International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

DeepFake Detection Using Deep Learning

Kirti Jeswani, Shivang Negi, Rahul Patil
Guide: Mrs. Savita Raut
Electronics and Telecommunication,
KJ Somaiya College of Engineering
Mumbai, India

Abstract

The growing computation power has made the deep learning algorithms so powerful that creating an indistinguishable human synthesized video popularly called as deep fakes hasbecome very simple. Scenarios where these realistic face swapped deep fakes are used tocreate political distress, fake terrorism events and blackmail people are easily envisioned. In this work, we describe a new deep learning-based method that can effectively distinguish AI-generated fake videos from real videos. Our method is capable of automatically detecting the replacement and reenactment of deep fakes. We are trying to use Artificial Intelligence (AI) to fight Artificial Intelligence (AI). Our system uses a
Res-Next Convolution neural network to extract the framelevel features and these features and further used to train the Long Short Term Memory(LSTM) based Recurrent Neural Network(RNN) to classify whether the video is subject to any kind of manipulation or not, i.e. whether the video is deep fake or real video. To emulate the real time scenarios and make the model perform better on real time data, we evaluate our method on large amount of balanced and mixed data-set prepared by mixing the various available data-set like FaceForensic++, Deep fake detection challenge, and Celeb-DF. We also show how our system can achieve competitive results using a very simple and robust approach.

Keywords: Res-Next Convolution neural network. Recurrent Neural Network (RNN). Long Short Term Memory (LSTM). Computer vision

Introduction

The increasing sophistication of smartphone cameras and the availability of good internet connection all over the world has increased the ever-growing reach of social media and media sharing portals have made the creation and transmission of digital videos more easy than ever before. The growing computational power has made deep learning so powerful that would have been thought impossible only a handful of years ago. Like any transformative technology, this has created new challenges. So-called "DeepFake" produced by deep generative adversarial models that can manipulate video and audio clips. Spreading of the DF over the social media platforms have become very common leading to spamming and peculating wrong information over the platform. Thesetypes of the DF will be terrible, and lead to threating, misleading of common people.
To overcome such a situation, DF detection is very important. So, we describe a new deep learning-based method that can effectively distinguish AI-generated fake videos (DF Videos) from real videos. It's incredibly important to develop technology that canspot fakes, so that the DF can be identified and prevented from spreading over the internet.
For detection the DF it is very important to understand the way Generative Adversarial Network (GAN) creates the DF. GAN takes as input a video and an image of a specific
individual ('target'), and outputs another video with the target's faces replaced with those of another individual ('source'). The backbone of DF are deep adversarial neural networks trained on face images and target videos to automatically map the faces and facial expressions of the source to the target. With proper post- processing, the resulting videos can achieve a high level of realism. The GAN split the video into frames and replaces the input image in every frame. Further it reconstructs the video. This process is usually achieved by using autoencoders. We describe a new deep learning-based method that can effectively distinguish DF videos from the real ones. Our method is based same process that is used to create the DF by GAN. The method is based on a properties of the DF videos, due to limitation of computation resources and production time, the DF algorithm can only synthesize face images of a fixed size, and they must undergo an
affinal warping to match the configuration of the source's face. This warping leaves some distinguishable artifacts in the output deepfake video due to the resolution inconsistency between warped face area and surrounding context. Our method detects such artifacts by comparing the generated face areas and their surrounding regions by splitting the video into frames and extracting the features with a ResNext Convolutional Neural Network (CNN) and using the Recurrent Neural Network (RNN) with Long Short Term Memory(LSTM) capture the temporal inconsistencies between frames introduced by GAN during the reconstruction of the DF. To train the ResNext CNN model, we simplify the process by simulating the resolution inconsistency in affine face wrappings directly.

Literature review

Deepfake Video Detection Using Recurrent Neural Networks

D. Güera and E. J. Delp

This paper is about detecting fake videos using recurrent neural networks. It discusses the problem of creating fake videos that are difficult to distinguish from real ones. The authors propose a system that uses a convolutional neural network (CNN) to extract features from frames of a video, and then uses a recurrent neural network (RNN) to classify the video as fake or real. The system was evaluated on a large dataset of fake videos and was found to be effective in detecting them.

A Hybrid CNN-LSTM model for Video Deepfake Detection by LeveragingOptical Flow Features

P. Saikia, D. Dholaria, P. Yadav, V. Patel and M. Roy

This above paper is based on the use of Optical Flow vectors with a pre-trained CNN model, appended with LSTM layers to model the inconsistent motion of each pixel of the frames of videos, which can be evaluated to classify a video into fake or real. To reduce the computational constraints, the experiment was performed on a subset of frames as considering all the frames of the videos require higher computational power. However, from the experimentation it is observed that the model performed better with an increasing number of frames per video. Our work paves the way for many possible future works: firstly the model can be improved by training on huge set of the frames of the videos. Secondly, more datasets can be incorporated for better performance so that the model can be trained to detect videos of all kinds of deepfake manipulation techniques. Further the comparable score of our proposed model with the reduced number of frames indicate the possible realization of early detection of the fake content. Thus, the application of optical flow field seems to be promising in this domain and can be further investigated on explain ability of ultra-realistic deepfakes.

Deepfake Detection through Deep Learning

D. Pan, L. Sun, R. Wang, X. Zhang and R. O. Sinnott

The paper is about deepfake detection through deep learning. It discusses the application of neural networks and deep learning to automatically detect deepfake videos. The authors propose two deepfake detection technologies, Xception and MobileNet. They utilize training and evaluation datasets from FaceForensics++. The results show high accuracy, varying between 91-98%. The authors also developed a voting mechanism to further improve accuracy.

DeepFake Detection for Human Face Images and Videos: A Survey A.Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan

The above paper offers a comprehensive survey of a new and prominent technology, namely, DeepFake. It communicates the basics, benefits and threats associated with DeepFake, GAN-based DeepFake applications. In addition, DeepFake detection models are also discussed. The inability to transfer and generalize is common in most existing deep learning-based detection methods, which implies that multimedia forensics has not yet reached its zenith. Much interest has been shown by different important organizations and experts that are contributing to the improvement of applied techniques. However, much effort is still needed to ensure data integrity, hence the need for other protection methods. Furthermore, experts are anticipating a new wave of DeepFake propaganda in AI against AI encounters where neither side has an edge over the other.

Deep fake Detection using deep learning techniques: A Literature ReviewA.Mary and A. Edison

This paper is a comprehensive overview of deep learning-based deepfake detection techniques, including their strengths, weaknesses, and potential applications. The first section discussed the existing programs and technologies that are extensively used to make fake photos and videos. And in the second section discuss the different type of techniques that are used for deep fake images and videos. Also, provide details of available datasets and evaluation metrics that are used for deep fake detection. Despite the fact that deep learning has done well in detecting deep fakes, the quality of deep fakes has been increasing. In order to recognize fake videos & photos properly must be enhanced current deep learning approaches. Furthermore, given present deep learning approaches, it is unknown how to identify the number of layers necessary and the

appropriate architecture for deep fake detection. To improve their capacity to cope with the ubiquitous impacts of deep fakes and mitigate their consequences, social media companies are integrating deep fake detection tools.

Deepfake Detection: Current Challenges and Next Steps

S. Lyu

This paper discusses the current challenge and the next steps in deepfake detection. Firstly, one critical disadvantage of the current DeepFake generation methods is that theycannot produce good details such as skin and facial hairs. This is due to the loss of information in the encoding step of generation. However, this can be improved by incorporating GAN models which have demonstrated performance in recovering facial details in recent works. Secondly, the synthesized videos can be more realistic if they areaccompanied with realistic voices, which combine video and audio synthesis together in one tool. In the face of this, the overall running efficiency, detection accuracy, and more importantly, false positive rate, have to be improved for wide practical adoption. The detection methods also need to be more robust to real-life postprocessing steps, social media laundering, and counter-forensic technologies. There is a perpetual competition of technology, know-hows, and skills between the forgery makers and digital media forensic researchers. The future will reckon the predictions we make in this work.

Proposed System

There are many tools available for creating the DF, but for DF detection there is hardly any tool available. Our approach for detecting the DF will be great contribution in avoiding the percolation of the DF over the world wide web. We will be providing a web-based platform for the user to upload the video and classify it as fake or real. This project can be scaled up from developing a web-based platform to a browser plugin for automatic DF detections. Even big application like WhatsApp, Facebook can integrate this project with their application for easy pre detection of DF before sending to another user. One of the important objective is to evaluate its performance and acceptability in terms of security, user-friendliness, accuracy and reliability. Our method is focusing on detecting all types of DF like replacement DF, retrenchment DF and interpersonal DF. figure.1 represents the simple system architecture of the proposed system:

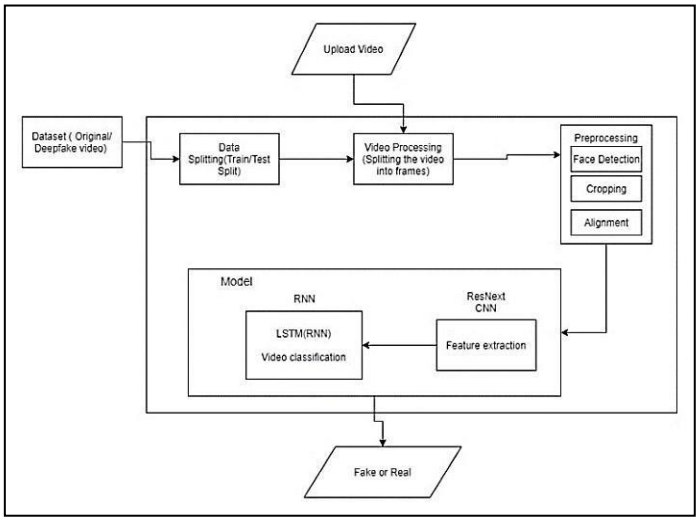

Fig. 1: System Architecture

Dataset

We are using a mixed dataset which consists of equal amount of videos from different dataset sources like YouTube, FaceForensics++, Deep fake detection challenge dataset. Ournewly prepared dataset contains 50% of the original video and 50% of the manipulated deepfake videos. The dataset is split into 70% train and 30% test set.

Preprocessing

Dataset preprocessing includes the splitting the video into frames. Followed by the face detection and cropping the frame with detected face. To maintain the uniformity in the number of frames the mean of the dataset video is calculated and the new processed face cropped dataset is created containing the frames equal to the mean. The frames that doesn't have faces in it are ignored during preprocessing.

As processing the 10 second video at 30 frames per second i.e total 300 frames will require a lot of computational

power. So for experimental purpose we are proposingto used only first 100 frames for training the model.

Model

The model consists of resnext50_32x4d followed by one LSTM layer. The Data Loader loads the preprocessed face cropped videos and split the videos into train and test set. Further the frames from the processed videos are passed to the model for training andtesting in mini batches.

ResNext CNN for Feature Extraction

Instead of writing the rewriting the classifier, we are proposing to use the ResNext CNN classifier for extracting the features and accurately detecting the frame level features. Following, we will be fine-tuning the network by adding extra required layers and selecting a proper learning rate to properly converge the gradient descent of the model. The 2048-dimensional feature vectors after the last pooling layers are then usedas the sequential LSTM input.

LSTM for Sequence Processing

Let us assume a sequence of ResNext CNN feature vectors of input frames as input and a 2-node neural network with the probabilities of the sequence being part of a deep fake video or an untampered video. The key challenge that we need to address is the de- sign of a model to recursively process a sequence in a meaningful manner. For this problem, we are proposing to the use of a 2048 LSTM unit with 0.4 chance of dropout, which is capable to do achieve our objective. LSTM is used to process the frames in a sequential manner so that the temporal analysis of the video can be made, by comparing the frame
at ‘t’ second with the frame of ‘t-n’ seconds. Where n can be any number of framesbefore t.

Predict

A new video is passed to the trained model for prediction. A new video is also preprocessed tobring 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 directlypassed 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 2.

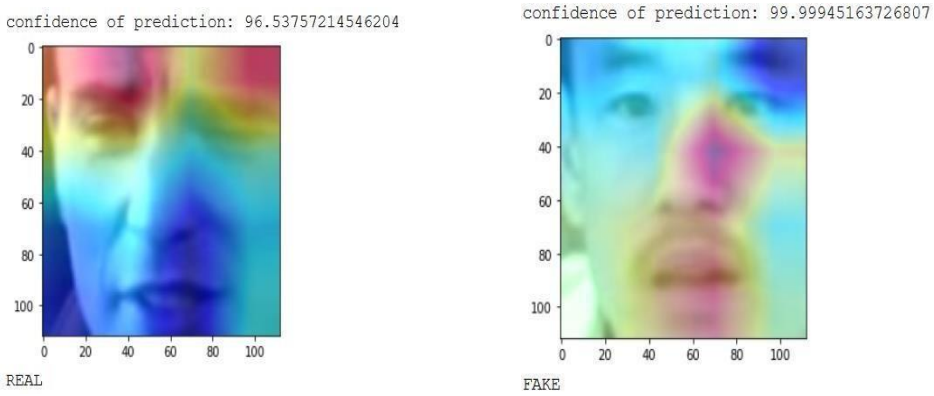

Fig 2: Obtained Results

Conclusion

We presented a neural network-based approach to classify the video as deep fake or real, along with the confidence of proposed model. The proposed method is inspired by the way the deep fakes are created by the GANs with the help of Autoencoders. Our method does the frame level detection using ResNext CNN and video classification using RNN along with LSTM. The proposed method is capable of detecting the video as a deep fake or real based on the listed parameters in paper. We believe that, it will provide a very high accuracy on real time data.

References

D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018, pp. 1-6, doi: 10.1109/AVSS.2018.8639163.

P. Saikia, D. Dholaria, P. Yadav, V. Patel and M. Roy, "A Hybrid CNN-LSTM model for Video Deepfake Detection by Leveraging Optical Flow Features," 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 2022, pp. 1-7, doi:0.1109/IJCNN55064.2022.9892905.

D. Pan, L. Sun, R. Wang, X. Zhang and R. O. Sinnott, "Deepfake Detection through Deep Learning," 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), Leicester, UK, 2020, pp. 134-143, doi: 10.1109/BDCAT50828.2020.00001.

A. Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan, "DeepFake Detection for Human Face Images and Videos: A Survey," in IEEE Access, vol. 10, pp. 18757-18775,2022, doi: 10.1109/ACCESS.2022.3151186.

A. Mary and A. Edison, "Deep fake Detection using deep learning techniques: A Literature Review," 2023 International Conference on Control, Communication and Computing (ICCC), Thiruvananthapuram, India, 2023, pp. 1-6, doi: 10.1109/ICCC57789.2023.10164881.

S. Lyu, "Deepfake Detection: Current Challenges and Next Steps," 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), London, UK,2020, pp. 1-6, doi: 10.1109/ICMEW46912.2020.9105991.