RESEARCH ARTICLE OPEN ACCESS

The Face-Off: Using Convolutional Neural Networks to Distinguish Real and Fake Faces

Shubham Dhamal,Vaishnavi Adsul, Priyanka Lokhande, Tushar Gavhane

Department of Information Technology

SVPM College of Engineering Baramati, Maharashtra, India

Email : shubhamdhamalsd@gmail.com, Vaishnaviadsul0705@gmail.com, lokhandepriyanka40@gmail.com, tusharg801@gmail.com

--************************----------------------------------

Abstract:

The creation of technology that can produce Deepfake videos is now advancing quickly. In fact, social networks frequently use compressed videos, including those from Twitter, Facebook, and Instagram. Understanding how to recognise compressed Deepfake videos is therefore essential. These movies can be made quickly, but doing so has negative consequences on one's reputation as well as the reputation of their business and contributes to financial loss. As a result, we suggested a detection method that can distinguish between actual and false videos. In order to train the convolutional neural network and categorise the films into real and fraudulent, we will use data from both actual and fake videos throughout this study. For the picture and video dataset, CNN provides greater accuracy. In this study, we will create a web application that accepts video as an input, pre-processes it using the CNN model, and outputs the final classification.

Keywords: - *Video forensics, Convolution Neural Network, Cyber Forensics, compressed Deepfake videos.*

--************************----------------------------------

1. INTRODUCTION

Most of the Internet traffic has shifted away from text sites and towards multimedia assets over the past ten years. Additionally, the emergence of large-scale multimedia social media platforms like WeChat, Instagram, and Snapchat has significantly changed our lives. It can not only improve people's lives but also give them the opportunity to communicate their experiences in more useful ways. Several uses for multimedia data are possible thanks to developments in video generating technologies. It improves social interaction, entertainment, and artistic expression but also jeopardises political stability, public safety, and individual privacy. The indistinguishability of digital media is considerably increased by the use of AI technology and forgery techniques [1].These films demonstrate how Obama has been tampered. The faces in Fig. 1 to the left are fakes,

and it is challenging to spot anything unusual with the unaided eye. Additionally, Online, there are false Obama-related videos can be found to make misleading claims.

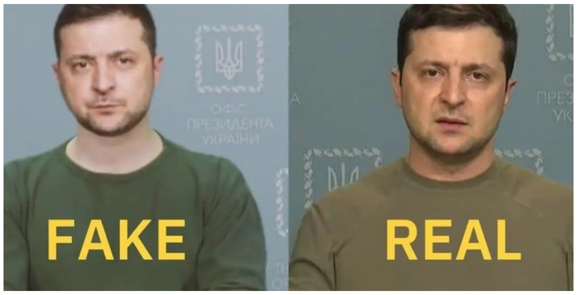


Fig 1. An illustration of a fake image (right) created using the Deepfakes method (left).

Due to the rapid dissemination of fraudulent information on social media, the impact of counterfeit information can suddenly grow by a factor of 10 million. The fake Obama movie's material has the potential to deceive viewers, which is bad for politics. A serious problem of public confidence is exacerbated by the introduction of counterfeit technologies, which also breeds public distrust. Furthermore, it can result in the disclosure of confidential information, telecommunications fraud, and harm to social justice.

In social networks today, compressed movies are frequently used. The uncompressed videos require a lot of storage, but our gadget only has a small amount of memory. Furthermore, the transmission speed of high-definition videos will be quite slow in the absence of a high network bandwidth [2]. When a user uploads a video to Instagram in social media, Instagram will compress the video. The size limitations of social media platforms like WeChat and Instagram force users to compress and reupload their films in order to send them. It will be challenging for us to identify the false video if thieves purposefully propagate compressed fake videos. The forensic of compressed Deepfake films becomes a crucial topic in order to address the issue of "seeing is not believing".

It becomes essential to determine the difference between the deepfake and the real video. We are employing AI to fight AI. Deep fakes are built with the use of pre-trained neural networks like GAN [3] or auto encoders utilising programmes like FaceApp and Face Swap. An LSTM-based artificial neural network [4] is used in our method to conduct the sequential temporal analysis of the video frames, and a pre-trained CNN extracts the frame-level properties. In order to evaluate whether a video is Deepfake or real, a convolution neural network first collects the frame-level properties. It then uses these attributes to train an artificial recurrent neural network based on long short-term memory.

It is necessary to simulate real-world scenarios in order to enhance the model's performance on real-time data. In order to recreate real-world situations and enhance the model's performance on real-world data, we trained our technique utilising a large variety of balanced and combinations of many available datasets, such as FaceForensics++ [5], Deepfake detection challenge, and Celeb-DF.

2. RELATED WORK

We explore a few conventional and deep learning methods for producing Deepfake videos. Here are some introductions to earlier, related studies on Deepfake videos and video-based digital media forensics.

2.1 DIGITAL MEDIA FORENSIC FOR VIDEO

Video-based Digital Media Forensics proposed a deep-learning network based on recompression error as a detection method for fake bitrate videos. techniques for identifying manipulation in interlaced and deinterlaced footage, as well. They evaluated the correlations that the deinterlacing methods used by the camera or software introduced into the deinterlaced footage. They offered a useful technique for determining motion in an interlaced video frame's field and between fields of surrounding frames. However, Deepfake movies cannot be located using these video-based forensics methods. Since no elements are removed, duplicated, or moved during the creation of the Deepfake videos, they are challenging to spot.

2.2 COMPRESSED DEEPFAKE VIDEOS ANALYSIS

Here, we have a look at the compressed Deepfake films at the temporal and frame levels. Compressed videos typically make some artefacts at the frame level worse when compared to high-definition videos. It was common practise to create temporal inconsistencies between frames in Deepfake movies.

3. LITERATURE SURVEY

The massive growth and criminal application of the deep fake video industry pose a severe danger to democracy, justice, and public trust. As a result, there is a greater requirement for fraudulent video analysis, detection, and response. A few words associated with deep fake detection are listed below:

Face Warping Artifacts in DF Videos [6] A method to detect artefacts was created by comparing the generated face areas and their surrounding regions with a particular Convolutional Neural Network model. In this investigation, there were two different kinds of facial artefacts.

Their strategy is based on the observation that the current DF technique can only provide images with a finite resolution, necessitating further transformation in order to match the faces that need to be replaced in the video.

Exposing AI Created Fake Videos by Detecting Eye Blinking [7] describes a novel method for revealing fake face videos created with deep neural network models. The method is based on recognising eye blinking, a physiological signal that is ineffectively displayed in artificially produced phoney videos. When used to identify movies made with Deep Learning software (DF), the method performs well when tested against benchmark datasets for eye-blinking detection.

The biological signals are extracted from facial regions on pairs of actual and fake portrait videos using the Biological Signals Approach for Synthetic Portrait Video Detection [8] technique. Transform feature sets [9] and PPG maps to extract the signal attributes, train a probabilistic SVM and a CNN, and analyse the spatial coherence and temporal consistency. The aggregate authenticity probabilities can then be used to decide whether the video is authentic or not.

No matter the generator, content, resolution, or video quality, Fraudulent Catcher can accurately detect fraudulent content. Since there is no discriminator, it is difficult to develop a differentiable loss function that adheres to the suggested signal processing techniques, which leads in the loss in their discoveries to retain biological signals.

4. PROPOSED SYSTEM

Using CNNs (Convolutional Neural Networks), this suggested system will classify discretized displacement patterns into many classes, using half of the CNNs for face binary classification and the other half for calibration. Evaluation of its performance and acceptance in terms of security, user friendliness, correctness, and reliability are one of the key goals.All Deepfakes, including interpersonal, replacement, and retrenchment Deepfakes, are targeted by our methodology.

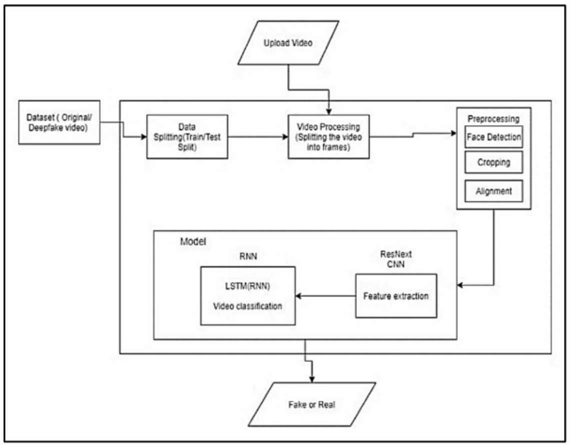


Fig. 2 Shows the Architecture of the System

4.1 DATASET

A mixed dataset that we are employing consists of an equal number of films from several dataset sources, such as FaceForensics++, YouTube, and the Deep fake detection challenge dataset [10]. Our most recent dataset contains 50% of the original video and 50% of the modified deepfake films. A test set makes up 30% of the dataset, and a train set makes up 70%.

4.2 DATA PREPROCESSING

As part of the pre-processing of the dataset, the video is split into frames. Next, a face is detected and the frame is cropped to include it. To preserve consistency in the number of frames, the mean of the dataset video is calculated, and a new processed face-cropped dataset is built using the frames that make up the mean. The frames that do not have any faces are ignored during pre-processing.

Therefore, we advise that the model be trained using just the first 100 frames for experimental purposes.

4.3 MODEL BUILDING

One LSTM [11] layer and then resnext50 32x4d make up the model. The data loader loads the pre-processed face-cropped videos and separates them into a train set and a test set. The modified video frames are also given to the model for training and testing in tiny batches.

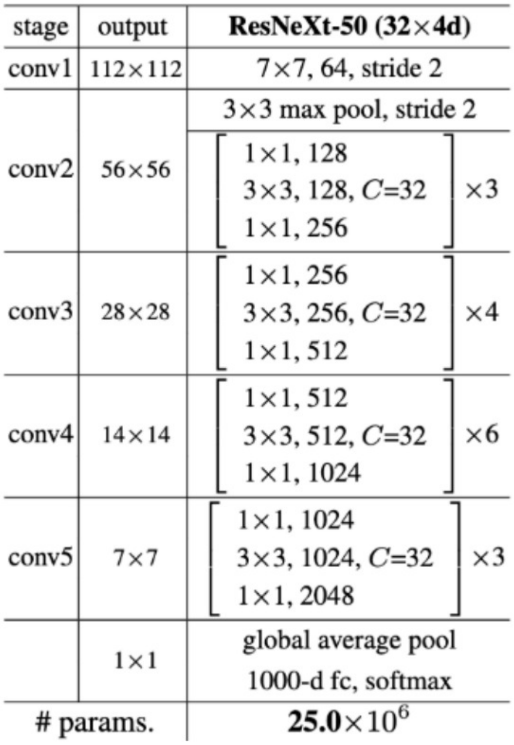
The model is composed of the following layers:

• LSTM Layer:

The LSTM layer is employed to examine sequences and spot temporal shifts between frames. The LSTM is fed fitted 2048-dimensional feature vectors. We are using a single LSTM layer with 2048 latent dimensions, 2048 hidden layers, and a 0.4 likelihood of dropout to achieve our objective. To do a temporal analysis of the video, the frames are processed sequentially using LSTM by contrasting the frame at second t with the one at second t-n. where n represents the number of frames prior to frame t.

• ResNext CNN:

This method uses a Residual Convolution Neural Network model that has already been trained. The model is known as resnext50 32x4d (). This model has 50 layers and 32 x 4 dimensions. Figure illustrates in detail how the model was applied.



stage	output	ResNeXt-50 (32×4d)	
conv1	112×112	7×7, 64, stride 2	
conv2	56×56	3×3 max pool, stride 2	
		1×1, 128 3×3, 128, C=32 1×1, 256	×3
conv3	28×28	1×1, 256 3×3, 256, C=32 1×1, 512	×4
conv4	14×14	1×1, 512 3×3, 512, C=32 1×1, 1024	×6
conv5	7×7	1×1, 1024 3×3, 1024, C=32 1×1, 2048	×3
	1×1	global average pool 1000-d fc, softmax	
# params.		**25.0**$\times 10^6$	

Fig. 3 ResNext Architecture

• Sequential Layer:

A sequential layer is a group of modules that can be operated concurrently and stacked on top of one another. In order to send the feature vector from the ResNext model to the LSTM sequentially, it is stored in an ordered way using a sequential layer.

•ReLU:

An activation function known as a "rectified linear unit" has a raw output in all other circumstances and an output of 0 when the input is less than 0. In other words, if the input is greater than 0, the output will be the same as the input. ReLU's functionality is more comparable to that of organic neurons. ReLU is non-linear and, in contrast to the sigmoid function, offers the advantage of having no backpropagation errors. For larger Neural Networks, ReLU-based models can be created rather quickly.

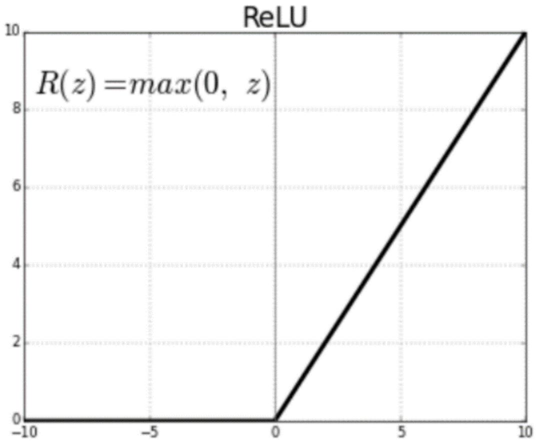


Fig. 4 ReLU Activation Function

• Dropout Layer:

The Dropout Layer, with a value of 0.4, is used to prevent overfitting in the model and can help with model generalisation by randomly setting the output for a particular neuron to 0. The cost function is more sensitive to surrounding neurons when the output is set to 0, which changes how the weights will be adjusted during the backpropagation procedure.

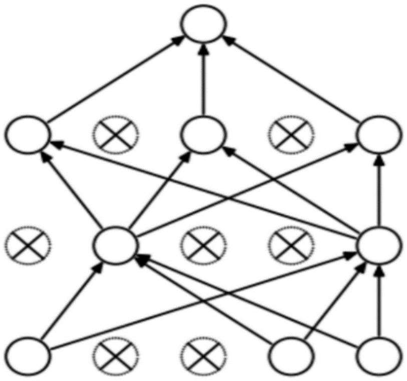


Fig. 5 Dropout Layer Overview

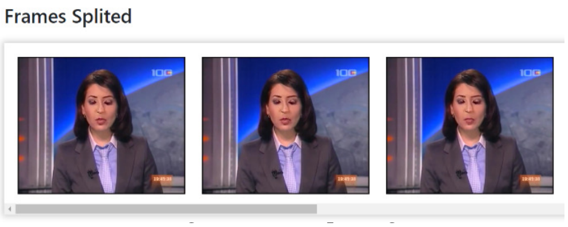
- **Adaptive Average Pooling Layer:** Low level neighbourhood data is gathered in order to lessen variation and computing complexity. A 2-dimensional Adaptive Average Pooling Layer is used in the model.

4.4 FEATURE EXTRACTION

Instead of building the classifier from scratch in order to extract the features, we advise using the CNN classifier for precisely detecting the frame level characteristics. In order to appropriately converge the gradient descent of the model, the network will then be fine-tuned by adding any extra layers that are required and selecting an acceptable learning rate. The 2048-dimensional feature vectors make up the sequential LSTM input layer after the final pooling layer.

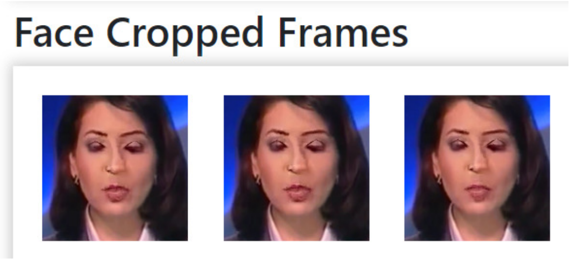
5 RESULT

A new video is given to the trained model for prediction.A raw video is additionally pre-processed to incorporate the format of the learned model. The video is splitting in number of frames according to sequence number provided by user.



Frames Splited

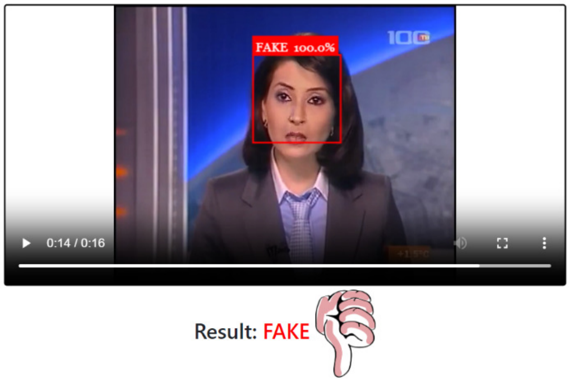
After the video is split into frames, the face is cropped.



Face Cropped Frames

Fig 07 Face Cropped Frame

After the cropping the cropped frames are given directly to the trained model for detection rather than being stored locally.



Result: FAKE

Fig. 08Result

6. LIMITATIONS

We did not use our method for the audio. Because of this, our method is unable to identify the audio deep fake. However, in the future, we advocate for identifying audio deep fakes.

7. CONCLUSION

In our study, we have introduced a new technique for detecting deep fake videos using a neural network-based method that includes measuring the confidence level of the model. The method is based on the generation of deep fake videos using GANs and Autoencoders. To detect deep fakes at the frame level, we have employed the use of CNN, and for analysing the entire video, we have utilized RNN and LSTM. By examining a set of parameters, our proposed method can determine whether a video is a deep fake or not with a high level of accuracy. We expect that the method will produce real-time results that are exceptionally reliable and consistent.

8. ACKNOWLEDGMENT

We would like to express our gratitude to **Prof. R. V. Chatse**, our BE Project Guide, whose valuable suggestions and support greatly contributed to the completion of this paper. We would also like to thank **Dr. Gawade J.S.**, Head of Department, and Honourable Principal**Dr. Mukane S.M.** for providing us with the opportunity and resources to undertake this project.

9. REFERENCES

[1] Luisa Verdoliva, "Media forensics and DeepFakes: An overview," IEEE Signal Processing, vol. 14, no. 5, pp. 910-932, 2020.

[2] C. Smansub, BoonchanaPurahong, P. Sithiyopasakul, C. Benjangkaprasert "A study of network bandwidth management by using queue tree with per connection queue" in April 2019Journal of Physics Conference Researchgate

[3] K. RemyaRevi, Vidya K R, M. Wilscy "Detection of Deepfake Images Created Using Generative Adversarial Networks: A Review" in researchgate

[4] LoliBurgueño, Jordi Cabot, Sébastien Gérard "An LSTM-Based Neural Network Architecture for Model Transformations" Conference: 2019 ACM/IEEE 22nd International, researchgate

[5] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, "FaceForensics++: Learning to Detect Manipulated Facial Images", arXiv:1901.08971 https://github.com/ondyari/FaceForensics

[6] Shruti Agarwal; Hany Farid; Tarek El-Gaaly; Ser-Nam Lim "Detecting Deep-Fake Videos from Appearance and Behavior," IEEE International Workshop.

[7] A Neural Network Algorithm Analyzes Eye Blinking to Detect Fake Videos (with 95% accuracy!) https://www.analyticsvidhya.com/blog/2018/08/algorithm-analyzes-eye-blinking-detect-fake-video/

[8] UmurAybarsCiftci, ˙Ilke Demir, Lijun Yin "Detection of Synthetic Portrait Videos using Biological Signals" in arXiv:1901.02212v2.

[9] Temporal and Spacial Coherence https://www.fullonstudy.com/temporal-spatial-coherence

[10] Deepfake Detection challenge dataset on kagglehttps://www.kaggle.com/c/deepfake-detection-challenge/data

[11] Sepp Hochreiter, Jürgen Schmidhuber "Long Short-term Memory", December 1997Neural Computation, Reserchgate

[12]An Overview of ResNet and its Variants: https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035

[13] Long Short-Term Memory: From Zero to Hero withPytorch:https://blog.floydhub.com/long-short-term-memory-from-zero-to-hero-with-pytorch/

[14] Sequence Models And LSTM Networks https://pytorch.org/tutorials/beginner/nlp/sequence_mod els_tutorial.html

[15] https://discuss.pytorch.org/t/confused-about-the-image-preprocessing-in-classification/3965

[16] Y. Qian et al. Recurrent color constancy. Proceedings of the IEEE International Conference on Computer Vision, pages 5459–5467, Oct. 2017. Venice, Italy.

[17] Nicolas Rahmouni, Vincent Nozick, Junichi Yamagishi, and Isao Echizen, "Distinguishing computer graphics from natural

images using convolution neural networks," in WIFS. IEEE, 2017.

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

[19] D. E. King, "Dlib-ml: A machine learning toolkit,"JMLR, vol. 10, pp. 1755–1758, 2009