-Attack dataset

Finetuned DCNN model	Test Accuracy
VGG16	97.59
VGG19	**98.11**
MobileNet	94.54
Xception	86.8
InceptionV3	81.76
DenseNet121	78.97

TABLE 6. Dissection of handcrafted and auto-extracted features across considered different feature combinations

Sr No.	Feature combination	VGG19 feature	Thepade's SBTC 10-ary features	Total dimension of feature vector
1.	VGG19 feature	256 (1/1 part)	0	256
2.	Thepade's SBTC 10-ary features	0	30 (1/1 part)	30
3.	VGG19 + Thepade's SBTC 10-ary feature fusion	256 (1/2 part)	30 (1/2 part)	286
4.	VGG19 + VGG19 + Thepade’s SBTC 10-ary feature fusion	512 (2/3 part)	30 (1/3 part)	542

TABLE 7. Percentage test accuracy of considered ML classifiers trained over different features for face liveness detection on the NUAA dataset

ML Classifier	Thepade's SBTC 10-ary features	Finetuned VGG19 feature	VGG19 + Thepade's SBTC 10-ary feature fusion	VGG19 + VGG19 + Thepade’s SBTC 10-ary feature fusion	Average across all types of features
RandomForest	70.31	93.92	93.95	94.56	88.19
Extratree	70.59	93.46	93.49	94.25	87.95
SVM	62.54	93.25	94.16	94.40	86.09
RandomForest + ExtraTree + SVM	69.96	93.37	93.93	94.09	87.83
Average across all ML classifiers	68.35	93.50	93.88	94.33	--

TABLE 8. Percentage test accuracy of considered ML classifiers trained over different features for face liveness detection on the Replay-Attack dataset

ML Classifier	Thepade's SBTC 10-ary features	Finetuned VGG19 feature	VGG19 + Thepade's SBTC 10-ary feature fusion	VGG19 + VGG19 + Thepade's SBTC 10-ary Feature fusion	Average across all types of features
Randomforest	91.66	95.78	95.96	96.47	94.97
Extratree	87.48	95.35	95.57	95.66	93.51
SVM	87.18	95.61	98.54	**98.67**	**95.00**
RandomForest+ ExtraTree + SVM	89.55	95.44	95.87	96.47	94.33
Average across all ML classifiers	88.97	95.55	96.49	**96.82**	

TABLE 9. Comparison of the existing face liveness detection methods of literature with the proposed method

Face Anti-spoofing Method	Pre-trained DCNN model	Performance Metric	Dataset Explored	Testing Accuracy (%)	HTER (%)	EER (%)
Proposed method **VGG19 + VGG19 + Thepade's SBTC 10-ary + RandomForest**	VGG19	HTER, Test Accuracy	Replay-Attack	**98.67**	**1.35**	-
Proposed method **VGG19 + VGG19 + Thepade's SBTC 10-ary +SVM**	VGG19	HTER, Test Accuracy	NUAA	**94.56**	**4.71**	-
ResNet50 + RI-LBP [20]	ResNet-50	HTER, EER	Replay-Attack	-	2.6	2.3
			NUAA	-	-	0.5
			CASIA-FASD	-	-	4.4
			MSU-MFSD	-	-	3.1
VGG16 + LBP [23]	VGG16	Test Accuracy	Replay-Attack	75.25	-	-
			SSIJRI	92.05	-	-
			3DMAD	96.97	-	-
			Replay-Mobile	90.52	-	-
VGG16 [24]	VGG16	HTER	Replay-Attack	-	2.5	-
			CASIA-FASD	-	0.0	-
			Replay-Mobile	-	0.0	-
ResNet-50 + CLBP [29]	ResNet-50	HTER, Test Accuracy	CASIA-FASD	94.65	8.68	-
			Replay-Attack	98.56	2.64	

The HTER and testing accuracy noticed in the suggested work provided here using VGG19; however, are better than the regarded existing equivalent attempts from the literature [20, 23, 24, 29] when one compares the explorations carried out utilising datasets NUAA and Replay-Attack (as presented in Table 9 for comparison of existing methods with the proposed approach).

7. CONCLUSION

One of the more clever methods for adapting the learned DCNN architecture of an existing system to newer applications is transfer learning. Depending on its specifics, each architecture may represent a different performance for newer applications. In the ambient computing environment of today, where person identity is confirmed using the collected contactless biometric features, human FPAD has assumed paramount importance. This study empirically evaluated six such pre-trained DCNN models (Xception, InceptionV3, MobileNet, DenseNet121, VGG16 and VGG19) to detect person face presentation attacks. To adapt all pre-trained DCNN models for use with human FPAD, certain changes have to be made. Performance measures for contrasting the effectiveness of the proposed method with current human FPAD methodologies include test accuracy and HTER. According to experimental findings, among all fine-tuned DCNNs taken into account, the VGG19 network architecture provides the highest test accuracy for FPAD across the NUAA and Replay-Attack datasets. Additionally, compared to using simply the deep learning features of VGG19, VGG19 + VGG19 + Thepade's SBTC 10-ary fused features improve the FPAD capacity for all classifiers with slight additional computations of feature extraction. Future research on datasets like 3DMAD and others can test the robustness of the approach suggested in the current work against mask attacks.

8. REFERENCES

1. Kekre, H., Thepade, S.D. and Maloo, A., "Face recognition using texture features extracted from walshlet pyramid", ***ACEEE International Journal on Recent Trends in Engineering and Technology (IJRTET)***, Vol. 5, No. 1, (2011), 186-190. https://doi.org/10.5120/1672-2256
2. Kekre, H., Thepade, S.D. and Chopra, T., "Face and gender recognition using principal component analysis", ***International Journal on Computer Science and Engineering***, Vol. 2, No. 4, (2010), 959-964.oi.
3. Shahbakhsh, M.B. and Hassanpour, H., "Empowering face recognition methods using a gan-based single image super-resolution network", ***International Journal of Engineering, Transactions A: Basics***, Vol. 35, No. 10, (2022), 1858-1866. https://www.ije.ir/article_150976.html
4. Asghari Beirami, B. and Mokhtarzade, M., "Ensemble of log-euclidean kernel svm based on covariance descriptors of multiscale gabor features for face recognition", ***International Journal of Engineering, Transactions B: Applications***, Vol. 35, No. 11, (2022), 2065-2071. https://www.ije.ir/article_153711.html
5. Thepade, S.D. and Bidwai, P., "Iris recognition using fractional coefficients of transforms, wavelet transforms and hybrid wavelet transforms", in 2013 International Conference on Control, Computing, Communication and Materials (ICCCCM), IEEE., (2013), 1-5.
6. Kekre, H., Thepade, S.D., Jain, J. and Agrawal, N., "Iris recognition using texture features extracted from walshlet pyramid", in Proceedings of the International Conference & Workshop on Emerging Trends in Technology., (2011), 76-81.
7. Thepade, D.S. and Mandal, P.R., "Novel iris recognition technique using fractional energies of transformed iris images using haar and kekre transforms", ***International Journal of Scientific & Engineering Research***, Vol. 5, No. 4, (2014), 305-308. https://www.ijser.org/researchpaper/Novel-Iris-Recognition-Technique-using-Fractional-Energies.pdf
8. Khade, S. and Thepade, S.D., "Novel fingerprint liveness detection with fractional energy of cosine transformed fingerprint images and machine learning classifiers", in 2018 IEEE Punecon, IEEE., (2018), 1-7.
9. Khade, S., Thepade, S.D. and Ambedkar, A., "Fingerprint liveness detection using directional ridge frequency with machine learning classifiers", in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), IEEE., (2018), 1-5.
10. Kekre, H., Thepade, S.D. and Maloo, A., "Eigenvectors of covariance matrix using row mean and column mean sequences for face recognition", ***International Journal of Biometrics and Bioinformatics (IJBB)***, Vol. 4, No. 2, (2010), 42-50.
11. Keramati Hatkeposhti, R., Yadollahzadeh Tabari, M. and GolsorkhtabariAmiri, M., "Fall detection using deep learning algorithms and analysis of wearable sensor data by presenting a new sampling method", ***International Journal of Engineering, Transactions A: Basics***, Vol. 35, No. 10, (2022), 1941-1958. https://www.ije.ir/article_151530.html
12. Fallah, A., Soleymani, A. and Khosravi, H., "A method for automatic lane detection using a deep network", ***International Journal of Engineering, Transactions A: Basics***, Vol. 35, No. 4, (2022), 802-809. https://www.ije.ir/article_143621.html
13. Azimi, B., Rashno, A. and Fadaei, S., "Fully convolutional networks for fluid segmentation in retina images", in 2020 International Conference on Machine Vision and Image Processing (MVIP), IEEE., (2020), 1-7.
14. Thepade, S.D. and Chaudhari, P.R., "Land usage identification with fusion of thepade sbtc and sauvola thresholding features of aerial images using ensemble of machine learning algorithms", ***Applied Artificial Intelligence***, Vol. 35, No. 2, (2021), 154-170. https://doi.org/10.1080/08839514.2020.1842627
15. Thepade, S.D., Chaudhari, P.R. and Das, R., "Identifying land usage from aerial image using feature fusion of thepade's sorted n-ary block truncation coding and bernsen thresholding with ensemble methods", ***International Journal of Engineering and Advanced Technology (IJEAT)***, Vol. 9, No. 3, (2020), 2612-2621. doi.
16. Thepade, S.D., Subhedarpage, K.S. and Mali, A.A., "Performance rise in content based video retrieval using multi-level thepade's sorted ternary block truncation coding with intermediate block videos and even-odd videos", in 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE., (2013), 962-966.
17. Madane, M. and Thepade, S., "Score level fusion based bimodal biometric identification using thepade's sorted n-ary block truncation coding with variod proportions of iris and palmprint traits", ***Procedia Computer Science***, Vol. 79, (2016), 466-473. doi: 10.1016/j.procs.2016.03.060.
18. Thepade, S.D. and Patil, P.H., "Novel visual content summarization in videos using keyframe extraction with thepade's sorted ternary block truncation coding and assorted similarity measures", in 2015 International Conference on Communication, Information & Computing Technology (ICCICT), IEEE. (2015), 1-5.
19. Lucena, O., Junior, A., Moia, V., Souza, R., Valle, E. and Lotufo, R., "Transfer learning using convolutional neural networks for face anti-spoofing", in Image Analysis and Recognition: 14th International Conference, ICIAR 2017, Montreal, QC, Canada, July 5–7, 2017, Proceedings 14, Springer., (2017), 27-34.
20. Chen, F.M., Wen, C., Xie, K., Wen, F.Q., Sheng, G.Q. and Tang, X.G., "Face liveness detection: Fusing colour texture feature and deep feature", ***IET Biometrics***, Vol. 8, No. 6, (2019), 369-377. https://doi.org/10.1049/iet-bmt.2018.5235
21. Tang, Y., Wang, X., Jia, X. and Shen, L., "Fusing multiple deep features for face anti-spoofing", in Biometric Recognition: 13th Chinese Conference, CCBR 2018, Urumqi, China, August 11-12, 2018, Proceedings 13, Springer., (2018), 321-330.
22. Tu, X. and Fang, Y., "Ultra-deep neural network for face anti-spoofing", in Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II 24, Springer. (2017), 686-695.
23. Das, P.K., Hu, B., Liu, C., Cui, K., Ranjan, P. and Xiong, G., "A new approach for face anti-spoofing using handcrafted and deep network features", in 2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), IEEE., (2019), 33-38.

24. Elloumi, W., Chetouani, A., Charrada, T.B. and Fourati, E., "Anti-spoofing in face recognition: Deep learning and image quality assessment-based approaches", ***Deep Biometrics***, (2020), 51-69. https://doi.org/10.1007/978-3-030-32583-1_4
25. Song, L. and Ma, H., "Face liveliness detection based on texture and color features", in 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), IEEE., (2019), 418-422.
26. Jagdale, P. and Thepade, S., "Face liveness detection using feature fusion using block truncation code technique", ***International Journal on Recent and Innovation Trends in Computing and Communication***, Vol. 7, No. 8, (2019), 19-22. https://doi.org/10.17762/ijritcc.v7i8.5348
27. Thepade, S., Jagdale, P., Bhingurde, A. and Erandole, S., "Novel face liveness detection using fusion of features and machine learning classifiers", in 2020 IEEE international conference on informatics, IoT, and enabling technologies (ICIoT), IEEE. (2020), 141-145.
28. Thepade, S.D., Chaudhari, P., Dindorkar, M., Bang, S. and Bangar, R., "Improved face spoofing detection using random forest classifier with fusion of luminance chroma", ***International Journal of Computer Information Systems and Industrial Management Applications***, Vol. 12, No. 2020, (2020), 374-386.
29. Abdullakutty, F., Johnston, P. and Elyan, E., "Fusion methods for face presentation attack detection", ***Sensors***, Vol. 22, No. 14, (2022), 5196.
30. Muhammad, U., Yu, Z. and Komulainen, J., "Self-supervised 2d face presentation attack detection via temporal sequence sampling", ***Pattern Recognition Letters***, Vol. 156, (2022), 15-22.
31. Abdullakutty, F., Elyan, E., Johnston, P. and Ali-Gombe, A., "Deep transfer learning on the aggregated dataset for face presentation attack detection", ***Cognitive Computation***, Vol. 14, No. 6, (2022), 2223-2233. https://doi.org/10.1007/s12559-022-10037-z
32. Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, (2014).
33. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., "Rethinking the inception architecture for computer vision", in Proceedings of the IEEE conference on computer vision and pattern recognition., (2016), 2818-2826.
34. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., "Densely connected convolutional networks", in Proceedings of the IEEE conference on computer vision and pattern recognition., (2017), 4700-4708.
35. Chollet, F., "Xception: Deep learning with depthwise separable convolutions", in Proceedings of the IEEE conference on computer vision and pattern recognition., (2017), 1251-1258.
36. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., "Mobilenets: Efficient convolutional neural networks for mobile vision applications", arXiv preprint arXiv:1704.04861, (2017).
37. Badre, S.R. and Thepade, S.D., "Novel video content summarization using thepade's sorted n-ary block truncation coding", ***Procedia Computer Science***, Vol. 79, (2016), 474-482. https://doi.org/10.1016/j.procs.2016.03.061
38. Chingovska, I., Anjos, A. and Marcel, S., "On the effectiveness of local binary patterns in face anti-spoofing", in 2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG), IEEE., (2012), 1-7.
39. Tan, X., Li, Y., Liu, J. and Jiang, L., "Face liveness detection from a single image with sparse low rank bilinear discriminative model", ECCV (6), Vol. 6316, (2010), 504-517. https://doi.org/10.1007/978-3-642-15567-3_37

Persian Abstract

چکیده

تکامل و پیشرفت‌های یادگیری عمیق برای تمام مشکلات تحقیقاتی که ممکن است به چالش‌های طبقه‌بندی در هر زمینه‌ای از زندگی .تبدیل شوند، استفاده می‌شوند .عمق و مقدار فراپارامترها خروجی هر طراحی شبکه عصبی پیچیده عمیق (DCNN)را تعیین می‌کند، به همین دلیل است که تعداد زیادی از آنها پیشنهاد شده‌اند ...چنین معماری‌های DCNN باید کاملاً از ابتدا توسعه داده شوند و نمی‌توان از آنها برای برنامه‌هایی که پیشرفته‌تر از آن‌هایی که برای آن طراحی شده‌اند استفاده کرد. این شبکه های از پیش آموزش دیده را می توان از طریق آموزش انتقال تغییر داد تا برای کاربردهای خاص تر مناسب تر شوند. این مقاله تلاش می‌کند تا اثربخشی و کاربرد معماری‌های DCNN از پیش آموزش‌دیده فعلی را برای تشخیص حملات نمایش چهره انسانی (FPAD) ارزیابی کند. ا توجه به توسعه محاسبات محیطی، که نیاز به شناسایی بدون تماس افراد با استفاده از ویژگی‌های بیومتریک آنها دارد، FPAD انسانی یکی از مهم‌ترین و حیاتی‌ترین حوزه‌های تحقیقاتی امروزی است. برای ارزیابی تجربی در FPAD انسانی، شش مدل DCNN از پیش آموزش‌دیده - MobileNet، Xception، DensNet۱۲۱، VGG۱۶، VGG۱۹ و InceptionV۳ در نظر گرفته می‌شوند. دو مجموعه داده معیار FPAD، NUAA و Replay-Attack، برای کاوش استفاده می‌شوند. به منظور بهبود قابلیت FPAD طبقه‌بندی‌کننده‌های یادگیری ماشینی تحلیل‌شده (ML)، ویژگی‌های ۱۰ ساله Thepade's Sorted Block Truncation Coding (SBTC) با ویژگی‌های یادگیری عمیق مشتق شده از DCNNهای با بالاترین عملکرد تنظیم‌شده ترکیب شده‌اند. با توجه به ادغام ویژگی های DCNN و Thepade SBTC ۱۰-ary، دقت FPAD طبقه بندی کننده های ML به طور قابل توجهی بهبود یافته است.