Deepfake Detection in Media Files - Audios, Images and Videos

Bismi Fathima Nasar
Dept. of Computer Science and Engineering
ER & DCI Institute of Technology
Thiruvanthapuram, Kerala - India
bismiputhuparambil12@gmail.com

Sajini T
Knowledge Resource Center
CDAC
Thiruvanthapuram, Kerala - India
sajini@cdac.in

Elizabeth Rose Lason
Dept. of Computer Science and Engineering
ER & DCI Institute of Technology
Thiruvanthapuram, Kerala - India
elizabeth@cdac.in

2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS) | 978-1-7281-9052-5/20/$31.00 ©2020 IEEE | DOI: 10.1109/RAICS51191.2020.9332516

***Abstract*—Recent advancement in deep learning has applied to solve various complex problems ranging from big data analytic to computer vision and human-level control. One among them is the deepfake technology which becomes a real threat to privacy, democracy, and national security. Deepfake is hyper-realistic digitally manipulated videos to depict people saying and doing things that never actually happened. This technology has been used in many fields in film industries for recreating videos without re-shooting, awareness video generation such as creating voices of those who have lost theirs or updating episodes of movies without re-shooting them at very low cost. This technology has many harmful usages in social media, pornographic sites, etc. to deface peoples which largely dominate the positive side of this application of deep learning. Also, the creation and spreading of these videos are increasing rapidly along all fields of media files. Therefore, it is very much important to develop efficient tools that can automatically detect the deepfake in these videos and thus reduce the public harm caused by such videos. In the early stages of deepfake detection, traditional technologies like signal processing, image processing, lip-syncing, etc were used but this provides very little accuracy when combined with the recent technologies of deep learning. So, here a system is proposed that can automatically detect the deepfake in media files such as images, videos, and audios. This uses an image processing approaches combined with deep learning which detects the inconsistency that exists in fake media.**

***Index Terms*—GAN, CNN, LSTM**

I. Introduction

The first known attempt of Deepfake was developed during 1865. It can be found in one of the iconic portraits of U.S President Abraham Lincoln. The lithography mixes Lincoln's head with the body of southern politician John Calhoun. After Lincoln's assassination, demand for lithographs of him was so great that engravings of his head on other bodies appeared almost overnight. Later on, in just over 18 months, a small subculture on Reddit focused on combining and superimposing images and videos with realistic and believable results that have exploded onto the national scene. Deepfake technologies have been a combination of Artificial Intelligence along with Machine Learning that allows to create fake videos which are very much difficult to differentiate from authentic videos. Deepfake algorithm used many deep learning methods like auto encoders and GAN has been widely used to train large dataset to generate the training models. These models are used in the testing phase of these videos to examine the facial expression, movements of a person, etc. The initial target of Deepfake videos were some public figures such as celebrities and politicians since they have a large number of videos and images available online. Thus these videos can be a threat to democracy since these methods were used to create fake speeches of political leaders which can cause impact over the election campaigns. Evidence has been obtained that even these Deepfake has misled military troops to lose the battle by creating a fake bridge across a river although the bridge was not there in reality. There is also positive side for Deepfakes such as this method can be used in creating videos which was lost or updating episodes of movies without re-shooting them. But the negative impact of these videos dominates over the positive usage of these videos. The various methods used in the creation of Deepfake include - Face Replacement, Face Re-enactment, Face Generation using GAN, Face Generation using the attributes that exist in the source image, and Speech Synthesis. Also, the process of creating those manipulated images and videos is also much simpler today as it needs as little as an identity photo or short video of a target individual. Recent technology can even create a Deepfake video with still images. Deepfake thus not only after the celebrities but also ordinary peoples. Also with the development of various applications like DeepNude, Fakeapp, etc. shows more spreading of the threats caused due to Deepfakes as it can transfer a person to non-consensual porn. These forms of falsification cause a real threat to privacy, identity, national security and affect many human lives. This paper basically consists of VII sections. Section I is the introduction, Section II contains the recent techniques involved in the creation and detection of the Deepfake multimedia, Section III covers the various works that has come up based on the Deepfake technology, Section IV states the proposed system detailing with a block diagram, Section V includes the detailing of the various stages in the development of the proposed system, Section VI covers the various challenges that we faced during the development of the proposed system along with the future scopes and finally the last section that is Section VII brings the entire paper into a conclusion phrase.

Authorized licensed use limited to: Nitte Meenakshi Institute of Technology. Downloaded on March 27,2021 at 15:03:50 UTC from IEEE Xplore. Restrictions apply.

II. TECHNICAL APPROACH IN DEEPFAKES

With the advancement in the deep learning techniques, the state of the art approach has also put up its footprint in the Deepfake creation and detection. This approach mainly includes AI-based techniques like CNN architecture, GAN, etc. explained in the section below. Also, Code snippets are readily available so it is very much easy to generate our own user-friendly software, require very less expert knowledge and the process is time-consuming. Many studies have been done to make it closer to the user expectation and to increase their usage by improving the techniques. The basic techniques involved in the state of the art approach used in this work is covered in this section below:

- CNN Network - CNN or the Convolutional Neural Network is a Deep Learning algorithm that takes the images of various categories as input and detect the point of difference between the categories which helps to differentiate them from one another. This technique is mainly deployed in the creation and detection of Deepfake. Also, the pre-processing required in a ConvNet is much lower as compared to other classification algorithms. CNN Network basically consists of two parts: One is a convolution tool layer that contains a convolution, max-pooling, and activation layer that splits the various features in the image and uses it for analysis. Secondly is a fully connected layer that uses the output of the convolution tool layer to predict the best description for the image. CNN uses the predictions from the layers to produce a final output that presents a vector of probability scores to represent the likelihood that specifies the distinct features of each class. The figure showing an example of the analysis over CNN is shown below.

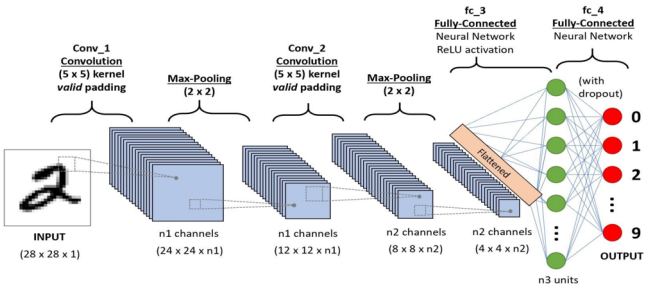

Fig. 1. CNN Network

- Generative Adversarial Network (GAN) - Generative Adversarial Networks (GANs) is a powerful class of neural networks that are used for unsupervised learning. Deepfake videos are usually created by using two competing GAN network-based AI systems -one is called the generator and the other is called the discriminator. The generator creates fake videos and makes the discriminator distinguish between fake and real video samples.Each time the discriminator accurately identifies the fake video samples, it gives the generator a hint about what not to do when creating the next fake samples. Conversely, as the discriminator gets better at detecting the fake video samples, the generator gets better at creating them. Together, the generator and discriminator form something called a generative adversarial network (GAN). The figure describing the working of the GAN is given below.

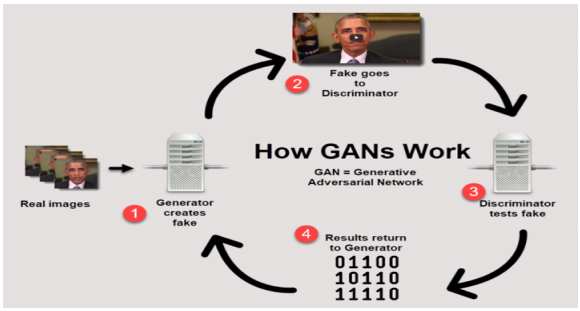

Fig. 2. GAN Network

III. RELATED WORKS

In this section, we are going to discuss the various literature work in the Deepfake creation and detection domain which we specified in our survey paper on deepfake [1]. In the paper [2], a detailed survey on the various passive blind video content authentication with the main focus on forgery detection, video recapture, and phylogeny detection. The auto encoders had the following disadvantages when compared to the recently used convolutional or neural networks [3]. First disadvantage is the Lack of temporal awareness which is the basic source of multiple abnormalities in the auto encoders. Next is the inconsistencies existing with the face encoder i.e. the Encoder is unaware of the skin tone or other background information. The third disadvantage is the visual inconsistency that exist due to the use of multiple cameras, different lighting conditions, or simply the use of different video codecs which make it tough for the auto encoder to create very accurate and realistic videos under different conditions. Finally, it is the inconsistency in choosing the illuminates between the different background with frames. This usually leads to blinking in the face region in the most of the Deepfake videos. So, to overcome these disadvantages over the auto encoder, another deep learning technique like the convolutional neural network (CNN), Recurrent Neural Network (RNN), Generative Adversarial Network (GAN), etc. were developed.

With the advancement in this convolutional network, there were many other schemes were developed in the creation and detection of Deepfake using Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), even the hybrid approaches of all the recent algorithm in the Deep Learning, etc. David Guera and Edward J Delp [3] brings up the first approach where the frame level features are extracted out after each processing in convolutional neural networks. These features are then fed into the recurrent

Authorized licensed use limited to: Nitte Meenakshi Institute of Technology. Downloaded on March 27,2021 at 15:03:50 UTC from IEEE Xplore. Restrictions apply.

neural network as training samples. The output from this RNN is the classification result. Along with the combination of CNN and RNN, a set of encoder-decoder is also used for dimensionality reduction and image compression in the training and generation phase. Another such approach was brought up by Ekraam Sabir in [4] were the face manipulation detection using RNN strategies was discussed. They uses a combination of variations in RNN models along with domain-specific face pre-processing techniques to obtain state-of-the-art performance on publicly available facial manipulation videos. The experimental evaluation shows an accuracy of 4.55 %. A similar approach was stated in [5] for the swapped face detection using Deep Learning and Subjective Assessment. In this paper, they proposed a swapped face detection system which shows 96 % positive results with few false alarms when compared with the other existing systems. Along with the detection of face-swapping, this model also evaluates the uncertainty in each prediction which is very much critical in the evaluation of the performance of a system. In order to improve this predictability, they have setup a website to review the human response over the dataset by collecting pair to pair comparison of images over the videos on humans. Based on these comparisons, images are classified as real or fake. It's experimental results show that this proposed model is much better when compared to the existing systems.

When it comes to the most advanced version of this deep neural network, a hybrid approach was implemented in [6]. A two-stream neural network was proposed trained by GoogLeNet. In the first stream, the tampering artifacts like strong edges near lips, blurred areas on the forehead, etc. was detected in image classification stream and in the second stream, a patch-based three-layer network is trained for capturing local noise components and camera characteristics. This network is designed to determine whether it comes from the same image or not.

Another concept in the hybrid model was pairwise learning [7] where a deep learning-based approach is used to identify manipulated images with contrastive loss. First, state-of-the-art GANs network is used to generate pair of fake and real images. Then, these pairs of image samples are fed into a common fake features network (CFFN) to learn the distinguish feature between the fake image and real image as a paired information. Then in the final stage, a small network will be used to combine these features to make the decision on whether it's fake or real. Experimental results show that the proposed method has high performance when compared to the existing state of the art image detection techniques. In the paper [8], a task-oriented GAN for PoISAR image classification and clustering techniques were used which consists of a Triplet network. Along with the generator and the discriminator, they are another network called the task network or T-net. The network in this proposed system basically has two task networks – one is called as the classifier and another is called as a clustered network. The first is the learning stage which has the two competing generator and discriminator network which work hand in hand as in GANs network. In the second phase, the generator network is adjusted and oriented as a Task network where some samples from the training samples are assigned with a specific task that is the generation of the manipulated data. This takes up the advantage of a GAN network and also overcomes the disadvantage of the GAN network.

Later on, a Hybrid combination of LSTM and Encoder-Decoder architecture [9] was developed for image forgery detection. In this system, a high-confidence architecture is used which utilizes re-sampling features used to capture artifacts. These artifacts include JPEG quality loss, up sampling, down sampling, rotation, etc., Long-Short Term Memory (LSTM) cells, an encoder-decoder network to distinguish whether the specific area of the image has tampered or not. Here they use a spatial map and the frequency domain correlation to determine the distinct characteristics of the manipulated and non-manipulated regions by combining the effort of LSTM and Encoder network. Finally, the decoder network learns how it is mapped from low-resolution feature maps to pixel-wise predictions for tamper detection. Through this work, they also present a dataset that can be used in further research work on media forensics. With several experiments conducted with different data sets, they brought up to the conclusion that their scheme was efficiently segmented various types of manipulations including copy-move, object removal, and splicing.

The active learning approach was also stated in this field of deep learning [10] as a wider advancement to acquire annotations for data from a human reaction by selecting informative samples with a high probability to enhance performance. This model is implemented to generate a label to the data in a cheaper manner. Here for each sample, a reward will be assigned by the classifier trained with these pre-existing labels and these rewards can be used to guide a conditional GAN to generate useful and informative samples with a higher probability for a certain label. Finally, with the evaluation of this model, the effectiveness of the model can be estimated showing that the generated samples are capable of improving the classification performance in popular image classification tasks. Then certain pre-processed authentic or fake images [11] can be used to train the CNN network in the generation which destroys the unstable low level noise cues on the manipulated images and the discriminative network is forced to learn more distinct features to classify the manipulated and real face images. A key difference with other GAN related methods is that here they use an image pre-processing step in the training stage to destroy low-level unstable artifacts of GAN images and force the discriminator to focus on more distinct clues and by doing so they improve the generalization capabilities. But this approach was difficult to implement and has got only some preliminary results. Thus, to improve the discriminative capabilities, face wrapping artifact detection technique was developed [12]. This method was developed based on the observation that current Deepfake algorithms can generate images of low resolution and further wrapping is essential to make the manipulated one with that of the original

Authorized licensed use limited to: Nitte Meenakshi Institute of Technology. Downloaded on March 27,2021 at 15:03:50 UTC from IEEE Xplore. Restrictions apply.

one. So, such transform leaves an artifact called resolution inconsistency along the line of fake one and these artifacts can be effectively captured using a CNN network for detecting the authenticity of the video.

Many other approaches with indirect implementation of Deep Learning were also implemented in parallel to the direct approaches but shows less accuracy when compared to other existing systems. One such approach was proposed in [13] where an automated system is developed that can detect the forgery in the videos being recorded in the camera and in its audio channel. Here they used the method in which they detect the audio-visual inconsistencies with certain artifacts like lip syncing, Dubbing inconsistencies etc. In the experimental evaluation, the proposed system is evaluated with various classifiers like the LSTM, GMM, PCA etc. but given a better result only with LSTM which flows as the drawback of the system. Another such approach in the detection of the forgery in the images and videos was a capsule forensic approach stated in the article [14] where they use a capsule network to detect the anomaly in the replay attack using printed images and videos to the computer-generated video using deep convolution neural networks. This experiment brings up the feasibility of generating a common detection technique that can be used to detect the forgery in the videos as well as images. Here the capsule network can be used in the domain along with computer vision where they use random noise samples in the training phase. The main aim of this work was to protect the random samples against machine attacks as well as mixed attacks. Another approach is the Eye blinking detection [15] where the temporal features of the eye and the inconsistency in the eye blinking is detected to identify the manipulation in the sample file using LSTM.

Finally, we provide a detail on a rare approach completely out from deep learning domain that can be used for the forgery detection in [16] where they use a PRNU (Photo Response Non-Uniformity) analysis for detecting the deep fake video manipulation. In this approach, the videos are divided into different frames and the frames corresponding to the face is cropped. Then the mean correlation is calculated between the authentic and Deepfake one is calculated to determine which one is fake and which is not. This method was also used to determine the amount of tampering in each Deepfake videos. PRNU analysis shows a notable difference in mean normalized cross-correlation scores between real and Deepfake medias. In the early stage of implementation of the detection techniques, there was no much academic paper found on the detection of Deepfake. Although efforts have been brought up to detect and remove these kinds of videos from websites such as Gyfcat [Matsakis, 2018]. Gyft attempts to use artificial intelligence and facial recognition software to mark inconsistencies in the facial region of an uploaded video.

IV. PROPOSED SYSTEM

Deepfake technology has advanced a great deal in recent years with the development of neural networks from Deep Mind doing an especially good job of creating realistic human-like voices, videos, or images. In order to reduce the harmful effects caused by these kinds of media files, we proposed a system that could detect the forgery in media files such as audios videos and images. There many existing systems coming up in this domain recently. The major issues with the existing techniques are that they deploy a detection method that can detect the forgery in any one of the media files. Most of them use traditional approach which needs expert knowledge and the process involved is time-consuming. In this work, we propose a system that uses a combinational approach using CNN architecture for image processing. The main objective of this system is to reduce the time involved in the processing to improve its performance when compared to the existing systems. The block diagram corresponding to the proposed system is shown in figure below.

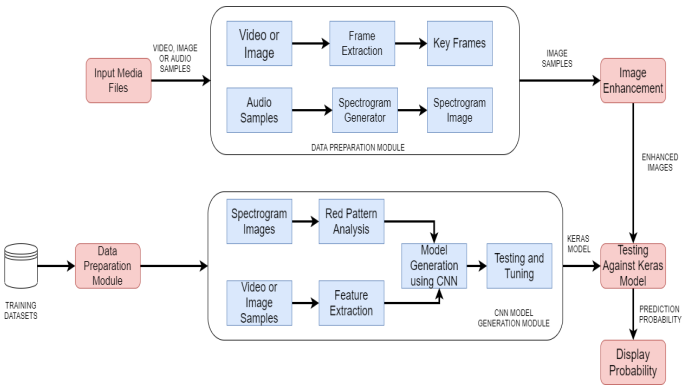

Fig. 3. Block Diagram of the Proposed System

The system consists of four modules namely Data Preparation module, Image Enhancement, CNN Model Generation, and Testing Phase or Detection phase. Initially, the user inputs the media file in the form of audio, video, or images. In the **Data Preparation**, the input media files get converted into images that is, the image will be taken as such, the videos will be converted into keyframes(Keyframes are usually the complete frames on which the data stream of a video is saved) and the audio is converted into spectrogram image. Then, these image files are given into an **Image Enhancement Module** where the image gets enhanced by removing the noise factor if present in the image files, and this process is called denoising. This enhanced image will be given for detection against the models being generated in the **CNN Model Generation Module**. The decision based on the testing will be given out in the form of prediction probability based on both the classes. This prediction probability has been calculated based on the comparison between the distinct features extracted out from the input media to the features saved in its corresponding model file. Based on this probability, the decision on forgery is displayed over the user interface.

The CNN Model Generation Module consists of two phases namely the training and testing phase. In the training phase,

Authorized licensed use limited to: Nitte Meenakshi Institute of Technology. Downloaded on March 27,2021 at 15:03:50 UTC from IEEE Xplore. Restrictions apply.

the patterns in the classes is learned by the model. Basically, three models are generated in this proposed system; one for images, another for the video, and the final one for the audio files. The model for the images is generated using a simple CNN network. The input to these network layers consists of a set of the fake and real image and the output file will be the Keras model called "model_image.h5" which contains features that can be used to classify between these two classes - fake and real. This model file is further used in the detection process in the case of images. The model for video detection is generated using the model for image forgery detection, but an improved version using transfer learning. This is given the name model _video.h5. For audio detection, we have generated another model called "model_audio.h5". This was generated by feeding spectrogram images to the CNN network which learn the variation in the normal energy distribution or intensity distribution which is further used in the testing phase. In the testing phase, the input sample was tested against the models been generated and made the decision based on the prediction probability. Then based on this prediction probability, the decision on fake or real is displayed. Display of meta-data corresponding to the input media file is added as an additional feature. This is stored in HTML format. The metadata file is generated by using an Exif Tool and the location of the corresponding metadata file will be also displayed over the interface.

V. RESULTS AND DISCUSSIONS

The main challenge we have faced during the development of the proposed system is the lack of dataset for training and testing in the CNN Network. To address these challenges, a image dataset is generated using face-swapping techniques The dataset for the image was generated by using the technique called face-swapping. This technique begins by detecting the face region in both source and target image and then cut off the face region from both of them. The face from the source is then placed over the empty space in the target image. As a base for the generation of the fake images, we use the face images of 50k celebrities from CelebA dataset. In the case of audio dataset generation, we use the various signal processing and manipulation techniques over real human voice recorders.

After the dataset generation, we start our implementation with the data preparation. In this proposed system, we use a basic image classifier in all the three cases. So, initially we need to convert all the three media files to images for the training and testing in CNN network for the development of Keras model. The videos are converted into sequence of key frames using *OpenCV* and the audio is converted into spectrogram images using *Matplotlib*. Then these converted image samples undergoes image enhancement with the *Librosa* package.The next phase is the training and testing phase over binary categorical CNN network. In the CNN training for images, the CNN network consists of two sequential layers and two dense layers each for each class of images. The generated dataset is fed over the CNN network to generate the model file named as model _image.h5.The generated model file has 538,508 parameters in which 524,428 are trainable parameters and the remaining are non-trainable parameters. The model training provided an accuracy ranging from 0.6 -1.0 and loss ranging from 0.54 - 0.0. The plot corresponding to accuracy and loss is given in the figures below.

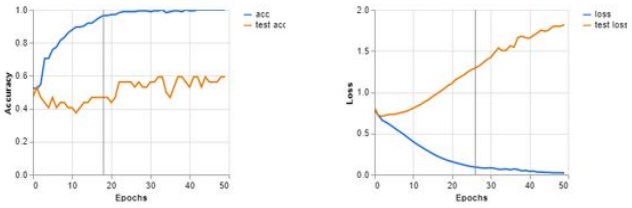

Fig. 4. Accuracy and Loss Corresponding to Image Modelling

In the case of video, the CNN training was done by adding more specific samples as training data and improved the performance of the model generated in image detection. The generated model file was named as "model video.h5". This model file was also tested against different data sets like Deepfake Timit, Forensics++, VidTimit dataset, etc, and it's performance evaluation which is tabulated, as shown in figure below.

Dataset	Real	Fake	Accuracy
VidTIMIT	0.10 – 0.15	0.75 – 0.99	99 %
Deepfake TIMIT	0.25 – 0.30	0.66 – 0.85	85 %
Face Forensics ++	0.12 – 0.20	0.70 – 0.90	90 %

Fig. 5. Evaluation Result corresponding to Video Modelling

In the case of audio, the CNN classification is done by analyzing the variation in the normal energy distribution or intensity distribution in the spectrogram. The training starts with the generation of the spectrogram image of the 1000+ samples of fake and real images that are given to the CNN network containing three convolution layers, three max-pooling layers, and five activation layers. The output from these layers is given to the fully connected convolution layer and is trained with an epoch = 15. The characteristics of red patterns corresponding to the fake and real samples are extracted out to generate the model audio.h5 file. This training provides an accuracy of 0.8611 and Loss of 0.290 when trained with 100 samples and further to increase the accuracy the number of samples is increased to more than 1000 samples. The accuracy and loss plot is shown in the figure below.

Authorized licensed use limited to: Nitte Meenakshi Institute of Technology. Downloaded on March 27,2021 at 15:03:50 UTC from IEEE Xplore. Restrictions apply.

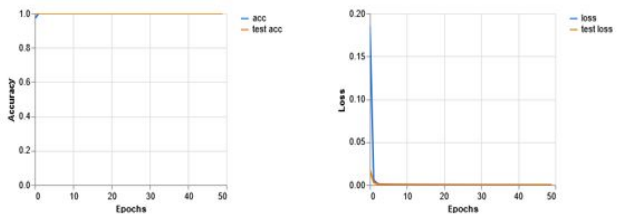

Fig. 6. Accuracy and Loss Corresponding to Audio Modelling

In the testing phase with all the three trained Keras model, we have tested with generated dataset as well as some real time dataset available like the VidTimit dataset, Face Forensics++ etc. and calculated the various parameter like the True positive, True negative, False positive, False negative, Accuracy, Precision and Recall. In all the three cases, we have got an accuracy nearly 0.9.

VI. CHALLENGES AND FUTURE SCOPES

The main challenges faced is the lack of availability of Deepfakes with high quality, properly classified and the original data. Both these data is required for training a supervised model. Available dataset normally contains either original or the fake. Another important challenge is the incompatibility of these detection techniques and their associated packages in the generic systems. Since, the system that is being used generally require high specification like high-quality graphics card, interfaces that support machine learning packages, high memory for the training process, etc. These challenges always pull the research works backward. This system with better dataset and training, it can be used in various fields like for women safety, Fake news detection, Fake checker over identity of an individuals and on investigation of various cyber crimes in future.

VII. CONCLUSION

Deepfakes are hyper-realistic video forgeries in which people say or do things that they have never actually said or done which have become a real threat to society. Mere visual verification is not enough to make a judgment on this kind of forgery. Since the visual quality of Deepfakes has become so flawless and most of the current technologies develop really high quality ones. Digger's solution is to develop a toolkit that can detect deep fakes in videos, images, and audios using the state of the art approach. So, here a scheme is developed to detect the forgery in the Deepfakes by deploying a combinational approach that uses image processing with the CNN network. The proposed system basically has four modules. The first one is the Data preparation module where the input samples are converted into image samples. Secondly is the Data Enhancement module where the noise component from the image is removed. Next is the CNN Network where the training and testing take place to generate the model and finally is the detection module, where the real-time detection against different media files taken from various platforms takes place. The accuracy of the model is also evaluated by testing against various datasets like the Deepfake Timit dataset, Face Forensic++, etc. and almost got the accuracy as 0.9 for all the three model files.

REFERENCES

[1] Bismi Fathima Nasar, Sajini. T, Elizabeth Rose Lalson, "A Survey on Deepfake Detection Techniques", International Journal of Computer Engineering in Research Trends, pp:49-55 ,August-2020.

[2] Raahat Devender Singh, Naveen Aggarwal," Video content authentication techniques: a comprehensive survey", Springer, Multimedia Systems, pp. 211- 240, 2018.

[3] David G'uera Edward J. Delp," Deepfake Video Detection Using Recurrent Neural Networks", Video and Image Processing Laboratory (VIPER), Purdue University,2018.

[4] Ekraam Sabir, Jiaxin Cheng, Ayush Jaiswal, Wael AbdAlmageed, Iacopo Masi, Prem Natarajan," Recurrent Convolutional Strategies for Face Manipulation Detection in Videos", In proceeding of the IEEE Xplore Final Publication, pp. 80-87, 2018.

[5] Xinyi Ding, Zohreh Razieiy, Eric C, Larson, Eli V, Olinick, Paul Krueger, Michael Hahsler," Swapped Face Detection using Deep Learning and Subjective Assessment", Research Gate, pp. 1-9, 2019.

[6] Peng Zhou, Xintong Han, Vlad I. Morariu Larry S. Davis," TwoStream Neural Networks for Tampered Face Detection", IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[7] Chih-Chung Hsu, Yi-Xiu Zhuang, and Chia-Yen Lee," Deep Fake Image Detection based on Pairwise Learning", MDPI, Applied Science,2020, doi:10.3390/app10010370.

[8] Fang Liu, Licheng Jiao, Fellow, IEEE, and Xu Tang, Member" TaskOriented GAN for PolSAR Image Classification and Clustering", IEEE Transactions On Neural Networks and Learning Systems, Volume 30, Issue 9, 2019.

[9] Jawadul H. Bappy, Cody Simons, Lakshmanan Nataraj, B.S. Manjunath, and Amit K. Roy-Chowdhury," Hybrid LSTM and EncoderDecoder Architecture for Detection of Image Forgeries", IEEE Transaction on Image Processing, Volume: 28 , Issue: 7 ,pp. 1-14, 2019.

[10] Xinsheng Xuan, Bo Peng, Wei Wang and Jing Dong," On the Generalization of GAN Image Forensics", Computer Vision and Pattern Recognition, Cornell University, Volume 1, pp. 1-8, 2019.

[11] Yuezun Li, Siwei Lyu," Exposing DeepFake Videos by Detecting Face Warping Artifacts", In Proceedings of the IEEE Xplore Final Publication, pp. 46- 52, 2019.

[12] Pavel Korshunov, S´ebastien Marcel," Speaker Inconsistency Detection in Tampered Video", 26th European Signal Processing Conference (EUSIPCO), 2018, ISBN 978-90-827970-1-5.

[13] Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen," CapsuleForensics: Using Capsule Networks to detect Forged Images and Videos", ICASSP, pp. 2307 – 2311, 2019

[14] Li, Y., Chang, M. C., and Lyu, S, "Exposing AI created fake videos by detecting eye blinking", In IEEE International Workshop on Information Forensics and Security (WIFS) (pp. 1-7). 2018.

[15] Steven Fernandes, Sunny Raj, Rickard Ewetz, Jodh Singh Pannu, Sumit Kumar Jha, Eddy Ortiz, Iustina Vintila, Margaret Salte," Detecting deepfake videos using attribution-based confidence metric", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 308-309), 2020.

[16] A.M. Rodriguez, Z. Geradts," Detection of Deepfake Video Manipulation", In Proceedings of the 20th Irish Machine Vision and Image Processing conference, Belfast, Northern Ireland, pp. 133-136, 2018, ISBN 978-0-9934207-3-3.

[17] Rohini Sawant and Manoj Sabnis," A Review of Video Forgery and Its Detection", IOSR Journal of Computer Engineering (IOSR-JCE) eISSN: 2278-0661, Volume 20, Issue 2, p-ISSN: 2278-8727, 2018.

[18] Thanh Thi Nguyen, Cuong M. Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen and Saeid Nahavandi," Deep Learning for Deepfakes Creation and Detection", IEEE, pp. 1-12, 2019.

Authorized licensed use limited to: Nitte Meenakshi Institute of Technology. Downloaded on March 27,2021 at 15:03:50 UTC from IEEE Xplore. Restrictions apply.