International Journal of Engineering Research and Science & Technology

IJERST

ISSN : 2319-5991



www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

DEEP FAKE DETECTION USING DEEP LEARNING

[1]Mr. Y. SHIVA RAO,[2]JELLA ESHWAR,[3]S PRASAD,[4]PENCHALA DIVYASRI,[5]K.ANAND

[1]Assistant Professor, Department of computer science & engineering Malla Reddy College of Engineering, secunderabad, Hyderabad.

[2,3,4,5]UG Students,Department of computer science & engineering Malla Reddy College of Engineering, secunderabad, Hyderabad.

ABSTRACT

Deep fakes are altered, high-quality, realistic videos/images that have lately gained popularity. Many incredible uses of this technology are being investigated. Malicious uses of fake videos, such as fake news, celebrity pornographic videos and financial scams are currently on the rise in the digital world. As a result, celebrities, politicians, and other well-known persons are particularly vulnerable to the Deep fake detection challenge. Numerous research has been undertaken in recent years to understand how deep fakes function and many deep learning-based algorithms to detect deep fake videos or pictures have been presented. This study comprehensively evaluates deep fake production and detection technologies based on several deep learning algorithms. In addition, the limits of current approaches and the availability of databases in society will be discussed. A deep fake detection system that is both precise and automatic. Given the ease with which deep fake videos/images may be generated and shared, the lack of an effective deep fake detection system creates a serious problem for the world. However, there have been various attempts to address this issue, and deep learning-related solutions outperform traditional approaches. These capabilities are used to train a ResNext which learns to categorizeif a video has been concern to manipulation or now no longer and is also capable of hit upon the temporal inconsistencies among frames presented by DF introduction tools.

Index Terms Deep Fakes, Deep Learning, Fake Generation, Fake Detection, Machine Learning.

I.INTRODUCTION

Motivation

The deep fake generation and detection technologies based on several deep learning algorithms are thoroughly assessed in this paper. Furthermore, the limitations of existing methodologies

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

and the accessibility of databases across society will be examined. An automated technique for deepfake detection that is accurate. The absence of an efficient deep fake detection system poses a major threat to the global community, given the simplicity with which deepfake movies and pictures may be created and distributed. There have been many efforts to solve this problem, however, and deep learning-related solutions work better than conventional methods.

Problem definition

Due to the huge loss of frame content during video compression, existing deep learning algorithms for image identification cannot effectively detect bogus videos. The severe deterioration of the frame data following video compression prevents the majority of image recognition techniques from being employed for videos. Additionally, videos provide a problem for techniques intended to identify only still fake images since their temporal features vary across sets of frames.

Objective of project

A framework on which low-level face manipulation defects are expected to further appear as temporal distortions with irregularities between the frames. However, deep learning algorithms frequently employ face photos from the internet that typically display people with wide eyes; fewer pictures of persons with closed eyes may be seen online. As a result, deep fake algorithms are unable to generate fake faces that blink often in the absence of photographs of actual people doing so. Deep fakes, in other words, have far lower blink rates than regular videos.

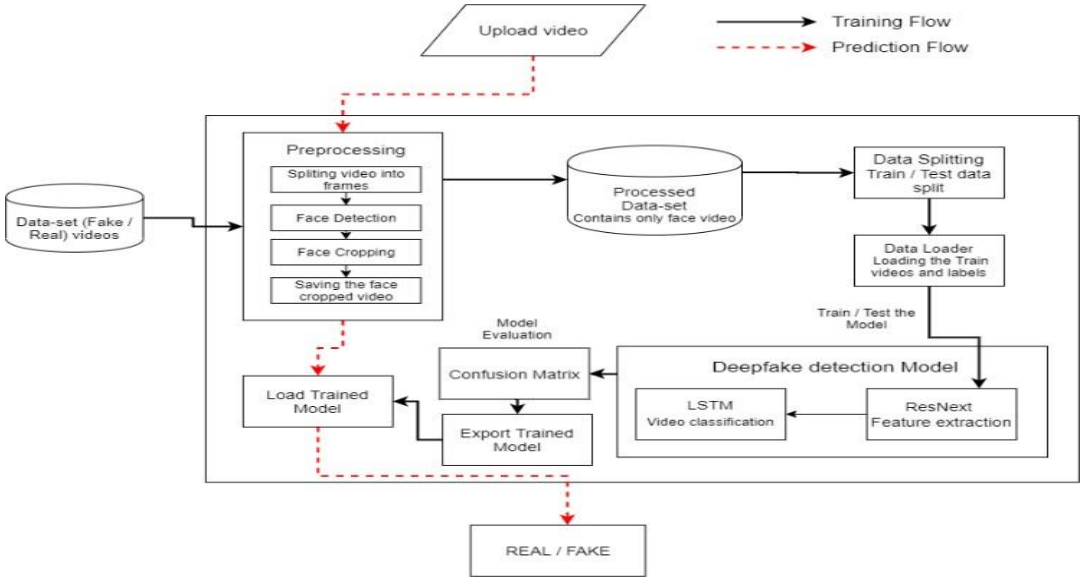
Scope of project

Detecting deep fake images and videos using deep learning techniques is an important and evolving area of research and development. The scope of this field is broad, encompassing both technological advancements and the societal implications of deep fake technology. Here are some key aspects to consider within the scope of deep fake detection using deep learning techniques:

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024



II. EXISTING SYSTEM

Zhao et al. recently introduced a methodology for deep fake detection utilizing the self-consistency of local source features, which are spatially-local, content-independent details of pictures. A CNN model employs a unique representation learning approach to extract these source features, which are represented as down-sampled feature maps referred to as pairwise self-consistency learning. This aims to punish feature vector pairings that correspond to areas in the same picture with poor cosine similarity scores. When dealing with false pictures created by technologies that output the entire image directly and whose source features are constant throughout each point inside each image, it could have a disadvantage.

In past months, free deep learning-based software tools have made the creation of credible face exchanges in videos that leave few traces of manipulation, in what are known as "DeepFake"(DF) videos.

Manipulations of digital videos has been demonstrated for many years through the good use of visual effects, recent advances in deep learning have led to a drastic increase in the making real looking of fake content and the accessibility in which it can be created.

Disadvantages of existing system:

- Since, fake image-based methods use error functions for real or fake image detection. For video, it needs lots of computational power and

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

is hence time-consuming by using such methods.

- Some poorly created deep fake videos keep some visual artifacts behind, which can be used for deepfake detection. Thus we can group methods used for classification based on classifiers used i.e either deep or shallow.

III. PROPOSED SYSTEM

There are many tools available for creating the DeepFakes, but for DeepFakes detection there is hardly any tool available. Our approach for detecting the DF will be a 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 for uploading the video and detect if its fake or real. This project is often scaled up from developing a webbased platform to a browser plugin for automatic DF detections. Even big applications like WhatsApp, Facebook can integrate this project with their application for easy pre-detection of DF before sending it to another user. One of the important objectives 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.

Advantages of proposed system

- Deep learning has shown considerable achievement in the identification of deep fakes.
- In order to recognize fake videos & photos properly must be enhanced current deep learning approaches.
- It primarily covers classic detection methods as well as deep Learning based methods such as CNN, RNN, and LSTM.

IV. MODULES:

Dataset: To built any machine learning and deep learning model we require a real-world data. First we collected data from different platform like Kaggle's Deepfake Detection challenge, Celeb-DF[8], FaceForensic. Kaggle's DeepFake detection challenge contains 3000 videos in which 50% data is real and 50% is manipulated data. Celeb-DF contains the videos of some famous

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

celebrities and there are a total of 1000 videos in which 500 are real and 500 are manipulated videos.FaceForensic ++ dataset contains a total of 2000 videos of which 1000 are real and the remaining are manipulated. Further this all three datasets are merged together and passed to the preprocessing of data.

Data Preprocessing: Preprocessing of data is a very important part as by doing preprocessing we actually try to get some important information from the data. We eliminate unnecessary data from original data. Splitting the movie into frames is part of the dataset preprocessing. Face detection is then performed, and the frame with the detected face is cropped. To preserve consistency in the number of frames, the mean of the video dataset is determined, and a new processed face cropped dataset containing the frames equal to the mean is constructed. During preprocessing, frames that do not include faces are ignored. Processing a 10-second movie at 30 frames per second, or 300 frames in total, will necessitate a significant amount of CPU power. So, for the sake of experimentation, we propose using only the first 100 frames to train the model.

Model: The model is made up of resnext50 32x4d and one LSTM layer. The Data Loader loads the preprocessed face cropped films and divides them into two groups: train and test. In addition, the frames from the processed videos are supplied to the model in tiny batches for training and testing.

ResNextCNN for Feature Extraction: We propose using the ResNext CNN classifier for extracting features and reliably recognizing frame-level characteristics instead of rewriting the classifier. Following that, we'll fine-tune the network by adding extra layers as needed and setting a correct learning rate to ensure that the gradient descent of the model is properly converged. LSTM for Sequence Processing: Assume a 2-node neural network with the probabilities of the sequence being part of a deep fake video or an untampered video as input and a sequence of ResNext CNN feature vectors of input frames as output. The main problem that we must solve is the design of a model that can recursively process a sequence in a meaningful way. For this task, we propose using a 2048 LSTM unit with a 0.4 likelihood of dropping out, which is capable of achieving our goal. The LSTM is used

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

to analyze the frames sequentially in order to do a temporal analysis of the video by comparing the frame at ‘t' second with the frame at ‘t' second.

Predict: The trained model is given a new video to forecast. A fresh video is also preprocessed to incorporate the trained model's format. The video is divided into frames, then face cropped, and instead of keeping the video locally, the cropped frames are sent immediately to the trained model for identification.

V. ALGORITHMS:

Long short-term memory (**LSTM**):

Long short-term memory is an artificial recurrent neural network (RNN) architecture 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, 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 trainingtraditional RNNs. Relative insensitivity to gap length is an advantage of LSTM over RNNs, hidden Markov models and other sequence learning methods in numerous applications

Training:

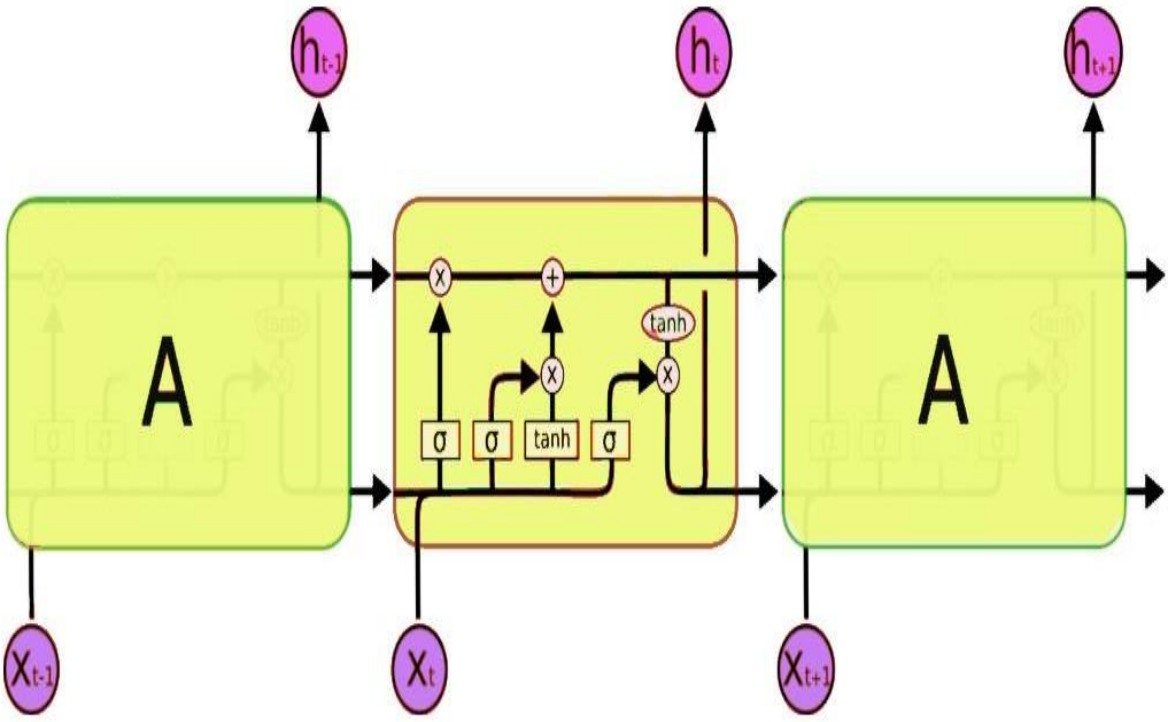
An RNN using LSTM units can be trained in a supervised fashion, on a set of training sequences, using an optimization algorithm, like gradient descent, combined with backpropagation through time to compute the gradients needed during the optimization process, in order to change each weight of the LSTM network in proportion to the derivative of the error (at the output layer of the LSTM network) with respect to corresponding weight.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

A problem with using gradient descent for standard RNNs is that error gradients vanish exponentially quickly with the size of the time lag between important events. However, with LSTM units, when error values are back-propagated from the output layer, the error remains in the LSTM unit's cell. This "error carousel" continuously feeds error back to each of the LSTM unit's gates, until they learn to cut off the value.



ResNeXt:

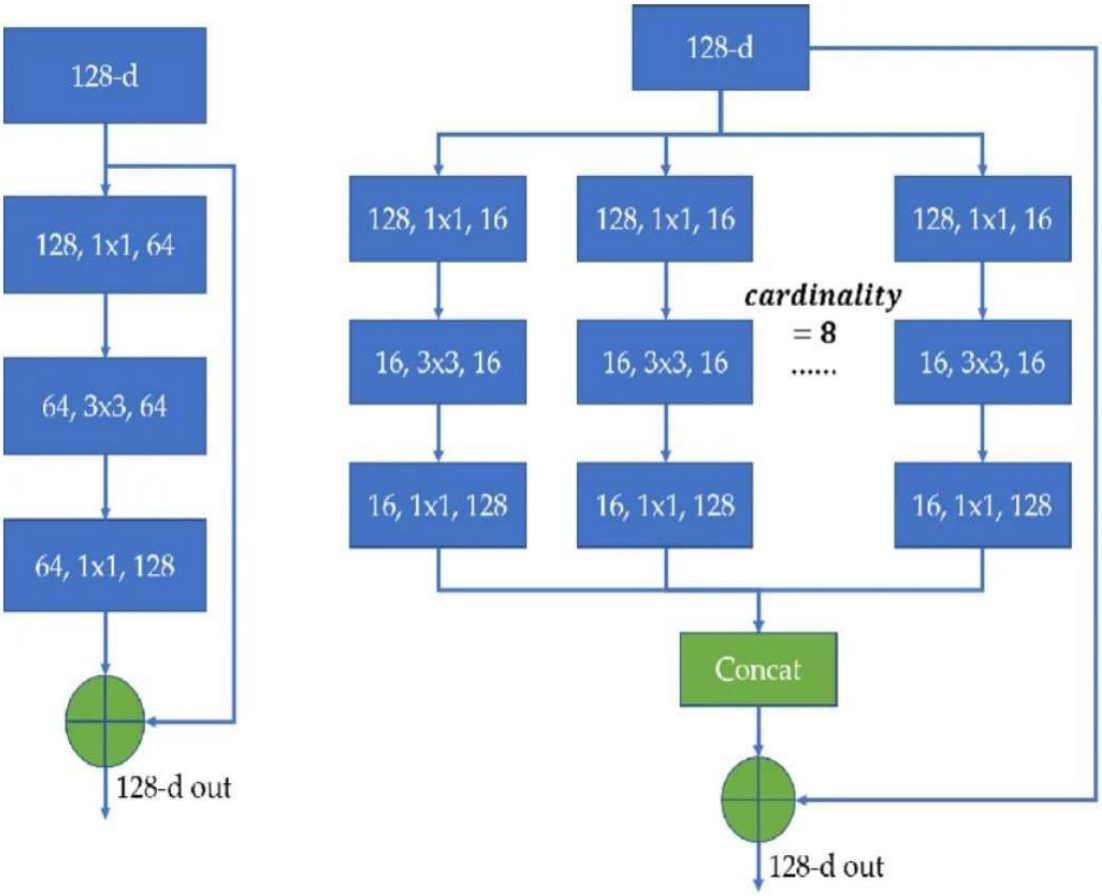
ResNeXt is a Convolutional Neural Network (CNN) architecture, which is a deep learning model. ResNeXt was developed by Microsoft Research and introduced in 2017 in a paper titled "Aggregated Residual Transformations for Deep Neural Networks."

ResNeXt uses the basic ideas of the ResNet (Residual Network) model, but unlike ResNet, it uses "groups" instead of many smaller paths. These groups contain multiple parallel paths, and each path is used to learn different features. This allows the network to learn more features more effectively, increasing its representational power.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024



The main features and advantages of ResNeXt are:

Parallel Paths: ResNeXt is based on the use of multiple parallel paths (or groups) in the same layer. This allows the network to learn a broader and more diverse set of features.

Depth and Width: ResNeXt combines two basic methods, both increasing the depth of the network and increasing the width of the network by increasing the number of groups in each layer. This allows using more parameters to achieve better performance.

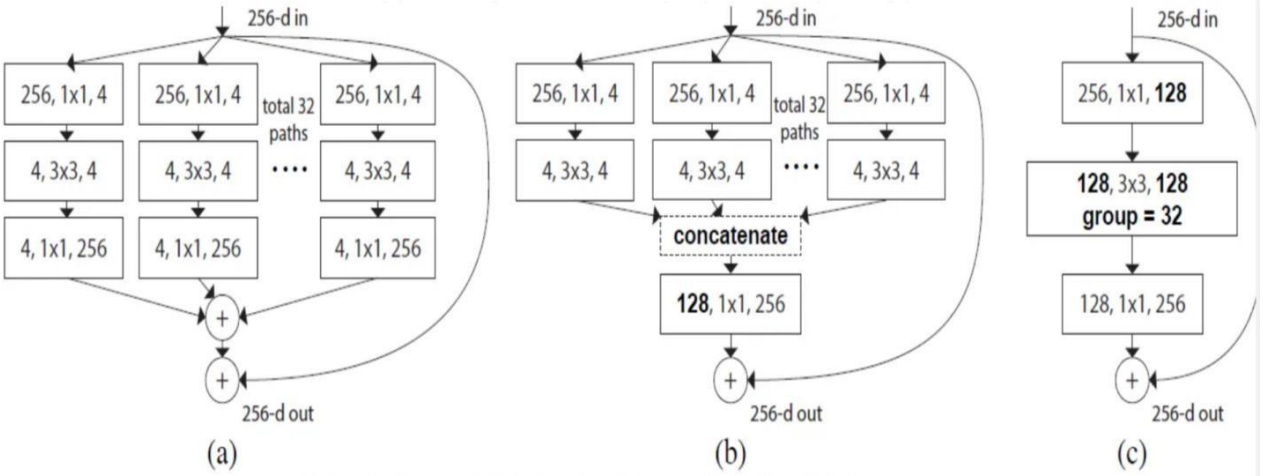
State-of-the-Art Performance: ResNeXt has demonstrated state-of-the-art performance on a variety of tasks. It has achieved successful results especially in image classification, object recognition and other visual processing tasks.

Transfer Learning: ResNeXt can be effectively used to adapt pre-trained models to other tasks. This is important for transfer learning applications.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024



ResNeXt is used in many application areas, particularly in deep learning problems working with visual and text data, such as image classification, object detection, face recognition, natural language processing (NLP) and medical image analysis. This model performs particularly well on large data sets and is also a suitable option for transfer learning applications.

VI. CONCLUSION:

Various researchers have created a number of deep-learning approaches for deep fake images and videos. Due to the extensive availability of photographs and videos in social media material, deep fakes had grown in popularity. This is especially crucial in social networking sites that make it simple for users to spread and share such fake information. Numerous deep learning-based approaches have recently been put out to deal with this problem and effectively identify fake images and videos. 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.

We provided a neural network-primarily based totally method to classify the

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 2, 2024

video as deep fake or actual, at the side of the self-assurance of the proposed model. Our approach does the frame stage detection the use of ResNext CNN and video class the use of LSTM. The proposed approach is successful in detecting the video as a deep fake or actual primarily based totally on the listed parameters in the paper. We consider that it'll offer a very excessive accuracy on actual time data.

VII. REFERENCES

[1] M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784, 2014.

[2] Y. Bengio, P. Simard, and P. Frasconi, "Long short-term memory," IEEE Trans. Neural Netw, vol. 5, pp. 157–166, 1994.

[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016. [4] S. Hochreiter, "Ja1 4 rgen schmidhuber (1997)."long short-term memory"," Neural Computation, vol. 9, no. 8.

[5] M. Schuster and K. Paliwal, "Networks bidirectional reccurent neural," IEEE Trans Signal Proces, vol. 45, pp. 2673–2681, 1997.

[6] J. Hopfield et al., "Rigorous bounds on the storage capacity of the dilute hopfield model," Proceedings of the National Academy of Sciences, vol. 79, pp. 2554–2558, 1982.

[7] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al., "Google's neural machine translation system: Bridging the gap between human and machine translation," arXiv preprint arXiv:1609.08144, 2016.