International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Detection of Deepfake Videos Using Transfer Learning

Manish Assudani[1], Chaitanya Gedam[2], Nilay Gajhbiye[3] Ishika Dorlikar[4], Chaitanya Sakhare[5]

[1] *Department of Computer Science & Engineering*
Anjuman College Of Engineering & Technology
Nagpur, India
[2] *Department of Computer Science & Engineering*
Anjuman College Of Engineering & Technology
Nagpur, India
[3] *Department of Computer Science & Engineering*
Anjuman College Of Engineering & Technology
Nagpur, India
[4] *Department of Computer Science & Engineering*
Anjuman College Of Engineering & Technology
Nagpur, India
[5]*Department of Computer Science & Engineering*
Anjuman College Of Engineering & Technology
Nagpur, India

Abstract - Deep learning algorithms have become so potent due to increased computing power that it is now relatively easy to produce human-like synthesised videos, sometimes known as "deep fakes." One may easily imagine scenarios where these realistic face swap deep fakes are used for negative purposes. In this paper, we provide a novel deep learning-based strategy for the efficient separation of fraudulent videos produced by AI from real videos. Automatically spotting replacement and recreation deep fakes is possible with our technology. To combat artificial intelligence, we are attempting to deploy artificial intelligence (AI). The frame-level features are extracted by our system using a Res-Next Convolution neural network, and these features are then used to train an LSTM-based recurrent neural network (RNN) to determine whether the video has been altered in any way or not, i.e. whether it is a deep fake or authentic video. We test our technique on a sizable, balanced, and mixed data set Deepfake detection challenge[1], in order to simulate real-time events and improve the model's performance on real-time data. We also demonstrate a very straightforward and reliable approach that allows our system to produce results that are competitive.

Key Words: Neural network with Res-Next technology.

Recurrent neural network (RNN).

Long vs. short-term memory (LSTM).

Computer Vision

1.INTRODUCTION

In our system, we have employed a PyTorch deepfake detection model that has been trained on an equal number of real and fake videos. This approach ensures that the model is not biased towards one type of video. The system architecture of our model is illustrated in the accompanying figure. During the development phase, we took a dataset and performed preprocessing steps to create a new processed dataset. This processed dataset specifically includes videos where the faces have been cropped. The frame-level features are extracted by our system using a Res-Next Convolution neural network, and these features are then used to train an LSTM-based recurrent neural network (RNN) to determine whether the video has been altered in any way or not, i.e. whether it is a deep fake or authentic video.

2. Literature Survey

Using a specific Convolutional Neural Network model, ExposingDF Videos by Identifying Face Warping Artifacts [1] employed a method to identify artefacts by comparing the generated face areas and its surrounding regions. Face Artifacts appeared twice in this work.Their approach is based on the observation that the present DF algorithm can only produce images of a certain resolution, which then requires additional transformation to match the faces to be replaced in the source video.

A novel technique for exposing false face videos made with deep neural network models is described in Revealing AI Generated Fake Videos by Detecting Eye Blinking [2]. The technique is based on the identification of eye blinking, a physiological signal that is poorly displayed in synthetically created phoney films. 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). Only the absence of blinking is used by their method as a detection hint. But, additional factors like teeth enchantment, facial wrinkles, etc. must be taken into account for the detection of the deep fake. All these factors are to be taken into account when using our strategy.

Using capsule networks to detect forged images and videos [3]employs an approach that employs a capsule network to detect forged, manipulated photos and videos in a variety of circumstances, including replay attack detection and computergenerated video detection. They used random noise

in the training phase of their approach, which is not a suitable solution. Even though the model performed well on their dataset, it may fail on real-time data due to noise in training.

Detection of Synthetic Portrait Videos using Biological Signals [5] 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, compute the spatial coherence and temporal consistency, capture the signal properties in feature sets and PPG maps, and compute the spatial coherence and temporal consistency. 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 fraudulent 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.

Deepfake Video Detection Using Neural Networks [22] utilised a similar method on a much larger dataset than ours, and we experimented with a couple of different datasets and altered the parameters to obtain a workable project in a much less burdened system.

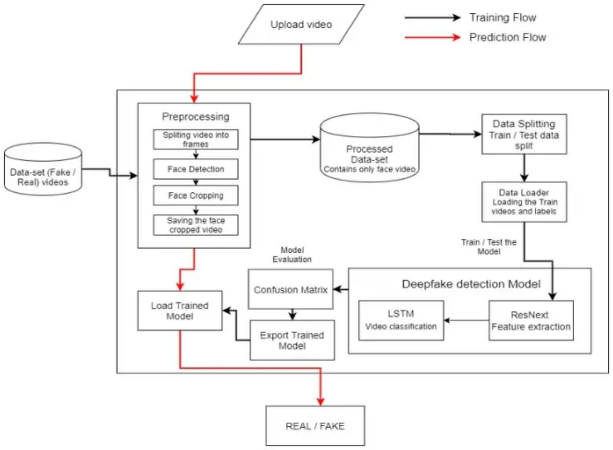
3. PROPOSED SYSTEM

Although there are numerous tools accessible for DF creation, there are very few ones available for DF detection. Our method for finding the DF will make a significant contribution to preventing the DF from spreading throughout the internet.

We'll offer a web-based platform where users may upload videos and mark them as bogus or authentic. From developing a web-based platform to a browser plugin for automatic DF detections, this project can be scaled up. Even popular apps like WhatsApp and Facebook can incorporate this project into their software for simple DF pre-detection prior to forwarding to another user. Reviewing 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 DF, including interpersonal, replacement, and retrenchment DF.

The suggested system's basic system architecture is shown in Figure 1.



A. Dataset:

We are using a mixed dataset made up of an equal number of films from various dataset sources, including the Deep fake detection challenge dataset[13].

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. Preprocessing:

The clip is divided into frames as part of the dataset preprocessing 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. Preprocessing ignores the frames that don't contain any faces. It will take a lot of computing power to process the 300 frames in a 10 second video at 30 frames per second. Therefore, we are suggesting that for experimental purposes, the model be trained using only the first 150 frames.

C. Model:

The model consists of one LSTM layer and then 50 32x4d resnext layers. The data loader loads the preprocessed face-cropped videos and separates them into a train set and a test set. Also, the model receives the frames from the altered videos in tiny batches for training and testing.

D. ResNext CNN for Feature Extraction

We suggest using the ResNext CNN classifier for properly recognising the frame level features 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.The sequential LSTM input is then made up of the 2048-dimensional feature vectors that remain after the final pooling layers.

E. LSTM for Sequence Processing

The key problem that needs to be solved is how to construct a model that meaningfully process a sequence recursively. To achieve our goal, we recommend a 2048 Long short - term memory unit with a 0.4 likelihood of dropout for this assignment .

F. Predict

The trained model receives a new video for prediction. Also, a fresh video is preprocessed to incorporate the trained model's format. After the footage is divided into frames, faces are cropped and the clipped frames are sent straight to the trained model for detection rather than being stored locally instead of in the video.

4.RESULT

The model's output will have a prediction confidence.

5. 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 ResNext CNN for frame level detection and RNN for video classification.together with LSTM. Based on the aforementioned criteria, the suggested approach may determine whether a video is a deep fake or real.parameters written down. We think it will deliver real-time data with extremely high accuracy.

6.. LIMITATIONS

In ths approach, fake audio detection is not possible as the approach is purely meant for video detection. In the future, audio and video detector can be implemented.

REFERENCES

[1] Yuezun Li, Siwei Lyu, "ExposingDF Videos By Detecting Face Warping Artifacts," in arXiv:1811.00656v3.

[2] Yuezun Li, Ming-Ching Chang and Siwei Lyu "Exposing AI Created Fake Videos by Detecting Eye Blinking" in arxiv.

[3] Huy H. Nguyen , Junichi Yamagishi, and Isao Echizen " Using capsule networks to detect forged images and videos ".

[4] Hyeongwoo Kim, Pablo Garrido, Ayush Tewari and Weipeng Xu "Deep Video Portraits" in arXiv:1901.02212v2.

[5] Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin "Detection of Synthetic Portrait Videos using Biological Signals" in arXiv:1901.02212v2.

[6] 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.

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

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

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

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

[11]Sequence Models And LSTM Networks https://pytorch.org/tutorials/beginner/nlp/sequence_models_tutorial.html

[12]https://discuss.pytorch.org/t/confused-about-the-imagepreprocessing-in-classification/3965

[13]https://www.kaggle.com/c/deepfake-detectionchallenge/data

[14]https://github.com/ondyari/FaceForensics

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

[16]P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-toimage translation with conditional adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5967–5976, July 2017. Honolulu, HI.

[17]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.