Deepfake Detection Model Based on Combined Features Extracted from Facenet and PCA Techniques

Article · January 2023

CITATIONS
0

READS
140

2 authors:

Duha Amir
University of Mosul
6 PUBLICATIONS **9** CITATIONS
SEE PROFILE

Laheeb Mohammad Ibrahim
University of Mosul
82 PUBLICATIONS **433** CITATIONS
SEE PROFILE

Al-Rafidain Journal of Computer Sciences and Mathematics (RJCM)

www.csmj.mosuljournals.com

Deepfake Detection Model Based on Combined Features Extracted from Facenet and PCA Techniques

Duha Amir Sultan[1],* Laheeb Mohammad Ibrahim[2]

Department of Biology, Education College for Girls, University of Mosul, Mosul, IRAQ [1] , Department of Software Engineering, College of Computer Science and Mathematics, University of Mosul, Mosul, IRAQ [2]
**Corresponding author. Email: duhaasultan@uomosul.edu.iq[1]*

Article information

Article history:
Received: 24/3/2023
Accepted: 4/7/2023
Available online:

Abstract

Recently, the increase in the emergence of fake videos that have a high degree of accuracy makes it difficult to distinguish from real ones. This is due to the rapid development of deep-learning techniques, especially Generative Adversarial Networks (GAN). The harmful nature of deepfakes urges immediate action to improve the detection of such videos. In this work, we proposed a new model to detect deepfakes based on a hybrid approach for feature extraction by using 128-identity features obtained from facenet_CNN combined with most powerful 10-PCA features. All these features are extracted from cropped faces of 10 frames for each video. FaceForensics++ (FF++) dataset was used to train and test the model, which gave a maximum test accuracy of 0.83, precision of 0.824 and recall value of 0.849.

Keywords:
CNN, Deepfake Detection, Deep learning, Facenet, PCA.

Correspondence:
Author: Duha Amir Sultan
Email: duhaasultan@uomosul.edu.iq

I. INTRODUCTION

Deepfake refers to the manipulated digital media, such as images or videos, where the image or video of a person is replaced with another person's likeness. Deepfakes can be generated by a class of deeplearning models called GANs [1], were two neural networks (Generator and Discriminator) compete against each other. The generator generates fake content based on an existing dataset, while the Discriminator learns to identify the difference between the real and fake content. This style of work led to the emergence of fabricated content with a high degree of accuracy, so that it is difficult to distinguish it from the real one. Later many types of GANs appeared, like FCGAN, DCGAN, StyleGAN,…and so on, see [2] for more details about GAN variations. As a result, several applications have appeared in the field of deepfake creation and unfortunately spread through the Internet. FakeApp was the first one that allows users to exchange faces with another person. By the time, more similar applications have been created such as FaceSwap, DeepFaceLab, DFaker and many more which make it very easy for everyone to create fake videos (even if they don't have any knowledge about it). Although Deepfake was majorly used for entertainment purposes, but its falsify content can be harm depending on the user intent. That is why there is an urgent need for methods or techniques for detecting this fake content. This topic has received a great attention for researchers, many studies and researches published dealing with methods of detecting fabricated videos. The common factor between all of these researches is that, they rely on machine learning and deep-learning in their work.[3]

In this work, we proposed a hybrid method for extracting features based on facenet CNN and Principle Component Analysis (PCA) in order to increase model accuracy for detecting fake videos. The research is divided into six parts. Part1 is the introduction. Part2 lists most of the previous works that are related to the topic. The proposed model and the practical works are explained in details in parts 4 and 5 respectively. Finally, and in part6, a brief paragraph summarizing the conclusion deduced after the completion of the

work

2. Related Works:

In deepfake videos, there will undoubtedly be some mismatch or inconsistency within the overall video frames. Finding these inconsistent artifacts is the objective of the researchers in this field. Depending on the type of the extracted features, video detection methods can be classified into two categories:

-Those that utilize visual artifacts inside frames.

-Those that utilize temporal features to take full advantages of the relationship among the multiple frames.

2.1. Visual Artifacts within a frame:

After doing the required preprocessing, such as Viola_Jones [4] to extract face region within each frame, Afchar D. et.al.[5] submitted a method built on two convolutional neural networks, name as Meso-4 and MesoInception-4 to extract features. These networks were trained on two datasets: FaceForensics (FF), with another dataset contains real and fake videos with the same resolution collected from the internet. The model achieved high detection rate for deepfake and Face2Face datasets. Another architecture introduced by Aya I. et.al.[6] named as YOLO-CNN-XGBoost that integrated the benefits of both XGBoost[7] and CNN. YOLO face detector[8] was used to determine faces in video frames, while InceptionResNetV2 model was used to extract the discriminant spatial-visual features. These features are then fed to the XGBoost classifier to distinguish between the real and fake videos. A method using optical flow vectors derived for two successive frames was presented by Amerini et al. [9]. ResNet and VGG16 models were used to extract features. The model's accuracy for ResNet50 and VGG16 was 81.61% and 75.46%, respectively, using the Face Forensics (FF++) dataset. Tran V.N. et al. [10] presented an architecture based on classifier network with manual attention target-specific regions to form distillation in order to enable the use of light model along with increasing the classification accuracy. Multitask cascaded convolutional neural networks (MTCNN)[11] used to detect the face within each frame, then facial landmarks were calculated for each face. The distillation sets are created, which contain a number of patches, to specify which areas of the face would be trained. So, the distillation sets worked as the classifier's input. Inception v3[12] was utilized to train the complete face, whereas MobileNet [13] was used to train individual face patches. Celebrities Deepfake Forensic_v2 (Celeb DFv2)[14] and DeepFake Detection Challenge (DFDC)[15] datasets are used to evaluate the model's performance, and the results showed high evaluation accuracy. Matern F.et. al.[16] concentrated in their works on eyes region . The features used by the researchers are : 1- Both eyes should have equal radii and colors. 2- The distance between the iris and the eye centers for the right and left eyes should be the same. 3- The eyes and teeth regions lack details and reflection. Before feature extraction, Hough circle transform and Canny edge detection were adopted to identify the eye region. For classification, logistic regression and neural networks were utilized. The authors established that classifying using the features of the eyes and teeth together produced better results than using the features alone. Deressa W. et.al.[17], proposed a model consisting of a CNN_Vision Transformer (CViT). VGG_like CNN was used to extract features, while the ViT was used for classification. DFDC dataset used to train and evaluate the model.

2.2. Temporal features across multiple frames:

To benefit the temporal relationships between successive frames and to increase accuracy, a two-stage deep learning model presented by Guera D. et al. [18] that combines CNN and LSTM networks to take advantage of CNN's feature extraction capabilities and LSTM's classification and memorizing abilities. 300 real videos were taken from the Hollywood Human Actions (HOHA) dataset [19], and 300 deepfake videos were acquired from various websites formed the dataset used by the model.

[20,21] presented similar works. A mixed model with ResNeXt[22] and an LSTM neural network was utilized by Abdul Jamsheed V. et al. [20]. DFDC, FF++, and Celeb DF datasets were combined to create the dataset that was used in this study. While Priti Y. et.al.[21] built a model using InceptionResNet v2 for feature extraction, and its output was used as an input to 2048 LSTM layer. The model was trained on dataset that has been collected from DFDC dataset for 20 and 40 epochs respectively. .

The eye area is transformed into discriminative features in Li Y. et al.'s research paper [23] by using a VGG16_CNN framework [24]. the output of VGG16 is received by LSTM. Closed Eyes in the Wild (CEW) dataset [25], which contains 1193/1232 images of closed and open eyes respectively, was used to evaluate the model.

Daichi Z. et al.'s [26] used Temporal Dropout 3-Dimensional Convolution Neural Network (TD-3DCNN) to fully utilize spatiotemporal information. Video frames volumes were sampled by temporal dropout operation and fed into a 3DCNN. Six detectors used for comparison, the model outperformed them and achieved a competitive performance on FF++, DFDC, and Celeb-DFv2 datasets.

A more generalized method was submitted by Yipin Z. and Ser-Nam L.[27] They discovered that when people speak, there is a significant correlation between the pronounced syllables and the lip motion. This led them to present a combined visual/auditory deepfake detection model. The model can be used with videos in various languages because it is language-neutral.

In Shruti A. et. al.[28], the researchers concluded that the shape of the mouth should be totally closed when pronouncing words with the sound M, B, or P. so, they focused of the mouth shape using three analysis approaches:1- manual (by analyst), 2- profile, and 3- CNN: where Xception architecture was used for classification.

3. Proposed model:

In this work, a hybrid method was suggested for extracting features to improve model ability to detect fake videos. The complete pipeline for the manipulation of deepfake video

detection can be divided into three major phases: 1-Preprocessing, 2-Feature extraction, and 3-Classification. Below Figure 1 expresses the steps of the overall process.

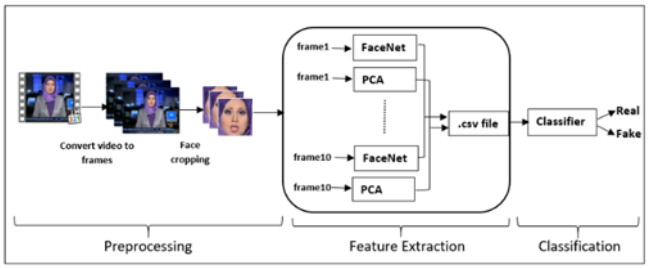

Figure 1. Detection pipeline of deepfake videos

3.1. Preprocessing:

It has been observed obviously in Figure 1 that the videos that need to be manipulated before being entered into the phase of feature extraction. Firstly, each video is converted into a number of frames where 10 successive frames for each video was extracted. Then , face region within each frame was detected and cropped using MTCNN. It is an effective algorithm for face detection, but it detects all faces and semi-faces found in the frame background. Avoiding this , only the face region with the largest area was taken, other detected faces were discarded, i.e. for each frame only one face was extracted and saved in a train or test folder for both real and fake videos for later use.

3.2. Feature Extraction:

The second step in the conducted method was to extract features from the cropped faces. Two algorithms were combined and used: 1-FaceNet [29]; which is a deep convolutional neural network takes a face image as input and outputs a vector of 128 numbers which represent the most important features of a face. This vector is called face embedding vector, (the embedding vector as a summary of the face can be thought of).In this research, the facenet used based on a variation of ResNet (ResNet-34). 2-Principle Component Analysis (PCA)[30]; a common method for feature extraction used to obtain the strongest and the best 10 features from each cropped face. The features extracted from these two methods were combined and saved in one feature vector.

These processes were applied on 10 frames for each video, specifically, 10 feature vectors were obtained for each video. To get only one feature vector per video, mean and standard deviation(std) values were calculated. This was done on the corresponding values of these 10-feature vectors associated with a specified video. As a result, each video was represented by a vector of 276 values. These values are 128 mean and 128 std for embedding vectors plus 10 mean and 10 std calculated for PCA features. All these values were saved in a .csv file.

3.3. Classification:

The final step is the classification process where a video is classified as a fake or real depending on the extracted features explained in step 3.2. A set of fully connected layers with dropout layers with a Sigmoid function used in the last layer form the structure of the classifier. Fig.2 represents the three major phases of the model with the clarification of classifier layers.

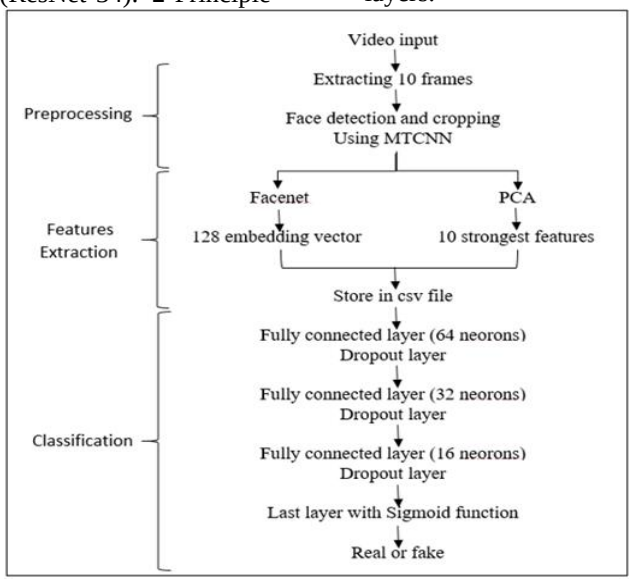

Figure 2. Practical work steps

4. Practical Work:

In this section, a description of the results of the training and testing processes of the proposed model on the selected datasets was conducted. The metrics used to evaluate the model. in Figure 2 shows the flowchart of the work steps, starting from entering a video until the video is classified using Sigmoid function.

4.1. Dataset:

The dataset used to train the model is FF++[31] comprises of 1000 real and 1000 fake videos. The fake videos were created using three from its four manipulation techniques: Deepfake, Face2Face, FaceSwap, and Neural Textures. By now, dataset contains 1000 fake videos for each manipulation technique.

4.2. Evaluation Metrics:

The model was evaluated and tested using the following three metrics, accuracy, precision, and recall [32]. These metrics were calculated through the confusion matrix after testing the model. eq.1 to eq.3 that represent the equations of the approved evaluation metrics:

Accuracy is the percentage of the correctly predicted classes.

$$\text{Accuracy}=\frac{\text{no.of correct predictions}}{\text{total no.of predictions}}=\frac{TP+TN}{TP+TN+FP+FN} \quad \ldots(1)$$

$$\text{Precision}=\frac{TP}{TP+FP} \quad \ldots(2)$$

$$\text{Recall}=\frac{TP}{TP+FN} \quad \ldots(3)$$

Where: TP : correctly predicts real videos
TN: correctly predicts fake videos
FP: incorrectly predicts real videos
FN: incorrectly predicts fake videos

These values can be obtained from the confusion matrix after testing the model , as evident from Figure 3.

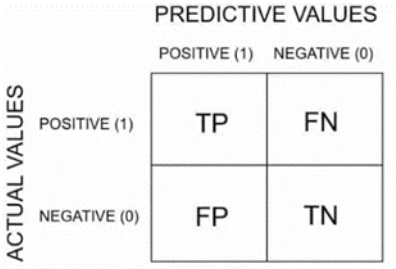

Figure 3. Confusion matrix for binary classification

5. Results:

Most of the previous works were relied on the use of one of the CNN networks pretrained on 'imagenet', whether using CNN only to extract visual features or CNN with LSTM to extract both visual and temporal features). But applying these methods on FF++ dataset, the obtained test accuracy was very poor. It did not exceed 0.56%, while the proposed model has achieved very better results as shown in Table1 where each of the two adopted methods, facenet and PCA, was first implemented separately, then combined and ended with a comparison between their results. Initially, a facenet was applied only to one frame (the first frame in each video), obtaining the features vector that contain the facial features present in the respective frame. Then, these features sent to the classifier. After making a prediction on this model, a test accuracy was somewhat acceptable, but not the best.

Secondly, the same facenet was applied, but on 10 consecutive frames for each video (10 feature vectors were obtained). So, to obtain one feature vector for each video, the values of these ten vectors were aggregated by using mean and std python functions. In this case, the test accuracy increased by approximately 0.05-0.1. At this point of the study, the same steps were applied using PCA, and the test accuracy is given in the table.

To enhance the extracted features and to increase the test accuracy, the extracted features from PCA were combined with the feature vector from the facenet. An increase in the test accuracy by an amount of 0.02-0.03 was noticed clearly. All these processes were applied to the FF++ dataset with the three techniques to produce faked videos: F2F, DF, FS. Also, the highest test accuracy obtained after training the model with the faked dataset DF.

Table 1: Results of the proposed model in a form of test accuracy compared with its elementary methods

		Facenet		**PCA**		**Facenet +PCA**
		Single frame	10 Frames	Single frame	10 frames	10 frames
accuracy	F2F	0.71	0.76	0.51	0.5477	0.79
	FS	0.59	0.766	0.53	0.53	0.74
	DF	0.68	0.766	0.62	0.588	0.814
precision	F2F	0.71	0.77	0.53	0.55	0.79
	FS	0.6	0.76	0.53	0.55	0.77
	DF	0.72	0.775	0.62	0.595	0.815
recall	F2F	0.71	0.765	0.51	0.55	0.79
	FS	0.59	0.765	0.53	0.53	0.74
	DF	0.74	0.765	0.62	0.59	0.815

For classification, a classifier with a number of dense layers of 64, 32 and 16 and 1 neurons were used as shown in Figure 2, with relu activation function for each layer, except the last layer where a sigmoid function was used. The model has been trained using binary crossentropy loss function for a total of 30 epochs. Adam optimizer is used for optimization. Figure 4 represents the training process with the training and validation loss and accuracy

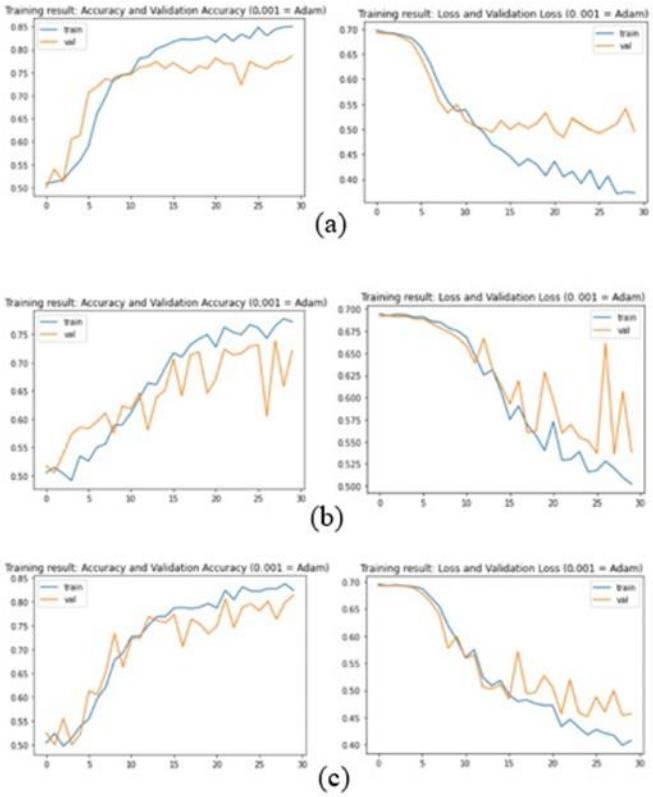

Figure 4. Training accuracy and loss of the after training on: (a) F2F (b) FS (c) DF datasets

The model was tested on 400 real and fake videos achieved a maximum accuracy of 0.789 and 0.79 on both F2F and DF fake datasets. Figure 5 shows the confusion matrices contents after testing the model through the prediction process.

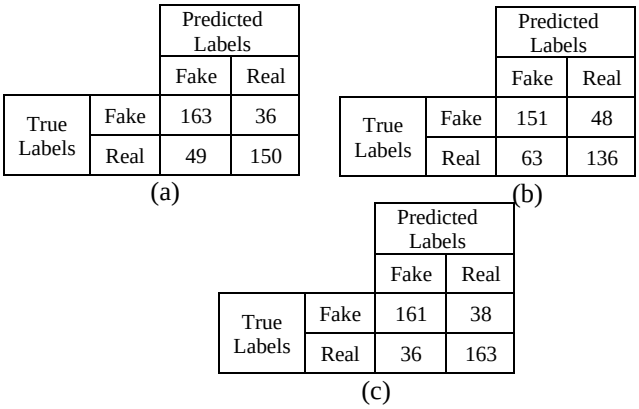

Figure 5. Confusion matrix obtained after testing the model using: (a) F2F (b)FS (c)DF datasets

Better accuracy results were obtained through different machine learning classifiers as shown in (fig.6) by applying a LazyClassifier provided in lazy predict python library. In (fig.6) all the used classifiers with the calculated accuracies arranged in descending order were enlisted. Highest accuracy 0f 83% with LDA, 80% with NuSVC, and 76% with LDA were obtained once tested using DF, FS, and F2F fake datasets respectively. After making a prediction on the test data using the model with the highest accuracy, the values of precision and recall were 0.824 and 0.849 respectively.

Table 2: comparison of the proposed model with equivalent ones

reference	**model**	**accuracy**
Self-experiment	ResNet50	0.56
Chandani et al.[33]	ResNet-152	0.767
Mittal et al.[34]	AlexNet	0.556
Sven van A. [35]	Exception	0.707
Afchar et al.[5]	MesoInception4	0.813
Proposed model	Facenet+Pca	0.83

In Table 2, the proposed model was compared with five different CNN models. The comparison was done using the value of test accuracy of the six models. The test accuracy of the first model obtained from the self-practical experience of the ResNet CNN after training on the *imagenet* dataset. Also, it was noted that the first two models were both ResNet but with different depths. The proposed model (using ResNet-34) gave better results than the previous ones despite being less in depth.

Additionally, the table contains the results obtained from the works of [34],[35],[36] and [5]. The superiority of the proposed model over the others was noted, referring that mixing more methods for extracting features gives better results. Worst result was gained from *'imagenet'* pretrained ResNet, indicates that the weights adopted in it did not serve the task of distinguishing fake faces from the real ones.

Classifier	Acc.
LinearDiscriminantAnalysis	0.83
RidgeClassifier	0.80
RidgeClassifierCV	0.80
CalibratedClassifierCV	0.79
NuSVC	0.79
LogisticRegression	0.79
LinearSVC	0.78
SVC	0.77
XGBClassifier	0.76
SGDClassifier	0.76
LGBMClassifier	0.75
Perceptron	0.75
GaussianNB	0.72
PassiveAggressiveClassifier	0.72
NearestCentroid	0.71
RandomForestClassifier	0.70
AdaBoostClassifier	0.69
ExtraTreesClassifier	0.69
BernoulliNB	0.68
BaggingClassifier	0.67
KNeighborsClassifier	0.65
QuadraticDiscriminantAnalysis	0.64
DecisionTreeClassifier	0.62
ExtraTreeClassifier	0.54
LabelSpreading	0.50
LabelPropagation	0.50
DummyClassifier	0.50

(a)

Classifier	Acc.
NuSVC	0.80
SVC	0.77
LinearSVC	0.76
RidgeClassifierCV	0.75
CalibratedClassifierCV	0.74
LinearDiscriminantAnalysis	0.74
LogisticRegression	0.74
RidgeClassifier	0.73
Perceptron	0.72
QuadraticDiscriminantAnalysis	0.71
PassiveAggressiveClassifier	0.71
LGBMClassifier	0.71
GaussianNB	0.70
SGDClassifier	0.69
XGBClassifier	0.69
AdaBoostClassifier	0.68
BernoulliNB	0.66
RandomForestClassifier	0.66
ExtraTreesClassifier	0.66
BaggingClassifier	0.65
NearestCentroid	0.64
DecisionTreeClassifier	0.58
KNeighborsClassifier	0.58
ExtraTreeClassifier	0.52
LabelSpreading	0.50
LabelPropagation	0.50
DummyClassifier	0.50

(b)

Classifier	Acc.
LinearDiscriminantAnalysis	0.76
SVC	0.75
NuSVC	0.75
LinearSVC	0.74
CalibratedClassifierCV	0.74
RidgeClassifier	0.74
LogisticRegression	0.74
RidgeClassifierCV	0.74
LGBMClassifier	0.73
XGBClassifier	0.72
AdaBoostClassifier	0.71
PassiveAggressiveClassifier	0.71
RandomForestClassifier	0.71
Perceptron	0.71
GaussianNB	0.71
NearestCentroid	0.70
SGDClassifier	0.69
BaggingClassifier	0.69
BernoulliNB	0.68
ExtraTreesClassifier	0.66
QuadraticDiscriminantAnalysis	0.60
KNeighborsClassifier	0.59
DecisionTreeClassifier	0.55
ExtraTreeClassifier	0.52
LabelPropagation	0.50
DummyClassifier	0.50
LabelSpreading	0.50

(c)

Figure 6. Test accuracy by using different classifiers after testing the model on (a) Deepfake (b) FaceSwap and (c) Face2Face fake datasets

Conclusion:

At this point, it can be concluded clearly that Fake contents in videos is increasing in social media and there is a crucial need for detecting such contents. Features extraction process is an influential factor in the success of such newly designed models. For this reason, focusing on this topic showed that merging two methods for extracting features can give better results for detecting fake videos than using each method alone. The combination was done between the facenet, which is a deep_CNN used originally

for face recognition, and a ML method, named as PCA, used for feature extraction and dimensionality reduction. The proposed method showed its superiority when compared with a number of typical and well-known CNN models.

References:

[1] Goodfellow I., Pouget_Abadie J., Mirza M., Xu B., Warde_Farly D., Ozair S., CourvilleA., and Bengio Y., (2014) "Generative Adversarial Mets", In Advances in Neural Information Processing Systems, pp. 2672-2680.

[2] Sultan, D. A. and Ibrahim, L. M. (2022). A Comprehensive Survey on Deepfake Detection Techniques. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), pp. 189–202.

[3] Shahzad HF, Rustam F, Flores ES, Luís Vidal Mazón J, de la Torre Diez I, Ashraf I. (2022). Review of Image Processing Techniques for Deepfakes. Sensors (Basel). 16; 22(12):4556. doi: 10.3390/s22124556. PMID: 35746333; PMCID: PMC9230855.

[4] Viola P. and Jones M. (2021). Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, volume 1, pages I–I.IEEE.

[5] Afchar, D., Nozick, V., Yamagishi, J., and Echizen, I. (2018). MesoNet: a compact facial video forgery detection network. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS) pp. 1-7 . IEEE.

[6] Aya I., Marwa E., Mervat S. Z., and Kamal E., (2021). A new deep learning-based methodology for video deepfake detection using Xgboost. Sensors,21, 5413.

[7] Chen T., Guestrin C. (2016), Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785-794.

[8] Joseph R., Santosh D., Ross G., and Ali F.(2016). You Only Look Once: Unified, Real-Time Object Detection. In proceedings of the IEEE Conference in Computer Vision and Pattern Recognition, pp 779-788.

[9] Amerini I., Galteri L., Caldelli R., Del Bimbo A.(2019) Deepfake Video Detection through Optical Flow Based CNN Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).

[10] Tran V.N, Lee S.H, Le H.S, Kwon K.R. (2021) High Performance DeepFake Video Detection on CNN-Based with Attention Target-Specific Regions and Manual Distillation Extraction.. Applied Sciences. 11(16):7678. https://doi.org/10.3390/app11167678.

[11] Zhang K., Zhang Z., Li Z., Qiao Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process. Lett. 23, pp. 1499–1503.

[12] Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–2 July 2016; pp. 2818–2826.

[13] Howard A.G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Adam H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861.

[14] Li, Y.; Yang, X.; Sun, P.; Qi, H.; Lyu, S.C.D. (2020). A large-scale challenging dataset for DeepFake forensics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19; pp. 14–19.

[15] Dolhansky, B.; Bitton, J.; Pflaum, B.; Lu, J.; Howes, R.; Wang, M.; Canton Ferrer, C. (2020). The deepfake detection challenge dataset. arXiv 2020, arXiv:2006.07397.

[16] Matern F., Riess C., and Stamminger M. (2019, January). Exploiting visual artifacts to expose deepfakes and face manipulations. In 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW) pp. 83-92. IEEE.

[17] Deressa W.,and Solomon A., (2021). Deepfake Video Detection Using Convolutional Vision Transformer. Computer Vision and Pattern Recognition, arXiv:2102.11126v3.

[18] Guera D., and Delp E. J. (2018, November). Deepfake video detection using recurrent neural networks. In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1-6. IEEE.

[19] Laptev I., Marszalek M., Schmid C., and Rozenfeld B. (2008). Learning realistic human actions from movies. Proceedings of the IEEE Conference on computer Vision and Pattern Recognition, pp. 1–8, June 2008. Anchorage, AK.

[20] Abdul Jamsheed V., and Janet B. (2021). Deep fake video detection using recurrent neural networks. International Journal of Scientific Research in Computer Science and Engineering. Vol.9(2), pp 22-26.

[21] Priti Y., Ishani J., Jaiprakash M., Vibhash C. and Gargi Kh., (2021). International Conference on Emerging Technologies: AI, IoT, and CPS for Science Technology Applications, 06-07.

[22] Saining X., Ross G., Piotr D., Zhuowen T., Kaiming H. (2017). Aggregated Residual Transformations for Deep Neural Networks. https://arxiv.org/abs/1611.05431v2.

[23] Li, Y., Chang, M. C., and Lyu, S. (2018, December). Inictu oculi: Exposing AI created fake videos by detecting eyeblinking. In 2018 IEEE International Workshop on InformationForensics and Security (WIFS), pp. 1-7. IEEE.

[24] Simonyan K. and Zisserman A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[25] Song F., Tan X., Liu X., and Chen S. (2014). Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. Pattern Recognition, vol. 47, no. 9, pp. 2825–2838.

[26] Daichi Z., Chenyo L., Fanzhao L., Dan Z., and Shiming G. (2021). Proceeding of the Thirtieth International Joint Conference on Artifitial Intelligence (IJCAI).

[27] Yipin Z. and Ser-Nam L., (2021). Joint Audio-Visual Deepfake Detection.IEEE/CVF International Conference on Computer Vision (ICCV), pp. 14780-14789, doi: 10.1109/ICCV48922.2021.01453

[28] Shruti A., Hany F., Ohad F.and Maneesh A., (2020). Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). DOI 10.1109/CVPRW50498.2020.00338.

[29] Florian S., Dmitry K., and James P., (2015) FaceNet: A Unified Embedding for Face Recognition and Clustering. arXiv:1503.03832v3 [cs.CV] 17 Jun 2015.

[30] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), pp. 417–441.

[31] https://doi.org/10.1037/h0071325 .

[32] Rossler A., Cozzolino D., Verdoliva L., Riess C., Thies J., Niebner M., (2019). Faceforensics++:Learning to detect manipulated facial images. In proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1-11.

[33] Hossin, M.1, Sulaiman, M.N. (2015). A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.2. pp. 1-11.

[34] Chandani, K., Arora, M. (2021). Automatic facial forgery detection using deep neural networks. In Advances in Interdisciplinary Engineering, (Springer), pp. 205–214.

[35] Mittal, H., Saraswat, M., Bansal, J.C. & Nagar, A. (2020). Fake-face image classifcation using improved quantum-inspired evolutionary-based feature selection method. in 2020 IEEE Symposium Series on Computational Intelligence (SSCI). Pp.989–995.

[36] Sven van A. (2021). Deepfake video detection using deep convolutional and hand-crafted facial features with long short-term memory network. M.Sc. thesis. Tilburg university, Tilburg, Netherlands.

نموذج اكتشاف التزييف العميق على أساس الميزات المجمعة المستخرجة من تقنيات facenet و pca

ضحى عامر سلطان
قسم علوم الحياة , كلية التربية للبنات , جامعة الموصل , الموصل , العراق
duhaasultan@uomosul.edu.iq

لهيب محمد ابراهيم
قسم هندسة البرمجيات , كلية علوم الحاسوب والرياضيات , جامعة الموصل , الموصل , العراق
laheeb_alzubaidy321966@uomosul.edu.iq

تاريخ الاستلام:24/3/2023 تاريخ القبول:4/7/2023

الملخص

نظرًا للتطور السريع لتقنيات التعلم العميق ، وخاصة شبكات الخصومة التوليدية (GAN)، فقد أدى ذلك إلى ظهور مقاطع فيديو مزيفة بدرجة عالية من الدقة ، بحيث يصعب تمييزها عن تلك الحقيقية منها. الطبيعة الضارة لملفات التزييف العميق قادت الى ضرورة اتخاذ إجراءات فورية لتحسين اكتشاف مقاطع الفيديو هذه. في هذا العمل ، اقترحنا نموذجًا جديدًا لاكتشاف التزييف العميق استنادًا إلى نهج هجين لاستخراج الميزات باستخدام identity 128 features والتي تم الحصول عليها من الشبكة العصبية التلافيفية facenet جنبًا إلى جنب مع أقوى 10ميزات مستخلصة عن طريق PCA .يتم استخلاص كل هذه الميزات من صور الوجوه المقتطعة من 10 إطارات لكل فيديو. تم استخدام مجموعة بيانات FF ++ لتدريب النموذج واختباره ، مما أعطى دقة اختبار قصوى تبلغ precision=0.824, 0.83 و recall=0.849.

الكلمات المفتاحية: الشبكة العصبية التلافيفية, كشف التزييف العميق, التعلم العميق, شبكة facenet, خوارزمية pc