An Efficient Deepfake Video Detection Approach with Combination of EfficientNet and Xception Models Using Deep Learning

Serhat ATAŞ, İsmail İLHAN and Mehmet KARAKÖSE

2022 26th International Conference on Information Technology (IT) | 978-1-6654-2127-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/IT54280.2022.9743542

Abstract— **Artificial intelligence is used in many areas and is constantly being developed. In recent years, videos made with deep fakes, which are often heard, have also developed. The use of videos made with deep fakes as blackmail in people's lives, manipulating the videos of important people to cause anxiety on people and etc. due to the fact that it poses a threat in many areas presents a big problem today. Efforts are being made to prevent this threat by detecting deep fake videos. Deep fake detection is still not fully resolved. For this reason, prominent technology companies provide support to researchers in this field and develop deep fraud detection by suggesting methods and organizing contests on most platforms such as Kaggle. In this article, a detection method is proposed to minimize the current concern of deep forgery. In the proposed method, the Xception model with high performance and speed and the EfficientNetB4 model with high accuracy were used. The proposed method aims to achieve better results and improvements in detecting fake videos.**

I. Introduction

Advances in deep learning have also led to growth in the production of deep fake videos. With the help of deep learning architectures such as GANs (Generative Adversarial Networks) and automatic encoders, and with enough footage of a target subject, believable fake videos can be easily created [1]. Three main methods such as face swapping, head puppetry and lip syncing are used in fake videos made with deepfake. As these fake videos became very realistic over time and left Deepfake Detection far behind, many technology companies came together to advance on the issue of deep fake detection. It has become convenient and useful for researchers that these technology companies come together and with the opportunities they provide to users, researchers can work in this field and easily access the proposed methods. Hsu et al. They carried out a deep learning-based study to detect fake images using contrast loss in gan-based samples [2]. Liang et al. summarized and analyzed the principles of deep forgery generation and detection methods on different types of forgery samples and datasets [3]. Yu et al. They explained in detail the fake image creation techniques. They compared the studies in this field by explaining the fraud detection methods, and then compared the data sets and discussed most of the issues in the literature [4]. Groh et al. made an evaluation in this area by comparing the effects of machine and human on deep fraud detection [5]. Guhagarkar et al. described and analyzed many studies on deep forgery [6]. Wodajo et al. They obtained experimental results on different datasets by describing the methods of deep forgery [7]. Ismail et al. They proposed a new method by comparing the face removal models and detection models by giving the information in the literature about deep forgery and by conducting studies on fraud detection on CELEB-DF and FaceForensics++ datasets [8]. Xia et al. they explained what deep forgery is, which technologies are used in deep forgery, and deep forgery detection [9]. Yasrab et al. They analyzed the upper body language of a person in a given image and performed the detection process on fake images. In their study, they used two neural networks as DNN (Deep Neural Network) and RNN (Recursive Neural Network) [10]. Samar Samir Khalil et al. They improved exposure estimation using HRNet and demonstrated fraud detection by proving that the YOLO v3 facial recognition model was better than the MTCNN facial recognition model [11]. Davide et al. They analyzed different solutions based on combinations of convolutional networks, especially EfficientNetB0 with different Video Transformers and compared their results with the latest available methods [12]. Ismail et al. examined the methods used to detect deep fake videos. They made performance impact analyzes of the applications. They have given data sets with different characteristics, detection applications with different methods and their properties in tables [13].

Detection of deep fake videos can be done in many ways, for example blink detection, head position detection, body position detection, inconsistent noise or sound detection, unnatural coloring, etc. There are many detection methods. In this article, we will describe the proposed method and convey the experimental results we have obtained in the detection of fake videos and the innovative aspect we have achieved.

The remaining of the manuscript is outlined as follows. Section 2, we try to target face detection and detection with two different models using trained models. The architecture and operation of the proposed method are explained. The stages and models of the method are explained. In section 3, we use DFDC, CELEB-DF, FaceForensics and our custom datasets in our work. The test results of the proposed method are evaluated. Finally in section 4, a general evaluation of the application was made.

978-1-6654-2127-0/22/$31.00 ©2022 IEEE

II. PROPOSED APPROACH

The proposed method will be done using CNN (Convolutional Neural Network). In our study, face extraction is performed with the Blazeface model. First of all, face extraction is done in the application, then detection is done on the features extracted with the Xception model because the Xception model is fast. If the video is fake in the first result, we will extract the features of the faces obtained with Blazeface by giving them to the EfficientNetB4 model. We will terminate it and there will be no need for re-detection. The block diagram of the proposed method is given in Fig. 1.

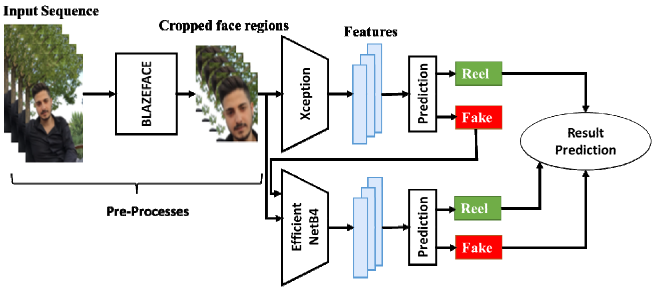

Figure 1. Block diagram of our proposed Deepfake Detection system: Faces in the video read with the BlazeFace method are extracted and we detect forgery only with Xception at first. As a result of the fraud detection we made with Xception, the program terminates if the image is real, but if the image is fake, the removed faces are loaded into the EfficientNetB4 model and the program results as a result of the estimation from the EfficientNetB4 model.

The reason why we use the EfficientNetB4 model in our study, and especially in fake videos, is that the EfficientNetB4 model has higher accuracy rates in fake videos than Xception in the tests we made with the special dataset we created, as well as the CELEB-DF, DFDC, FaceForensics datasets. The reason for using the Xception model in the first place is because it is faster than the EfficientNetB4 model and increases the performance of the program. The Xception model, on the other hand, has higher accuracy rates in real videos compared to the EfficientNetB4 model as a result of the tests. For this reason, if the videos are real, we will not need to re-detect them. As a result of the tests, we saw that we achieved higher accuracy rates when we equalized the number of frames (fps) to be extracted from each video to 29 before detecting. The stages of our study are given in Fig. 2.

A. Pre-Processes

In the proposed method, frames are obtained from the video by processing the video, and the face extraction process from the frames is done with the face extraction model. We use the BlazeFace model as the face extraction model. BlazeFace is a very fast and effective face removal model produced by Google that can also be used on phones. Amit Kumar et al. have proven the speed and accuracy of BlazeFace with their work [14].

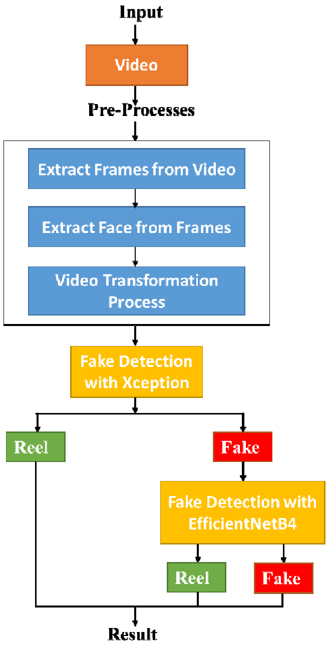
Input
Video
Pre-Processes
Extract Frames from Video
Extract Face from Frames
Video Transformation Process
Fake Detection with Xception
Reel
Fake
Fake Detection with EfficientNetB4
Reel
Fake
Result

Figure 2. Recommended Method Stages

B. Xception and EfficientNetB4 with Fake Detection

Two models, EfficientNetB4 and Xception, are used in application. Because the Xception model has fast and high accuracy rates, it is the model we use for the first detection in the method. Since the Xception model is better than the EfficientNetB4 model in real videos, it has increased the accuracy rate for real videos by using the Xception model on real videos. Yuval et al. they have proven the accuracy and speed of the Xception model with their work [15]. The architecture of the Xception model is given in Fig. 3 [16].

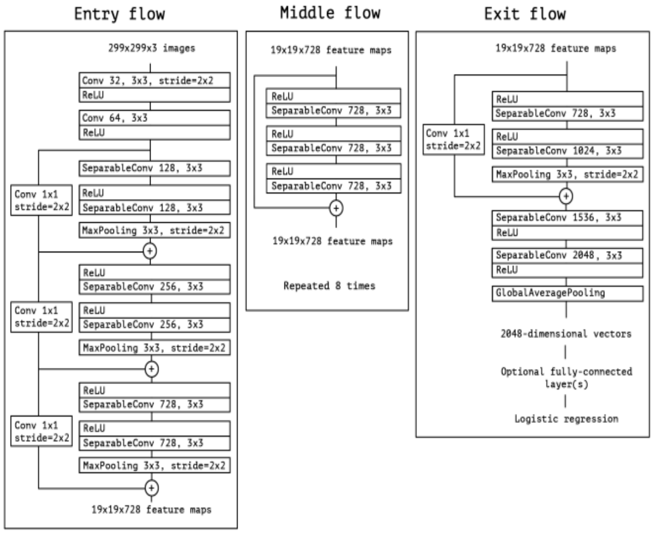

Figure 3. Architecture of Xception Model [16]

Since the EfficientNetB4 model has better accuracy rates in fake videos than the Xception model, if the videos that pass the Xception model's control are fake, forgery will be detected again with EfficientNetB4, and we have increased the accuracy rate on fake videos with this process. Ismail A et al. they have proven that EfficientNet models have high accuracy rates [8]. The architecture of the EfficientNetB4 model is given in Figure 4 [17].

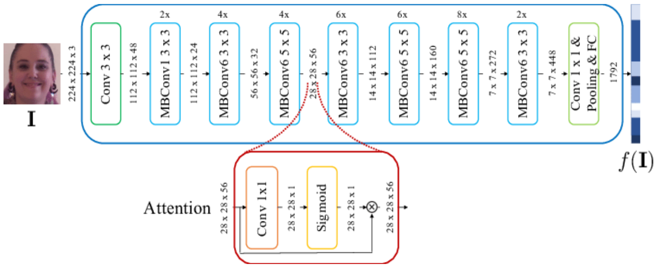

Figure 4. Architecture of EfficientNetB4 Model [17]

III. Experimental Results

The EfficientNetB4 and Xception models used in the proposed method were pre-trained on the DFDC dataset. Tests were conducted with our special dataset consisting of 20 videos, the CELEB-DF dataset, the FaceForensics dataset, the dataset in the DeepFake Detection Challenge competition on Kaggle. The images in our special dataset consisting of 20 videos are between 480p-1080p and the sizes are between 1MB and 10MB, and the videos we used from the DFDC dataset are between 1080p and their sizes are between 3MB-15MB. By comparing the datasets we received with other datasets, it was determined that the DFDC dataset was better than other datasets because it consisted of 128,000 images and included the most forgery methods. In addition, it has been determined that the dataset with the most videos after the DFDC dataset is CELEB-DF, and the dataset that processes the most methods is the FF(FaceForensics) dataset.

The FaceForensics dataset includes 4 different forgery methods: FaceSwap, DeepFakes, Face2Face and NeuralTextures.

The DFDC dataset was shot by many different actors with various lighting conditions, head poses and various backgrounds [18].

There are 5639 fake and 590 real videos in the CELEB-DF dataset. The actual videos in the dataset were created from videos collected from Youtube by celebrities of different genders, ages and races. Comparison of data sets is given in Table 1.

In the test results we have done, both models were first tested on fake videos, then both models were tested on real videos and the results were reported.

The proposed method has been tested in on videos consisting of both fake and real titles, and the reports on these videos have been converted into numerical values by testing them on different data sets as shown in Table II and Table III.

In the results obtained, it has been confirmed that the Xception model has higher accuracy rates than the EfficientNetB4 model in real videos, and on the contrary, the EfficientNetB4 model has better accuracy rates than Xception in fake videos. The accuracy rates of the models on real and fake videos are given in Table II and Table III.

TABLE I
Comparison of Data Sets

Data Set	**Comparison of Datasets**	
	Number of Video	***Number of Method***
DFDC	128154	8
CELEB-DF	6229	1
FF+DF	5000	4
UADFV	98	1
Google DFD	3000	1
DF-TIMIT	960	2

TABLE II
Accuracy Rates of Models According to Different Data Sets on Fake Videos

Data Set	**Model Performances in Fake Videos**	
	Xception	***EfficientNetB4***
DFDC+Private	96.50%	**97.80%**
CELEB-DF	53.39%	**87.47%**
FaceForensics	61.08%	**79.32%**

TABLE III
Accuracy Rates of Models According to Different Data Sets on Real Videos

Data Set	**Model Performances in Reel Videos**	
	Xception	***EfficientNetB4***
DFDC+Private	**92.10%**	87.00%
CELEB-DF	**98.27%**	92.42%
FaceForensics	**78.71%**	55.11%

Our study detects forgery on the images divided into frames, converts the numerical value obtained in these frames into an average, and if the resulting numerical value is between 0.5-1.0, the image is fake and between 0.0-0.5 it detects that the image is real. The numerical results we have obtained according to the frames are shown in the graph in Fig. 5.

IV. Conclusion

As a result of the studies we have done, it has been emphasized how much of a concern deep forgery is today and some studies in this field are explained. By obtaining our proposed method and the experimental results of our proposed method on different data sets, it has been tried to benefit from fake detection.

In the application, it has been tried to contribute to deep fraud detection by using a CNN-based deep learning algorithm, by making deepfake detection according to the areas where Xception and EfficientNetB4 models are good.

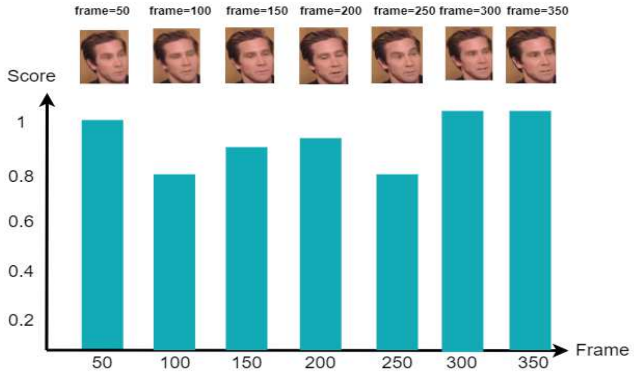

Figure 5. Extracting detection results from frames

As it showed us in our study, if we only used the Xception model in the DFDC data set, there would be a loss of 1.3% in practice, and on the contrary, if we only used the EfficientNetB4 model, there would be a loss of 5.10%. If we evaluate the same situation for the CELEB-DF dataset, there would be a loss of 34.08% on fake videos if we only used the Xception model, and 5.85% on real videos if we used the EfficientNetB4 model only.

Due to the high accuracy of the Xception model on real videos and the high accuracy of the EfficientNetB4 model on fake videos, the combination of the two models provided more effective results. Thus, in our proposed method, were achieved 92.10%-97.80% real-fake accuracy rates in the DFDC dataset, 98.27%-87.47% in the CELEB-DF dataset, and 78.71%-79.32% in the FaceForensics dataset.

References

[1] Digvijay Yadav, Sakina Salmani, “Deepfake: A Survey on Facial Forgery Technique Using Generative Adversarial Network”, Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019).IEEE Xplore Part Number: CFP19K34-ART; ISBN: 978-1-5386-8113-8.

[2] Hsu, C.-C.; Zhuang, Y.-X.; Lee, C.-Y. Deep Fake Image Detection Based on Pairwise Learning. Appl. Sci. 2020, 10, 370. https://doi.org/10.3390/app10010370

[3] LIANG, Ruigang, et al. A survey of audiovisual deepfake detection techniques. Journal of Cyber Security, 2020, 5.2: 1-17.

[4] YU, Peipeng, et al. A Survey on Deepfake Video Detection. IET Biometrics, 2021.

[5] GROH, Matthew, et al. Deepfake detection by human crowds, machines, and machine-informed crowds. Proceedings of the National Academy of Sciences, 2022, 119.1.

[6] Mr. Neeraj Guhagarkar, Ms. Sanjana Desai, Mr. Swanand Vaishampayan, Prof. Ashwini Save, "DEEPFAKE DETECTION TECHNIQUES: A REVIEW", VIVA-IJRI Volume 1, Issue 4, Article 2, pp. 1-10,.

[7] Wodajo, D., & Atnafu, S. (2021). Deepfake Video Detection Using Convolutional Vision Transformer. ArXiv, abs/2102.11126.

[8] Ismail, A., Elpeltagy, M., S Zaki, M., & Eldahshan, K. (2021). A New Deep Learning-Based Methodology for Video Deepfake Detection Using XGBoost. Sensors (Basel, Switzerland), 21(16), 5413. https://doi.org/10.3390/s21165413

[9] XIA, Jiyu; HUA, Man. Numerical Analysis and Optimization Application Trend of Deepfake Video Inspection Technology. In: Journal of Physics: Conference Series. IOP Publishing, 2021. p. 012068.

[10] Yasrab, R.; Jiang, W.; Riaz, A. Fighting Deepfakes Using Body Language Analysis. Forecasting 2021, 3, 303-321.

[11] KHALIL, Samar Samir; YOUSSEF, Sherin M.; SALEH, Sherine Nagy. iCaps-Dfake: An Integrated Capsule-Based Model for Deepfake Image and Video Detection. Future Internet, 2021, 13.4: 93.

[12] Coccomini, D., Messina, N., Gennaro, C., & Falchi, F. (2021). Combining efficientnet and vision transformers for video deepfake detection. arXiv preprint arXiv:2107.02612.

[13] İlhan, İ., Karaköse, M. (2021). A Comparison Study for The Detection and Applications of Deepfake Videos. Adıyaman University Journal of Engineering Sciences, 8(14), 47-60.

[14] D. S. Brar, A. Kumar, Pallavi, U. Mittal and P. Rana, "Face Detection for Real World Application," *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*, 2021, pp. 239-242, doi: 10.1109/ICIEM51511.2021.9445287.

[15] NIRKIN, Yuval, et al. DeepFake detection based on discrepancies between faces and their context. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[16] CHOLLET, François. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 1251-1258.

[17] Bonettini, Nicolo & Cannas, Edoardo & Mandelli, Sara & Bondi, Luca & Bestagini, Paolo & Tubaro, Stefano. (2020). Video Face Manipulation Detection Through Ensemble of Cnns.

[18] Liu, Jiarui, et al. A lightweight 3D convolutional neural network for deepfake detection. International Journal of Intelligent Systems, 2021, 36.9: 4990-5004.