Detect the Fake Faces from Images and Videos Using Convolutional Neural Network

Shubham Dhamal[1] Vaishnavi Adsul[2] Tushar Gavhane[3] Priyanka Lokhande[4]
[1,2,3,4]Department of Information Technology
[1,2,3,4]SVPM College of Engineering Baramati, Maharashtra, India

Abstract— Nowadays, the development of technologies that can generate Deepfake videos is expanding rapidly. In fact, compressed videos are common in social networks, such as videos from Instagram, Facebook and Twitter. Therefore, learning how to spot compressed Deepfake videos becomes crucial. These videos can be easy to create but due to this it effects on Personal reputation, Business Reputation and aids to financial loss. Therefore, we proposed a detection system which identifies the real video and fake. Throughout this paper we will use real and fake videos data to train the Convolutional neural network and classify their categories into real and fake. CNN gives better accuracy on image and video dataset. In this paper, we will be building a web application that takes image or video as input and pre-process that with help of CNN model and gives the final classification as an output.

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

I. INTRODUCTION

Within the last ten years, the majority of Internet traffic has moved away from text pages and toward multimedia assets. Additionally, the rise of expansive multimedia social media platforms like Instagram, Snapchat, and WeChat has huge alterations in our lives. It can not only make people's lives better but also enable them to share their experiences in more practical ways. Multimedia data can be used for a variety of purposes as a result of advancements in video generating technology. In addition to enhancing entertainment, artistic expression, and social connection, it also threatens political stability, public safety, and individual privacy AI technology combined with forgery tactics enhances the indistinguishability of digital media significantly [1]. The faces in Fig. 1 to the left are forged, and with the naked eye, it is difficult to detect anything strange. Additionally, fake videos regarding Obama can be found online. These films demonstrate how Obama has been tampered

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

with to make misleading claims. The impact of forged information can instantly increase by a factor of 10 million times to the spread of false information through social media. Viewers of the bogus Obama film can be duped by its contents, which has a bad impact on politics. Additionally, the emergence of counterfeit technology increases public mistrust and contributes to a significant crisis of public confidence. Additionally, it might lead to the revelation of private information, telecommunications fraud, and social justice harm.

Nowadays, compressed videos are widely used in social networks. The reason is that the uncompressed videos take up a lot of storage, but our device has limited memory. Furthermore, if there is no high network bandwidth [2], the transmission speed of high-definition videos will be quite slow. In social media, when a user uploads a video to Instagram, the video will be compressed by the Instagram. If a user sends a video with social software such as WeChat and Instagram, the size of the videos is so restrictive that the users have to compress it and upload it again. If criminals deliberately spread compressed fake videos, it will make it difficult for us to detect the forgery video. In order to solve the problem that seeing is not believing, the forensic of compressed Deepfake videos becomes an important issue.

Finding the difference between the deepfake and the authentic video becomes crucial. To combat AI, we are using AI. Tools like FaceApp and Face Swap are used to build deep fakes, which are created using pre-trained neural networks like GAN [3] or auto encoders. Our approach processes the sequential temporal analysis of the video frames using an LSTM-based artificial neural network [4], and a pre-trained CNN extracts the frame-level characteristics. Convolution neural network captures the frame-level characteristics, and then uses these features to train an artificial Recurrent Neural Network based on Long Short-Term Memory to determine whether the video is Deepfake or real.

In order to improve the model's performance on real-time data, real-world scenarios must be simulated. We trained our technique using a vast number of balanced and combination of multiple available datasets including FaceForensics++ [5], Deepfake detection challenge, and Celeb-DF in order to simulate real time scenarios and improve the model's performance on real time data.

II. RELATED WORK

For the purpose of creating Deepfake videos, we discuss a few traditional and deep learning techniques. Here are some introductions to earlier related works on video-based digital media forensics and Deepfake videos.

A. Digital Media Forensic For Video

Video-based A deep-learning network based on recompression error was suggested by Digital Media Forensics as a detecting tool for bogus bitrate videos. Additionally, methods for spotting signs of tampering in interlaced and deinterlaced footage. They measured the correlations that the camera or software deinterlacing techniques introduced into deinterlaced footage. They

suggested a practical method for measuring motion inside a single frame's field and between fields of neighbouring frames in interlaced video. These video-based forensics techniques, however, cannot be utilised to find Deepfake videos. The Deepfake movies are difficult to detect since they are created using artificial intelligence without the removal, duplication, or movement of any items.

B. Compressed Deepfake Videos Analysis

Here, we examine the compressed Deepfake videos from the frame- and temporal-level perspectives. When compared to high-definition videos, compressed videos mostly worsen some artefacts at the frame level. When making Deepfake films, temporal discrepancies between frames would be produced at the temporality level.

III. LITERATURE SURVEY

Democracy, justice, and public trust are seriously threatened by the deep fake video industry's tremendous expansion and criminal usage. The need for fake video analysis, detection, and action has increased as a result. The following is a list of some words that are connected to deep fake detection:

Finding Face Warping Artifacts [6] in DF Videos By comparing the generated face areas and their surrounding regions with a specific Convolutional Neural Network model, developed a method to identify artefacts. There were two types of face artefacts in this work.

Their approach is based on the observation that the present DF technique can only produce images with a finite resolution, which then requires additional transformation to match the faces that need to be replaced in the video.

A novel technique for exposing false face videos made with deep neural network models is described in Exposing AI Created Fake Videos by Detecting Eye Blinking [7]. The technique is based on the identification of eye blinking, a physiological signal that is poorly displayed in synthetically created fake videos. The method is tested against benchmark datasets for eye-blinking detection and exhibits good results when it comes to identifying films produced using Deep Learning software (DF).

The Biological Signals Approach for Synthetic Portrait Video Detection [8] method extracts biological signals from facial areas on pairs of real and false portrait videos. Utilize transformations to train a probabilistic SVM and a CNN, determine the spatial coherence and temporal consistency, and capture the signal properties in feature sets [9] and PPG maps. Afterward, use the overall authenticity probabilities to determine whether the video is real or not.

With excellent accuracy and regardless of the generator, content, resolution, or video quality, Fraudulent Catcher can identify fake content. It is not easy to create a differentiable loss function that follows the suggested signal processing procedures since there is no discriminator, which results in the loss in their discoveries to preserve biological signals.

IV. PROPOSED SYSTEM

In this proposed system we will be using CNN (Convolutional Neural Network) in which half CNN's out of them for facial binary classification and half CNNs for calibration, this is formulated as multiclass classification of discretized displacement pattern.

Evaluating its performance and acceptance in terms of security, user friendliness, accuracy, and reliability is one of the key goals.

Our approach focuses on identifying all types of Deepfakes, including interpersonal, replacement, and retrenchment Deepfakes.

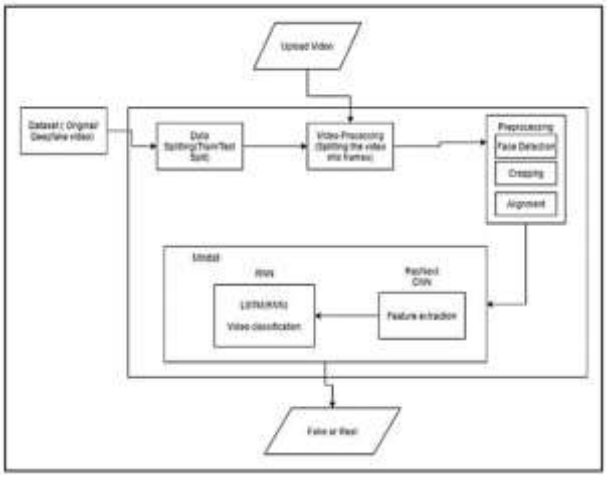

Fig. 2: Shows the Architecture of the System

A. Dataset

We are using a mixed dataset made up of an equal number of films from various dataset sources, including YouTube, FaceForensics++ , and the Deep fake detection challenge dataset [10]. 50% of the original video and 50% of the altered deepfake videos are included in our recently created dataset. The dataset is divided into a 30% test set and a 70% train set.

B. Data Preprocessing

The video is divided into frames as part of the dataset pre-processing procedure Face detection and cropping the frame to include the found face come next. The mean of the dataset video is determined in order to maintain consistency in the number of frames, and a new processed face-cropped dataset is constructed using the frames that make up the mean. Pre-processing ignores the frames that don't contain any faces.

Therefore, we are suggesting that for experimental purposes, the model be trained using only the first 100 frames.

C. Model Building

The model comprises of one LSTM [11] layer followed by resnext50 32x4d. The pre-processed face-cropped videos are loaded by the data loader, who divides them into a train set and a test set. Additionally, the model receives the frames from the edited videos for training and testing in small batches.

There are the following layers in the model:

– ResNext CNN: This technique makes use of a pre-trained Residual Convolution Neural Network model. Name of the model is resnext50 32x4d (). This model has 32 x 4 dimensions and 50 layers. Figure depicts how the model was implemented in detail.

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.		25.0×10^6

Fig. 3: ResNext Architecture

- Sequential Layer: A sequential layer is a collection of modules that can be stacked on top of one another and executed simultaneously. The feature vector provided by the ResNext model is stored in an ordered manner using a sequential layer. so that it can be delivered consecutively to the LSTM.
- LSTM Layer: This layer is used to analyse sequences and identify temporal changes between frames. Fitted 2048-dimensional feature vectors are used as the LSTM's input. In order to accomplish our goal, we are utilising a single LSTM layer with 2048 latent dimensions, 2048 hidden layers, and a 0.4 likelihood of dropout. The frames are processed sequentially using LSTM in order to do a temporal analysis of the video by comparing the frame at second t with the frame at second t-n. In which n is the number of frames before to t.
- ReLU: A Rectified Linear Unit is an activation function with a raw output in all other cases and an output of 0 if the input is less than 0. In other words, the output is identical to the input if the input is higher than 0. ReLU's operation is more similar to how organic neurons function. ReLU is non-linear and has the advantage of having no backpropagation mistakes, unlike the sigmoid function. It is also relatively quick to develop models based on ReLU for bigger Neural Networks.

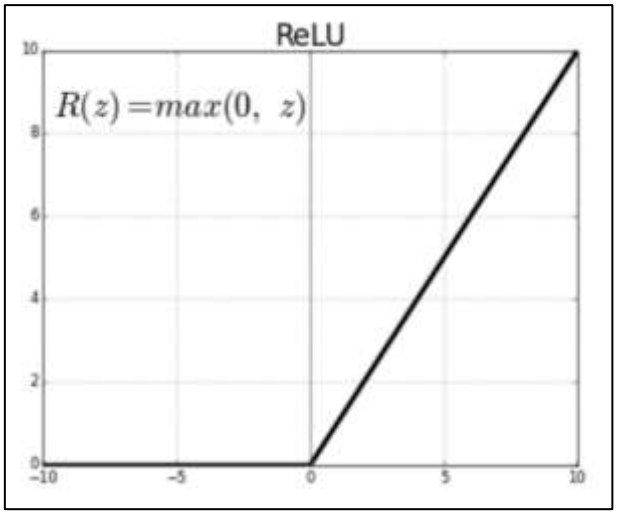

Fig. 4: ReLU Activation Function

- Dropout Layer: By randomly changing the output for a specific neuron to 0, the Dropout Layer, with a value of 0.4, is employed to prevent overfitting in the model and can aid in model generalisation. When the output is set to 0, the cost function becomes more susceptible to nearby neurons, which alters how the weights will be changed during the backpropagation process.

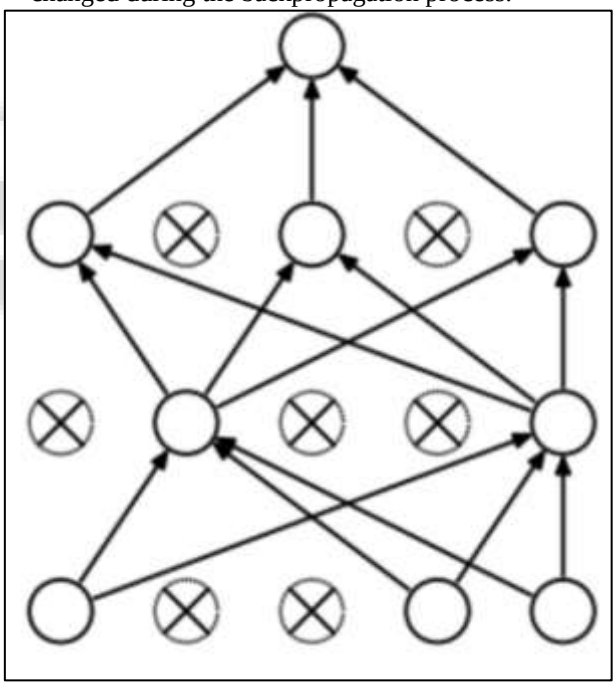

Fig. 5: Dropout Layer Overview

- Adaptive Average Pooling Layer: It is used to collect low level information from the neighbourhood and reduce variation and computation complexity. The model makes use of a 2-dimensional Adaptive Average Pooling Layer.

D. Feature Extraction

We suggest using the CNN classifier for accurately detecting the frame level characteristics rather than constructing the classifier from scratch in order to extract the features. The network will then be fine-tuned by adding any additional necessary layers and choosing an appropriate learning rate to properly converge the gradient descent of the model.

Following the last pooling layers, the sequential LSTM input is then made up of the 2048-dimensional feature vectors.

E. Expected Result

The trained model receives a new video for prediction. Additionally, a raw video is pre-processed to incorporate the trained model's format. Face cropping is done after the video is divided into frames, and the cropped frames are then provided straight to the trained model for detection rather than being stored locally.

Fig. 3: Expected Result

V. LIMITATIONS

We did not work for the audio in our method. Due to this, the audio deep fake cannot be detected by our approach. But in the future, we suggest achieving the identification of audio deep fakes.

VI. CONCLUSION

We provided a neural network-based method for determining if a video is a deep fake or the real thing, along with the model's level of confidence. The deep fakes produced by GANs with the aid of Autoencoders serve as an inspiration for the suggested strategy. Our approach uses CNN for frame-level detection and RNN and LSTM for video classification. Based on the factors stated in the study, the suggested method is capable of determining if a video is a deep fake or real. We think it will deliver real-time data with remarkably high accuracy. Research work in running paragraphs. Some content related to your research work in running paragraphs. Some content related to your research work in running paragraphs. Some content related to your research work in running paragraphs.

REFERENCES

[1] Luisa Verdoliva, "Media forensics and DeepFakes: An overview," IEEE Signal Processing, vol. 14, no. 5, pp. 910-932, 2020.
[2] C. Smansub, Boonchana Purahong, 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. Remya Revi, Vidya K R, M. Wilscy "Detection of Deepfake Images Created Using Generative Adversarial Networks: A Review" in researchgate
[4] Loli Burgueño, 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] Umur Aybars Ciftci, ˙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 kaggle https://www.kaggle.com/c/deepfake-detection-challenge/data
[11] Sepp Hochreiter, Jürgen Schmidhuber "Long Short-term Memory", December 1997 Neural 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 with Pytorch: 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_models_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

