DEEPFAKE DETECTION USING LSTM AND RESNEXT

Dr.CH.V.Phani Krishna, HOD&Professor,Department of CSE,

Sowmya Arukala,BTech, Department of CSE, a.sowmyagoud456@gmail.com

Vishanth Reddy Battula, BTech, Department of CSE, reddyvishanth2@gmail.com

M.Bhavya Sri, BTech, Department of CSE, madhumuthyapothula@gmail.com

ABSTRACT: Deepfakes are synthetic media in which a person in an existing image or video is replaced with someone else's likeness. Generative Adversarial Networks, or GANs, are a deep-learning-based generative model. More generally, GANs are a model architecture for training a generative model, and it is most common to use deep learning models in this architecture. In the case of GANs, the generator model applies meaning to points in a chosen latent space, such that new points drawn from the latent space can be provided to the generator model as input and used to generate new and different output examples.. Thus we can easily use GANs to create deepfakes which can then be misused in a number of places. Deepfakes are concerning everyone out there in the digital world. The project deals with detection of deepfakes using Renext and LSTMs and packages the benefits of deep learning to detect deepfakes in the form of a Django web Application, To detect deepfakes we gather the frames from the video uploaded and split the video into desired number of frames. Following that we make use of python face recognition libraries and other C++ visual libraries to detect the face of the character from the video. We then apply our models ,which are trained for different number of frame sequences to predict if the video is a deepfake or Real.

1. 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. These types 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 can spot 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.

Fig.1: Example figure

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.

2. LITERATURE REVIEW

Exposing DeepFake Videos By Detecting Face Warping Artifacts

In this work, we describe a new deep learning based method that can effectively distinguish AI-generated fake videos (referred to as {\em DeepFake} videos hereafter) from real videos. Our method is based on the observations that current DeepFake algorithm can only generate images of limited resolutions, which need to be further warped to match the original faces in the source video. Such transforms leave distinctive artifacts in the resulting DeepFake videos, and we show that they can be effectively captured by convolutional neural networks (CNNs). Compared to previous methods which use a large amount of real and DeepFake generated images to train CNN classifier, our method does not need DeepFake generated images as negative training examples since we target the artifacts in affine face warping as the distinctive feature to distinguish real and fake images. The advantages of our method are two-fold: (1) Such artifacts can be simulated directly using simple image processing operations on a image to make it as negative example. Since training

a DeepFake model to generate negative examples is time-consuming and resource-demanding, our method saves a plenty of time and resources in training data collection; (2) Since such artifacts are general existed in DeepFake videos from different sources, our method is more robust compared to others. Our method is evaluated on two sets of DeepFake video datasets for its effectiveness in practice.

Exposing AI Created Fake Videos by Detecting Eye Blinking

The new developments in deep generative networks have significantly improve the quality and efficiency in generating realistically-looking fake face videos. In this work, we describe a new method to expose fake face videos generated with deep neural network models. Our method is based on detection of eye blinking in the videos, which is a physiological signal that is not well presented in the synthesized fake videos. Our method is evaluated over benchmarks of eye-blinking detection datasets and shows promising performance on detecting videos generated with DNN based software DeepFake.

Using capsule networks to detect forged images and videos

Recent advances in media generation techniques have made it easier for attackers to create forged images and videos. State-of-the-art methods enable the real-time creation of a forged version of a single video obtained from a social network. Although numerous methods have been developed for detecting forged images and videos, they are generally targeted at certain domains and quickly become obsolete as new kinds of attacks appear. The method introduced in this paper uses a capsule network to detect various kinds of spoofs, from replay attacks using printed images or recorded videos to computer-generated videos using deep convolutional neural networks. It extends the application of capsule networks beyond their original intention to the solving of inverse graphics problems.

Image-to-image translation with conditional adversarial networks

We investigate conditional adversarial networks as a general-purpose solution to image-to-image translation problems. These networks not only learn the mapping from input image to output image, but also learn a loss function to train this mapping. This makes it possible to apply the same generic approach to problems that traditionally would require very different loss formulations. We demonstrate that this approach is effective at synthesizing photos from label maps, reconstructing objects from edge maps, and colorizing images, among other tasks. Indeed, since the release of the pix2pix software associated with this paper, a large number of internet users (many of them artists) have posted their own experiments with our system, further demonstrating its wide applicability and ease of adoption without the need for parameter tweaking. As a community, we no longer hand-engineer our mapping functions, and this work suggests we can achieve reasonable results without hand-engineering our loss functions either.

DeepFakE: improving fake news detection using tensor decomposition-based deep neural network

Social media platforms have simplified the sharing of information, which includes news as well, as compared to traditional ways. The ease of access and

sharing the data with the revolution in mobile technology has led to the proliferation of fake news. Fake news has the potential to manipulate public opinions and hence, may harm society. Thus, it is necessary to examine the credibility and authenticity of the news articles being shared on social media. Nowadays, the problem of fake news has gained massive attention from research communities and needed an optimal solution with high efficiency and low efficacy. Existing detection methods are based on either news-content or social-context using user-based features as an individual. In this paper, the content of the news article and the existence of echo chambers (community of social media-based users sharing the same opinions) in the social network are taken into account for fake news detection. A tensor representing social context (correlation between user profiles on social media and news articles) is formed by combining the news, user and community information. The news content is fused with the tensor, and coupled matrix-tensor factorization is employed to get a representation of both news content and social context. The proposed method has been tested on a real-world dataset: BuzzFeed. The factors obtained after decomposition have been used as features for news classification. An ensemble machine learning classifier (XGBoost) and a deep neural network model (DeepFakE) are employed for the task of classification. Our proposed model (DeepFakE) outperforms with the existing fake news detection methods by applying deep learning on combined news content and social context-based features as an echo-chamber.

3. IMPLEMENTATION

Detection of Synthetic Portrait Videos using Biological Signals [5] approach extract biological signals from facial regions on authentic and fake portrait video pairs. Apply transformations to compute the spatial coherence and temporal consistency, capture the signal characteristics in feature sets and PPG maps, and train a probabilistic SVM and a CNN. Then, the aggregate authenticity probabilities to decide whether the video is fake or authentic. Fake Catcher detects fake content with high accuracy, independent of the generator, content, resolution, and quality of the video. Due to lack of discriminator leading to the loss in their findings to preserve biological signals, formulating a differentiable loss function that follows the proposed signal processing steps is not straight forward process.

DISADVANTAGES:

Due to lack of discriminator leading to the loss in their findings to preserve biological signals, formulating a differentiable loss function

Deepfakes are concerning everyone out there in the digital world. The project deals with detection of deepfakes using Renext and LSTMs and packages the benefits of deep learning to detect deepfakes in the form of a Django web Application, To detect deepfakes we gather the frames from the video uploaded and split the video into desired number of frames. Following that we make use of python face recognition libraries and other C++ visual libraries to detect the face of the character from the video. We then apply our models ,which are trained for different number of frame sequences to predict if the video is a deepfake or Real.

ADVANTAGES:

We presented a LSTM based approach to detect the video as deep fake or real, by processing 1 sec of video with good accuracy.

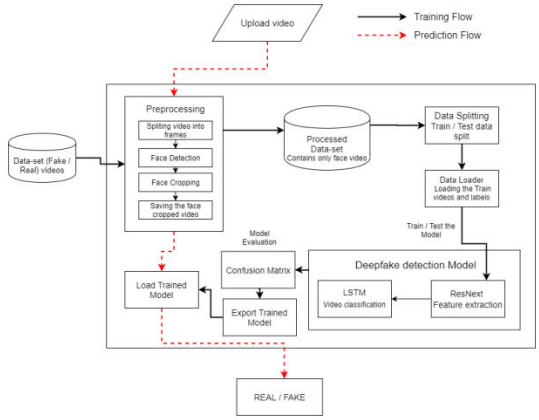

Fig.2: System architecture

MODULES:

Dataset:

We are using a mixed dataset which consists of equal amount of videos from different dataset sources like YouTube, FaceForensics++[14], Deep fake detection challenge dataset. Our newly 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 proposing to 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 and testing 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 used as 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 frames before t.

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.

4. METHODOLOGY

ALGORITHM USED:

Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised. Deep-learning architectures such as deep neural networks, deep belief networks, deep reinforcement learning, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.

CONVOLUTIONAL NEURAL NETWORK

To demonstrate how to build a convolutional neural network based image classifier, we shall build a 6 layer neural network that will identify and separate one image from other. This network that we shall build is a very small network that we can run on a CPU as well. Traditional neural networks that are very good at doing image classification have many more parameters and take a lot of time if trained on normal CPU. However, our objective is to show how to build a real-world convolutional neural network using TENSORFLOW.

Neural Networks are essentially mathematical models to solve an optimization problem. They are made of neurons, the basic computation unit of neural networks. A neuron takes an input (say x), do some computation on it (say: multiply it with a variable w and adds another variable b) to produce a value (say; z= wx+b). This value is passed to a non-linear function called activation function (f) to produce the final output(activation) of a neuron. There are many kinds of activation functions. One of the popular activation function is Sigmoid. The neuron which uses sigmoid function as an activation function will be called sigmoid neuron. Depending on the activation functions, neurons are named and there are many kinds of them like RELU, TanH.

If you stack neurons in a single line, it's called a layer; which is the next building block of neural networks. See below image with layers.

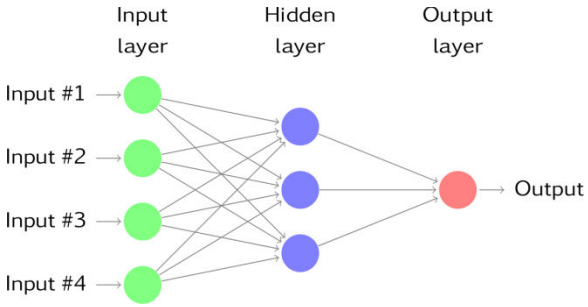

Fig.3: CNN model

To predict image class multiple layers operate on each other to get best match layer and this process continues till no more improvement left.

LSTM:

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture[1] used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points (such as images), but also entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected handwriting recognition,[2] speech recognition[3][4] and anomaly detection in network traffic or IDSs (intrusion detection systems). A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs. Relative insensitivity to gap length is an advantage of LSTM over RNNs, hidden Markov models and other sequence learning methods in numerous applications.

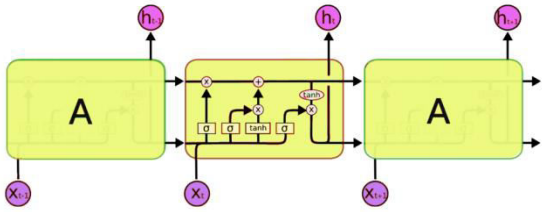

Fig.4: LSTM model

5. EXPERIMENTAL RESULTS

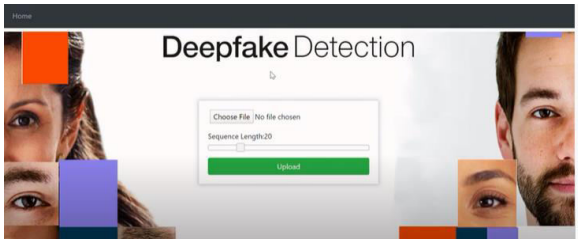

Fig.5: Output screen

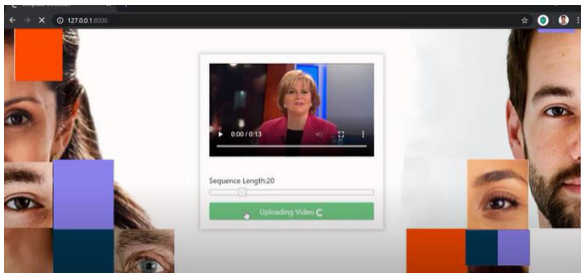

Fig.6: Output screen

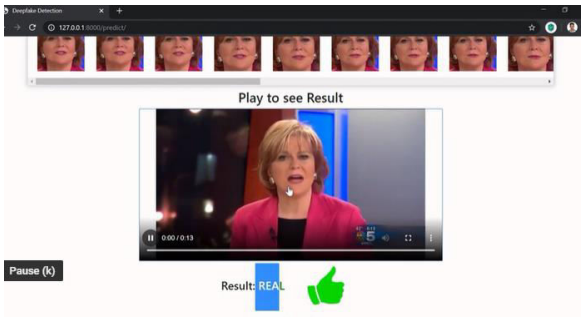

Fig.7: Output screen

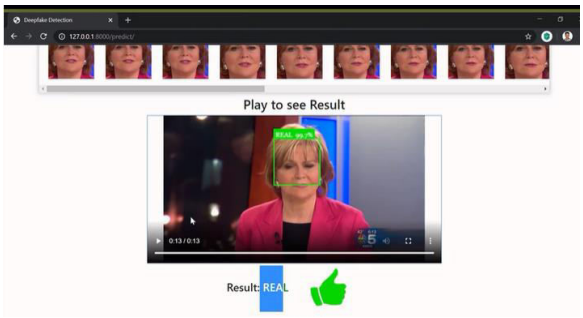

Fig.8: Output screen

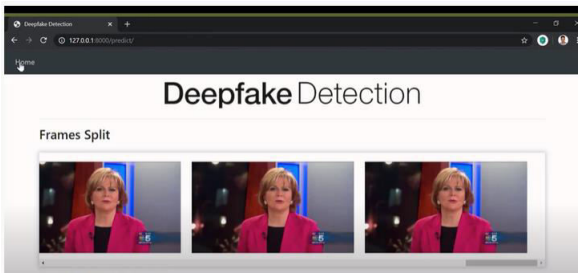

Fig.9: Output screen

6. 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.We presented a LSTM based approach to detect the video as deep fake or real, by processing 1 sec of video with good accuracy.

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-resnet-and-its-variants-5281e2f56035

[10] Long Short-Term Memory: From Zero to Hero with Pytorch:

https://blog.floydhub.com/long-short-term-memory-from-zero-to-hero-with-pytorch/