International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 07 | July - 2022 Impact Factor: 7.185 ISSN: 2582-3930

THE DETECTION OF VIDEO MANIPULATION OF FACES USING A NETWORK OF CONVOLUTIONAL NEURAL NETWORKS

V.VENKATA RAMANJANEYULU[1]
ASSISTANT PROFESSOR
DEPT.OF CSE
TEEGALA KRISHNA REDDY ENGINEERING COLLEGE, MEERPET
P.MEENAKSHI[2], P.MADHUSMITA[3], PASAM BHARGAVA SAI[4]
DEPT. OF CSE
TEEGALA KRISHNA REDDY ENGINEERING COLLEGE, MEERPET

ABSTRACT

These days, it is not difficult to conceive scenarios that these lifelike face swapped deep fakes are used to influence political circumstances, organise terrorist acts, or blackmail people. In this paper, we offer a novel approach based on deep learning that is able to efficiently differentiate between actual movies and fraudulent ones created by artificial intelligence. Our approach employs a Res-Next Convolution neural network to extract the frame-level characteristics, and these features are then utilised to train a Long Short Term Memory (LSTM) based Recurrent Neural Network (RNN) to classify if the video has been subjected to any form of manipulation or not, i.e. whether the video is a deep fake or true video. Our technique is evaluated using a large quantity of balanced and mixed data sets, which were created by combining a number of other publicly accessible data sets, such as Face-Forensic++, the Deepfake detection challenge, and Celeb-DF.

Keywords: LSTM,RNN, data-set, Res-Next Convolution neural network, deep learning, AI

1.INTRODUCTION:

In the rapidly evolving world of social media platforms, deep fakes are viewed as the greatest danger posed by artificial intelligence. Realistic face swap deep fakes may be used in a variety of settings to create political unrest, conduct phoney terror acts, or blackmail individuals. Examples include Brad Pitt and many more.

Differentiating deepfake from real video is a pressing concern. We're employing AI to fight AI. Deep fakes may be created with the use of software like FaceApp and Face Swap, both of which rely on pre-trained neural networks like GAN and Auto encoders. Our approach employs an LSTM-based convolutional neural network for processing the sequential temporal analysis of video frames, and a pre-trained Res-Next CNN for extracting frame-level features. ResNext Convolution neural network captures frame-level information to detect whether a video is Deepfake or genuine. These attributes are then utilised to teach a short-term memory-based artificial recurrent neural network. In order to better prepare the movies for the customers' usage, we have developed a front-end application that allows users to submit the videos. Once the video has been processed by the model, the model's confidence in its deepfake/real verdict and its rendered result will be presented back to the user.

2.LITERATURE SURVEY:

By comparing the generated face areas and their surrounding regions with a specific Convolutional Neural Network model, Face Warping Artifacts [15] developed a method to identify artefacts. There were two types of face artefacts in this work. They developed their technique in response to the realisation that the current deepfake algorithm can only produce images of a certain resolution, which must then be further modified to match the faces to be substituted in the source video. The temporal analysis of the frames was not taken into account in their methodology.The article Detection by Eye Blinking proposes a novel technique for classifying videos as deepfakes or pristine by using the eye blinking as a key characteristic. The cropped frames of eye blinking were temporally analysed using the Long-term Recurrent Convolution Network (LRCN). Since today's deepfake creation algorithms are so advanced, the absence of eye blinking cannot be the only indicator of a deepfake. For the detection of profound fakes, additional factors like teeth enchantment, facial wrinkles, incorrect brow positioning, etc. must be taken into account.Capsule networks to detect forged images and videos [17] employs a technique that uses a capsule network to find fake, altered images and videos in a variety of situations, such as replay attack detection and computer-generated video detection. Their approach uses random noise during the training phase, which is not a good choice. Even so, the model showed promise in their dataset, but it may falter on real-time data due to training noise. It is suggested that our approach be trained on real-time, noiseless datasets.Independent of the creator, content, resolution, and video quality, False Catcher accurately identifies fake content. Formulating a differentiable loss function that follows the suggested signal processing steps is not a simple task because the lack of a discriminator results in a loss in their discoveries to preserve biological signals.

3.OBJECTIVE:

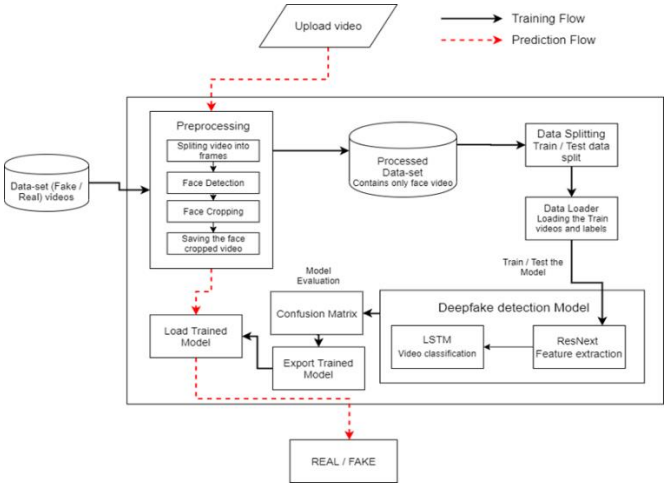
However For a long time, visual effects were the only way to show off changes to digital images and videos that were convincing. However, recent advances in deep learning have greatly increased the realism of false material while also

making it easier to generate. Our studies aim to locate the distorted realities of deep fakes, to classify films as either deep fakes or immaculate, and to provide a straightforward method for submitting movies and establishing their authenticity.

4.EXISTING SYSTEM:

To However, one of the most important applications of face detection is in the field of facial recognition. Facial recognition is a biometric technique that goes well beyond simple facial detection. Indeed, it tries to figure out whose face it is. The process involves comparing a digital image of a person's face (often from a video frame) with images already stored in a database. When there is a high likelihood that a face matches one in the database, facial recognition works, despite its limitations. Methods for recognising faces may be categorised by how they identify faces, such as appearance, features, information, or templates. There are good points and bad points to each: In rule-based or knowledge-based methods, a face is described using predefined criteria. The biggest problem of this method is how hard it is to establish uniform standards. The Multi-Task Cascaded Convolutional Neural Network (MTCNN) is now one of the most popular facial recognition methods in use. However, there are drawbacks to these approaches as well, including high convolutional cost and processing cost, the necessity for a larger number of support vectors to achieve high accuracy, and poor detection rates, among others.

5.SYSTM ARCHITECTURE:



Diagrams of a system's architecture are used to depict the flow of information inside a company's database visually. It explains how a system gathers information from an input and stores and uses it in a database and generates reports. To accomplish certain business tasks, a logical system diagram depicts how data moves through the system. That which makes the logical system work is shown in the physical diagram.

6.MODULES:

6.1 Data-setGathering:

As a means of enhancing the model's capacity for real-time prediction. We acquired the information from numerous publicly available data-sets, including FaceForensic++, the Deepfake Detection Challenge, and Celeb-DF. Then, in order to accomplish exact and rapid detection on different sorts of videos, we integrated the datasets we had gathered. In order to counteract the model's training bias, we used a dataset consisting of 50% actual and 50% fake videos. For the sake of brevity, this research does not address audio deep fakes, hence the dataset for the deep fake detection task only contains a subset of movies with audio notifications. We preprocessed the DFDC dataset by removing the audio-altered movies using a Python script.

6.2 Pre-processing:

During the preprocessing stage, the movies have all clutter and interruptions removed. Video is just kept long enough to see the face. The video is segmented into frames as the beginning stage in the preparation process. After the video has been cut up into individual frames, the faces in each frame are extracted and the frames are cropped to just show the faces. Subsequently, the individual frames are pieced back together to form a whole new video. This approach is repeated for every single film, ultimately yielding a dataset consisting exclusively of films containing faces. The frames without faces are disregarded during preprocessing.

6.3Data-set split:

The dataset is separated into a train dataset and a test dataset, with a proportion of 70 percent train movies (4,200) to 30 percent test videos (1,800). The train and test split is a balanced split, which means that each split contains fifty percent of the genuine films and fifty percent of the phoney ones.

6.4 Model Architecture:

The CNN and RNN are both included into our approach. We have used the Pre-trained ResNext CNN model to extract the features at the frame level, and on the basis of the features that have been retrieved, an LSTM network has been trained to identify the video as either pristine or deepfaked. When working with training pilot movies and the Data Loader, the labels of the videos are loaded and fitted into the model for training.

6.5 Hyper- parameter tuning:

Selecting optimal hyper-parameters is what allows for the most precise results.Following several iterations of the model. We choose the optimal hyper-parameters for our data collection. Adam[21] optimizer with the model parameters is utilised to allow the adjustable learning rate.

7. PROPOSED SYSTEM:

To address the abovementioned restrictions, we discuss our suggested approach for video face modification detection, i.e.,

given a video frame, to recognise whether faces are genuine (pristine) or artificial. The foundation of the suggested approach is the idea of ensembling. To achieve this, we base our work off of the unique technique to automated scaling of CNNs introduced in, the EfficientNet family of models. This collection of architectures delivers greater accuracy and efficiency with regard to other state-of-the-art CNNs, and really proven to be highly beneficial to fill hardware and temporal limits given by DFDC. Our approach makes use of both convolutional neural networks and recurrent neural networks. We utilised a pre-trained version of the ResNext CNN model to extract the features at the frame level. An LSTM network is then trained to classify the video as either deepfake or perfect using the recovered data. For the training video split, the Data Loader is used to import the film labels and incorporate them into the model. Leaky Relu activation function is also a part of this concept. The model can learn the average rate of correlation between the input and the output thanks to a linear layer consisting of 2048 input features and 2 output features. The model employs a layer of adaptive average polling with a single output parameter. Whichgivesthe the intended output size of the picture of the form H x W. A Sequential Layer is used to process frames in order. When doing batch training, a batch size of 4 is often employed. A SoftMax layer is utilised to obtain the confidence of the model during predication.

8.IMPLEMENTATION:

8.1 ResNext CNN:

The pre-trained model of Residual Convolutional Neural Network is used.The model name is resnext50_32x4d(). This model consists of 50 layers and 32 x 4 dimensions. Figure shows the detailed implementation of model.

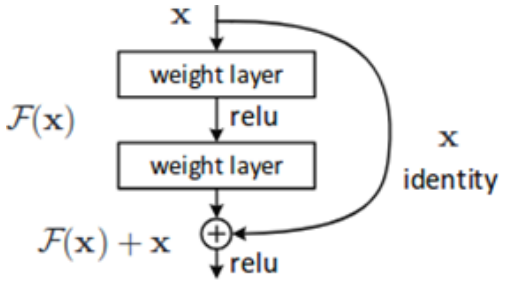


Fig 8.1 : ResNext Working

8.2 LSTM :

Long short-term memory (LSTM) is utilised to evaluate sequences and detect temporal differences between images. The input to the LSTM is a vector of features with a size of 2048. For this purpose, we use a single LSTM layer with 2048 latent dimensions, 2048 hidden layers, and a 0.4 percent dropout probability. By comparing the frame captured at time 't' with the frame captured at time 't-n', LSTM enables a temporal analysis of the video to be performed in a sequential fashion. To the extent that n is a positive integer greater than zero, the time interval preceding time t is completely variable.

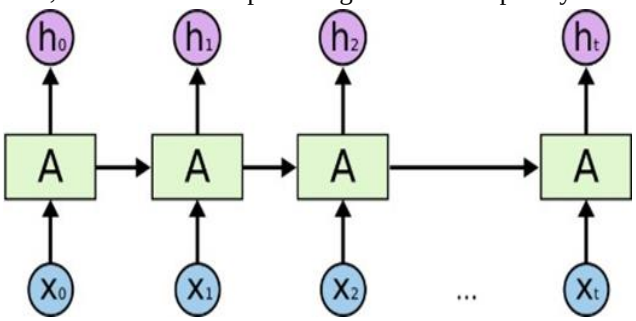
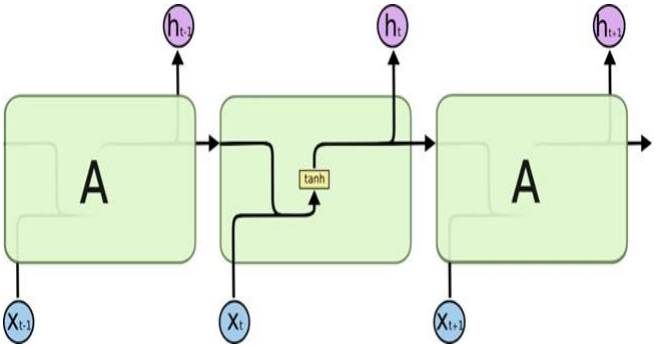


Fig 8.2(a) Overview of LSTM



. **Fig 8.2(b) Internal LSTM Architechture**

8.3 ReLU:

When the activation function's input is less than 0, the output is zero, and the output is raw otherwise. If the input is non-zero, the output is the same as the input. ReLU's functioning is more akin to that of real neurons in the human brain. Since ReLU is non-linear, it avoids the problem of backpropagation mistakes that plague the sigmoid function. Furthermore, the speed at which models based on ReLU may be constructed in large Neural Networks is quite high.

9.RESULT:

9.1 OUTPUT SCREENS :

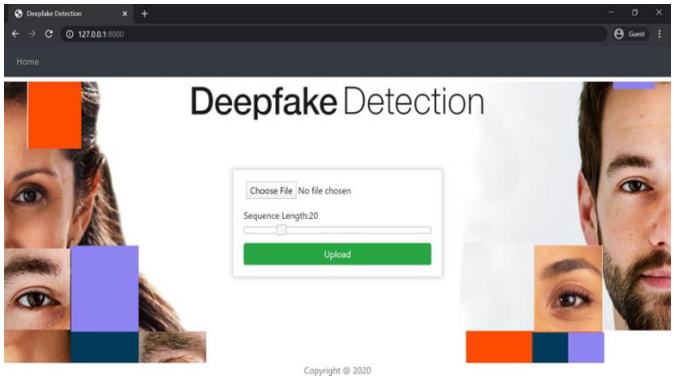


Fig (a). Home Page

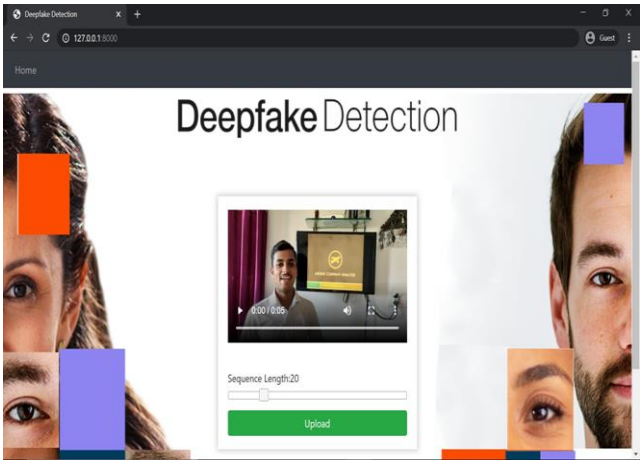
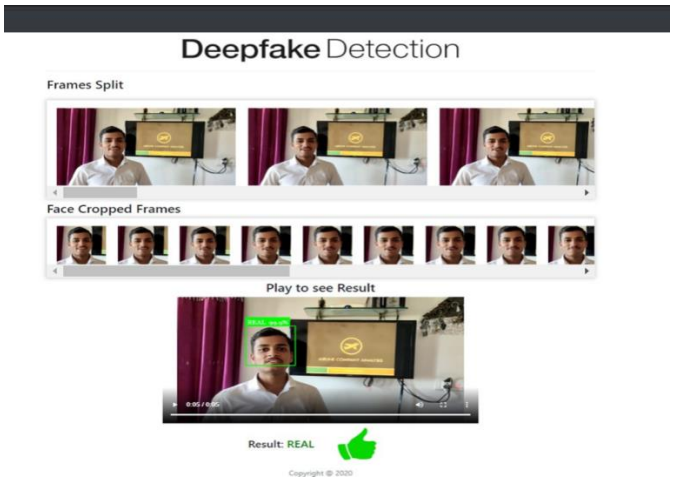


Fig (b). Uploading a real video



Fig(c).Real Video Output

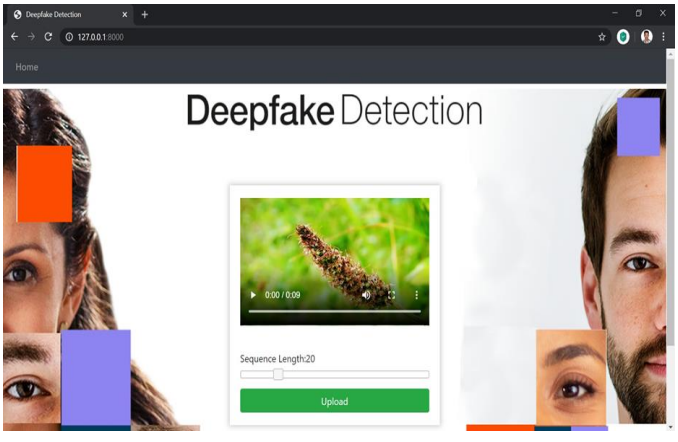


Fig (d).Output of Uploaded video with no faces

10.EXPERIMENTAL EVALUATION:

10.1 Experimental Data Sets

The following datasets are utilised in the experiments performed to assess the effectiveness of the proposed technique. We have collected information from a variety of publically accessible data-sets, including Face Forensic++ (FF), Deepfake detection challenge (DFDC), and Celeb-DF. To facilitate precise and instantaneous video content recognition, we have also combined the collected datasets to produce our own unique dataset. We have taken into account 50% real and 50% false videos in order to prevent the model from being overtrained. Since the focus of this study is on video deepfakes, the Deep Fake Detection Challenge (DFDC) dataset only comprises a subset of movies with audio alerts. We preprocessed the DFDC dataset by removing the audio-altered movies using a Python script. From the preprocessed DFDC dataset, we extracted 1500 authentic and 1500 sham videos. Among these are the FaceForensic++(FF) dataset, which has 2,000 films, 1,000 of which are fake, and the Celeb-DF dataset, which contains 500 real and 500 fake movies. This gives us a dataset of 6,000 films, of which 3,000 are authentic and 3,000 are shams.

10.2 Evaluation Metrics:

Accuracy metric is used to evaluate the performance of model over the different datasets.

10.3 Experimental Results and Analysis:

The accuracy of the model varies with the sequence length of the datasets i.e. the no of frames .

Dataset	Sequencelength	Accuracy
Face Forensic++	40	95.22613
Celeb-DF+Face Forensic++	100	93.97781
OurDataset	40	89.34681

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Where,

TN = True Positives
TF = True Negatives
FP = False Positives
FN = False Negatives

11.CONCLUSIONS:

Along with the confidence of the suggested model, we offered a neural network-based method for classifying videos as either deep fakes or actual videos. Our approach is capable of accurately predicting the output based on the processing of one second of video at a rate of ten frames per second. We built the model by making use of a pre-trained ResNext CNN model to extract the frame classification performance and LSTM for temporal sequence processing in order to identify the differences between the t frame and the t-1 frame. Our computational model is able to analyse the video in the frame sequence of 10, 20, 40, 60, 80, and 100.

12.REFERENCES:

[1] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies,Matthias Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images" in arXiv:1901.08971.

[2] Deepfake detection challenge dataset :https://www.kaggle.com/c/deepfake-detection-challenge/data Accessed on 26 March, 2020

[3] YuezunLi,XinYang,PuSun,HonggangQiandSiweiLyu" Celeb-DF:A Large-scaleChallenging Dataset for DeepFake Forensics" in arXiv:1909.12962

[4] Deepfake Video of Mark Zuckerberg Goes Viral on Eve of House A.I. Hearing : https://fortune.com/2019/06/12/deepfake-mark-zuckerberg/ Accessed on 26 March, 2020

[5] 10 deepfake examples that terrified and amused the internet : https://www.creativebloq.com/features/deepfake-examples Accessed on 26 March,2020

[6] TensorFlow:https://www.tensorflow.org/(Accessedon26 March,2020)

[7] Keras:https://keras.io/(Accessedon26March,2020)

[8] PyTorch:https://pytorch.org/(Accessedon26March,2020)

[9] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. arXiv:1702.01983, Feb. 2017

[10] J. Thies et al.Face2Face: Real-time face capture and reenactment of rgb videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2387–2395, June 2016. Las Vegas, NV.

BIOGRAPHIES

P.MEENAKSHI (18R01A05E7)

MADHUSMITA (18R01A05D9)

PASAM BHARGAVA SAI (18R01A05F1)