Archives available at journals.mriindia.com

International Journal of Recent Advances in Engineering and Technology

ISSN: 2347-2812
Volume 14 Issue 1s, 2025

Deepfake Video Detection using Neural Networks

Vaibhav Nangare[1], Bhosale Sachin2, Anand Khatri[3], Bhalchandra Mundhe[4]

[1,2,3,4]Computer Engineering, Jaihind College of Engineering, Kuran, India
vaibhavnangare5@gmail.com[1],sachinbhosale@gmail.com[2],khatrianand@gmail.com[3], mundheraj.mundhe@gmail.com[4]

Peer Review Information

Submission: 20 Jan 2025
Revision: 24 Feb 2025
Acceptance: 27 March 2025

Keywords

Recurrent Neural Networks
Convolutional Neural Networks
Deepfake Video Detection

Abstract

Recent developments in free deep learning software tools have made it easier to create realistic facial interactions in "DeepFake" (DF) videos, which are videos with minimal evidence of manipulation. Although the effective use of visual effects has been used to manipulate digital media for many years, recent developments in deep learning have drastically increased the realism of fake content and made it easier to make. DF stands for "AI-synthesized media."

It is easy to create the DF with artificially intelligent tools. However, it is quite difficult to discover these DF because the algorithm is difficult to train to detect the DF. With the use of recurrent and convolutional neural networks, we have advanced the detection of the DF. Convolutional neural networks (CNNs) are used by the system to extract features at the frame level. A recurrent neural network (RNN) is trained using these features to determine whether a video has been altered and to identify the temporal discrepancies between frames that the DF generation tools introduce. A big collection of phony movies gathered from a standard data set was compared to the expected outcome.

We demonstrate how, with a straightforward architecture, our system may achieve competitive outcomes in this job.

INTRODUCTION

The rising sophistication of smartphone cameras and the global availability of reliable internet connections have expanded social media's reach and made it easier than ever to create and share digital videos. Deep learning is now more powerful than it was a few years ago due to the increasing computing power.

As with every revolutionary invention, this has brought up new difficulties. "DeepFake" is a term used to describe the manipulation of audio and video footage by deep generative adversarial models. It is now increasingly usual for the DF to spread on social media, which encourages spam and the spread of false information. These DF members will be awful and will threaten and deceive the general public. [2]

DF detection is crucial to resolving such a scenario. In order to efficiently differentiate AI-generated false videos (DF Videos) from authentic videos, we provide a novel deep learning-based technique. The development

of technology that can detect fakes is crucial in order to identify the DF and stop it from propagating online.

The way the Generative Adversarial Network (GAN) generates the DF is crucial to understanding its detection. A GAN receives as input a video and a picture of a particular person (the "target"), and then produces a second movie in which the faces of the target are swapped out with those of a different person (the "source"). Deep adversarial neural networks that have been trained on target movies and face images form the foundation of DF. These networks automatically translate the source's faces and facial expressions to the target.

High levels of realism can be achieved in the final videos with the right post-processing. The input image is changed in each frame once the GAN splits the movie into frames. It also reconstructs the video. Usually, autoencoders are used to accomplish this operation. We provide a brand-new deep learning technique that successfully separates DF films from authentic ones. Our approach is based on the same methodology that GAN uses to generate the DF.

Based on the characteristics of the DF movies, the DF algorithm can only create face pictures of a specific size as a result of production time and computation resource limitations. These images must then go through affinal warping to fit the source's facial configuration. The resolution discrepancy between the warped face area and the surrounding context causes some recognizable artifacts to remain in the deepfake video output.

In order to identify these artifacts, we first split the video into frames, then we use a ResNext Convolutional Neural Network (CNN) to extract the features. Next, we use a Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM) to capture the temporal inconsistencies between frames that the GAN introduced during the reconstruction of the DF. By explicitly modeling the resolution discrepancy in affine face wrappings, we streamline the process of training the ResNext CNN model. [4]

LITERATURE SURVEY

Deepfake video's rapid proliferation and illicit use pose a serious danger to justice, democracy, and public confidence. The need for bogus video analysis, detection, and action has grown as a result. The following is a list of some terms that are related to deepfake detection: By comparing the generated face areas and its surrounding regions with a specialized Convolutional Neural Network model, ExposingDF Videos by Detecting Face Warping Artifacts [1] employed a method to identify artifacts. There were two types of face artifacts in this piece.

Based on the fact that the present DF technique can only produce images with a limited resolution, their method requires further transformation to match the faces that need to be replaced in the original movie. The article Exposing AI Created false Videos by Detecting Eye Blinking [2] outlines a novel technique for exposing deep neural network-generated false face videos. Eye blinking, a natural indication that is poorly represented in the artificially created phony videos, is the basis for the method's identification.

In tests using benchmarks of eye-blinking detection datasets, the approach demonstrates encouraging results in identifying films produced by the Deep Neural Network-based program DF.

Their approach solely relies on the absence of blinking as a detection clue. To detect the deepfake, however, a few additional factors need to be taken into account, such as facial wrinkles and tooth enchantment. We suggest using our method to take all of these factors into account.

In order to identify modified and forged images and videos in many contexts, such as replay attack detection and computer-generated video detection, a technique known as "using capsule networks to detect forged images and videos" [3] can be applied.

The training portion of their approach uses random noise, which is not a good choice. Even if the model worked well on their dataset, noise in the training process could cause it to perform poorly on real-time data. Our approach is suggested to be trained on real-time and noiseless datasets. Biological signals from facial areas in real

and false portrait video pairings are extracted utilizing the Detection of Synthetic Portrait Videos using Biological Signals [5] technique. To calculate temporal consistency and spatial coherence, capture signal properties in feature sets and PPG maps, and train a CNN and a probabilistic SVM, apply transformations. After that, the total authenticity probabilities are used to determine if the video is real or not.

Fraudulent Catcher accurately identifies fraudulent content regardless of the video's source, content, resolution, or quality. The procedure of creating a differentiable loss function that adheres to the suggested signal processing processes is not simple because the absence of a discriminator hindered their ability to maintain biological signals.

PROPOSED SYSTEM

There are numerous tools available for DF creation, but very few are accessible for DF detection. Our method of DF detection will be very helpful in preventing the DF from spreading throughout the internet. We will offer a web-based platform where users may post videos and determine if they are authentic or not. From creating a web-based platform to creating a browser plugin for automatic DF detections, this project can be expanded. This solution may be integrated with popular apps like Facebook and WhatsApp to easily detect DF before sending it to another user. Evaluating its performance and acceptance in terms of security, usability, correctness, and dependability is one of the key goals.

Our approach focuses on identifying all forms of DF, including interpersonal, retrenchment, and replacement DF. The suggested system's basic system architecture is shown in Figure 1. -

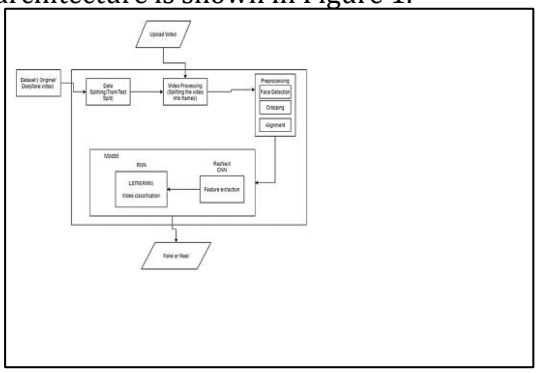

Fig. 1: System Architecture

Dataset:

We are working with a mixed dataset, which includes an equal number of videos from various dataset sources, such as FaceForensics++ [14], YouTube, and the Deep Fake Detection Challenge dataset [13]. Half of the original video and half of the altered deepfake videos are included in our recently created dataset. 30% of the dataset is used for testing, while 70% is used for training.

Preprocessing:

As part of the preprocessing of the dataset, the video is divided into frames. After that, the face is detected, and the frame with the detected face is cropped. In order to ensure consistency in the quantity of frames, the average of the video dataset is determined, and a new processed face cropped dataset is produced with the mean number of frames. The preprocessing technique ignores the frames that contain no faces.

A significant amount of computing power will be needed to process the 10-second movie at 30 frames per second, or 300 frames altogether. Therefore, for experimental purposes, we suggest training the model using just the first 100 frames.

Model:

The model is made up of one LSTM layer after resnext50_32x4d. The preprocessed face-cropped films are loaded by the data loader, which also separates the videos into train and test sets. Additionally, the model receives the frames from the processed videos in small batches for training and testing.

ResNext CNN for Feature Extraction

For feature extraction and precise frame-level feature detection, we are suggesting using the ResNext CNN classifier rather than building the classifier again. After that, we'll be fine-tuning the network by adding the necessary extra layers and choosing the right learning rate to ensure that the model's gradient descent converges correctly. The final pooling layers' 2048-dimensional feature vectors are then fed into the sequential LSTM.

LSTM for Sequence Processing

Assume that a sequence of ResNext CNN feature vectors of input frames is fed into a

2-node neural network. The probability of the sequence indicate if it is part of a genuine or deepfake movie. Our biggest challenge is to develop a model that can handle a sequence recursively in a meaningful way. To achieve our objective, we propose to solve this problem with a 2048 LSTM unit with a 0.4 dropout probability. By comparing the frame at "t" second with the frame at "t-n" seconds, the temporal analysis of the video may be performed by using LSTM to process the frames sequentially. where the number of frames preceding t, n, might be any value.

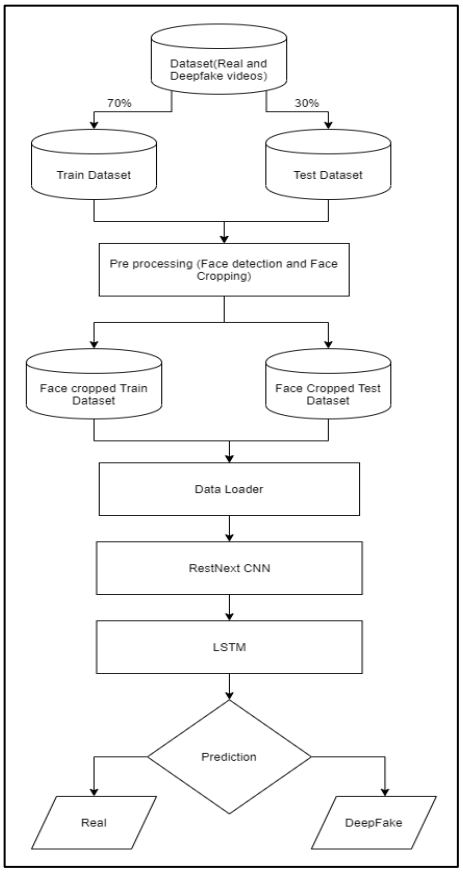

Fig. 2: Training Flow

Fig. 3: Expected Results

Predict:
A new video is passed to the trained model for prediction. A new video is also preprocessed to bring in the format of the trained model. The video is split into frames followed by face cropping and instead of storing the video

into local storage the cropped frames are directly passed to the trained model for detection.

RESULT
The output of the model is going to be whether the video is deepfake or a real video along with the confidence of the model. One example is shown in the figure 3.

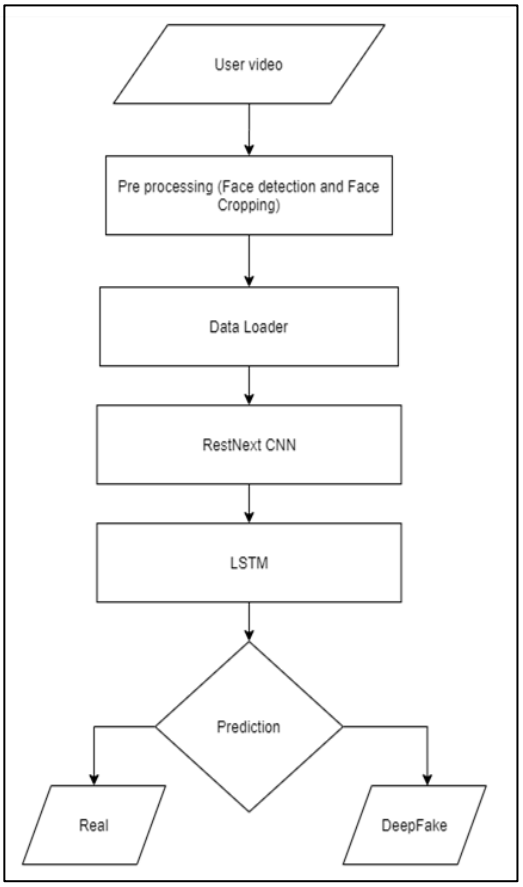

Fig. 4: Prediction flow

CONCLUSION

We demonstrated a neural network-based method for determining if a video is real or deepfake, as well as the suggested model's level of confidence. The suggested approach draws inspiration from how GANs use autoencoders to produce deepfakes. Our approach uses ResNext CNN for frame-level detection and RNN and LSTM for video classification. Using the criteria mentioned in the study, the suggested approach can determine whether the video is real or a deepfake. We think it will give real-time data a very high level of precision.

LIMITATIONS

The audio has not been accounted for in our method. Our approach won't be able to identify the audio deepfake because of this. However, we are putting up the idea of detecting audio deepfakes in the future.

References

Yuezun Li, Siwei Lyu, “ExposingDF Videos By Detecting Face Warping Artifacts,” in arXiv:1811.00656v3.

Yuezun Li, Ming-Ching Chang and Siwei Lyu “Exposing AI Created Fake Videos by Detecting Eye Blinking” in arxiv.

Huy H. Nguyen , Junichi Yamagishi, and Isao Echizen “ Using capsule networks to detect forged images and videos ”.

Hyeongwoo Kim, Pablo Garrido, Ayush Tewari and Weipeng Xu “Deep Video Portraits” in arXiv:1901.02212v2.

Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin “Detection of Synthetic Portrait Videos using Biological Signals” in arXiv:1901.02212v2.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In NIPS, 2014.

David G¨uera and Edward J Delp. Deepfake video detection using recurrent neural networks. In AVSS, 2018.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016.

An Overview of Reset
and its Variants
: https:

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

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

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

https://www.kaggle.com/c/deepfak
e-detection- challenge/data

https://github.com/ondyari/FaceForensics

Y. Qian et al. Recurrent color constancy.

Proceedings of the IEEE International Conference on Computer Vision, pages 5459–5467, Oct. 2017. Venice, Italy.

P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to- image translation with conditional adversarial networks.

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5967–5976, July 2017. Honolulu, HI.

R. Raghavendra, Kiran B. Raja, Sushma Venkatesh, and Christoph Busch, “Transferable deep-CNN features for detecting digital and print-scanned morphed face images,” in CVPRW. IEEE, 2017.

Tiago de Freitas Pereira, Andr´e Anjos, Jos´e Mario De Martino, and S´ebastien Marcel, “Can face anti spoofing countermeasures work in a real world scenario?,”in ICB. IEEE, 2013.

Nicolas Rahmouni, Vincent Nozick, Junichi Yamagishi, and Isao Echizen, “Distinguishing computer graphics from natural images is using convolution neural networks,” in WIFS. IEEE, 2017.

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.
D. E. King, “Dlib-ml: A machine learning toolkit,” JMLR, vol. 10, pp. 1755–1758, 2009.