e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

DEEPFAKE DETECTION USING XCEPTION AND LSTM

Adil Mohammed Parayil[*1], Ameen Masood V[*2], Muhammed Ajas P[*3], Tharun R[*4], Usha K[*5]

[*1,2,3,4]APJ Abdul Kalam Technological University, Computer Science And Engineering, NSS College Of Engineering, Palakkad, Kerala, India.

[*5]Department Of Computer Science And Engineering, NSS College Of Engineering, Palakkad, Kerala, India.

DOI : https://www.doi.org/10.56726/IRJMETS41123

ABSTRACT

Deepfakes are fabricated works of art where a person appearing in an earlier photograph or motion picture is modified to exhibit the face of another person. A deep learning-based generative model called a Generative Adversarial Network (GAN) is a model architecture for training generative models to create fake videos. In the context of GANs, the generation model lends significance to points in a predetermined latent space, enabling fresh points pulled from the latent space to be fed to the generator model as input and utilized to produce brand-new and distinctive output instances. Due to this, it is simple to use GANs to build deep fakes that can be used inappropriately in various contexts. Due to the engrossing misuse of deepfakes, there is a need for appropriate detection tools. Deepfake detection requires a large amount of data to train models and test them. They require large datasets that contain thousands of videos, both real and fake at equal ratios to avoid biased results. This paper works with deep learning-based deep fake detection using Convolutional Neural Network(CNN) and Recurrent Neural Network(RNN), CNN utilizing the Xception network, and ytLSTM as RNN. In the algorithm, the spatial features are recognized by the Xception and LSTM identifies the temporal inconsistency between frames. The models are trained against three standard datasets. The results are also validated with standard datasets resulting in deepfake prediction with a minimal computational time and nominal accuracy.

Keywords: Deepfake Detection, Convolutional Neural Network, Recurrent Neural Network, Xception, LSTM.

I. INTRODUCTION

A rapidly developing technology entitled Deepfake uses cutting-edge machine learning and artificial intelligence algorithms to produce incredibly lifelike and frequently undetected fake photos, videos, and audio recordings of people. Deep learning algorithms are used to construct deepfakes, which are difficult to distinguish aside from the genuine entity because they analyze and recreate patterns in enormous volumes of data, including voice patterns, body language, and facial expressions. The abuse of deepfake technology has caused serious concerns, despite the fact that it has several legal uses, such as in the entertainment sector and for training AI algorithms. Deepfakes may be used to produce false information, including propaganda and fake news, that can spread quickly over social media and other internet platforms, endangering public confidence and societal norms. Therefore, it is essential to create efficient instruments and methods to identify and stop deepfakes as well as to spread knowledge about their possible abuse. To achieve this, deepfake detection systems will need to continue to be researched and developed, as well as legal and regulatory actions to address the improper use of this technology.

For detecting the deepfake it is very important to understand the way Generative Adversarial Network (GAN) creates the deepfake. 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 is 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.

Various detection models have been arising featuring the detection of deepfakes caused by eye-blinking irregularities, evaluated on eye-blinking detection datasets, and showing promising results for videos generated using Deep Neural Network based software.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

However other features like teeth, wrinkles, artificial accessories, etc should also be considered. The proposed method considers all such artifacts also for detection.

Other methods include deepfake face recognition using optical flow vectors. This approach involves estimating the optical flow between pairs of consecutive frames and using the resulting flow maps as input to a CNN for classification. Although this approach gave a promising result, the approach only focused on the spatial aspects and not the temporal properties.

A robust approach was proposed based on ResNext and LSTM for deepfake video detection. This is a promising method using both spatial and temporal artifacts from videos. The model was trained for various standard datasets and a combination of them with an equal number of real and fake videos, thus eliminating the bias. This model resulted in high-accuracy prediction standard deepfake detection datasets. Since ResNext is a huge network and focuses on improving accuracy, the model resulted in a higher computation time. Our model provides a solution by replacing the CNN model with Xception CNN which is equivalently accurate and has a lesser computational time. Thus incorporating both spatial and temporal features with a better computational time without compromising accuracy.

Due to limitations in computing resources, the detection algorithm proposed could only detect images of faces with a fixed size, the model uses wrapping for the frames to match the configuration of the source face. Undergoing wrapping produce inconsistency in resolution between the wrapped area and its surrounding, which will leave some distinguishable artifacts. The proposed model compares these artifacts from the given input by extracting the videos into frames. Each frame is passed through the Xception network using a time-distributed layer and using the feature vectors thus produced, LSTM identifies the temporal inconsistencies.

II. METHODOLOGY

Although there are several tools accessible for DF creation, there are very few ones available for DF detection. Our method of DF detection will significantly contribute to preventing the spread of DF over the internet. We'll offer a web-based platform where users may post videos and mark them as fake or authentic.

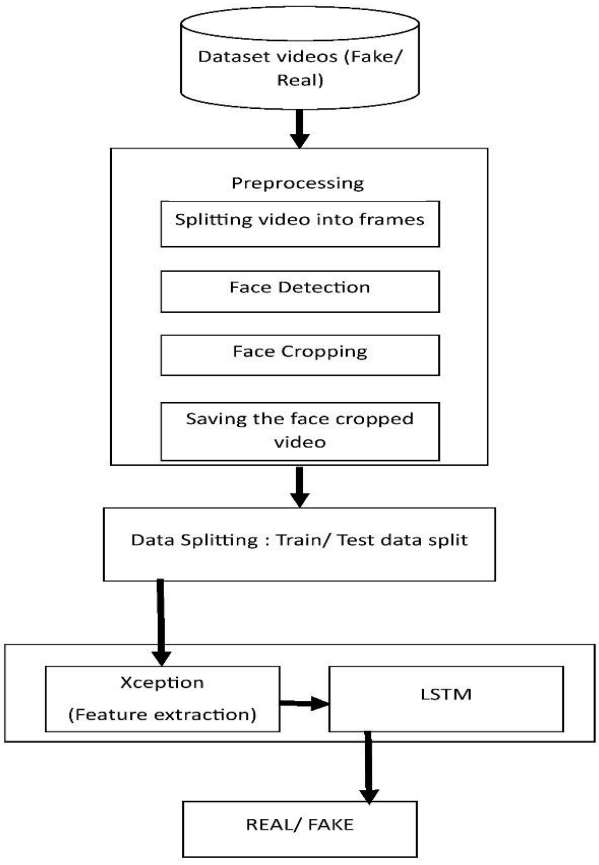

Figure 1. System Architecture

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

A. Dataset

The information was acquired using a variety of data sources, including FaceForensic++ (FF), and Deepfake detection challenge (DFDC). Then, to do precise and immediate detection of various types of movies, we combined the datasets we had collected. We have taken into account 50% real and 50% false videos to prevent the model's training bias. We have taken 160 genuine and 160 deceived movies from DFDC and FaceForensic++(FF) data sources. This results in a total of 320 movies in our dataset. 70% of the videos in the training dataset (224) and 30% of the test dataset (96) make up the training dataset. 50 percent actual videos and 50 percent fake videos make up each split in the train and test split, which is balanced.

B. Preprocessing

The video has been split into frames as part of the dataset preparation. Followed by face detection and cropping the frame with the detected face. The mean of the dataset video is determined in order to preserve 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 include any faces. As processing the 300 frames of the 10-second movie at 30 frames per second will take a lot of computer power. So, for experimental reasons, we suggest training the model using only the first 100 frames.

Figure 2: Frame Extracted from a Fake Video

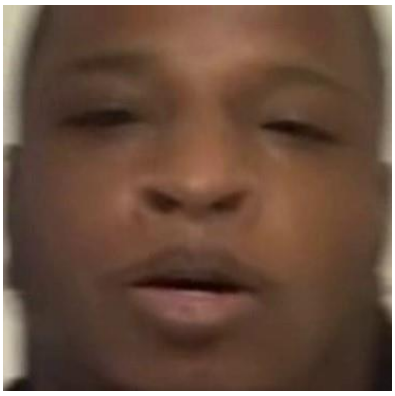

Figure 3: Frame after Face Cropping

C. Model

The model includes an Xception followed by an LSTM layer. The preprocessed face-cropped films are loaded by the data loader. Further divides them into a train set and a test set. Then it is passed to the model.

D. Xception CNN for Feature Extraction

Xception uses depthwise separable convolutions instead of standard convolutions, which significantly reduces the number of parameters and computations while maintaining high accuracy. This design has made Xception a popular choice for applications with limited computational resources. We are proposing to use the Xception

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

CNN classifier for extracting the features and accurately detecting the frame level features with least computationl effort. Following, we will be finetuning the network by adding extra required layers. The 2048-dimensional feature vectors after the last pooling layers are then used as the sequential LSTM input.

E. LSTM for Sequence Processing

A 2-node neural network using a sequence of Xception CNN feature vectors of input frames as input and the probability of the sequence is part of a deep fake video or a real video. Designing a model that meaningfully processes a series recursively is the main issue that has to be addressed. We suggest using a 2048 LSTM unit with a 0.5 risk of dropout for this reason which could achieve our objective. By comparing the frame at second 't' with the frame at second 't-n', LSTM is used to sequentially process the frames in order to do a temporal analysis of the video. Where n can be any number of frames before t.

F. 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 in local storage, the cropped frames are directly passed to the trained model for detection.

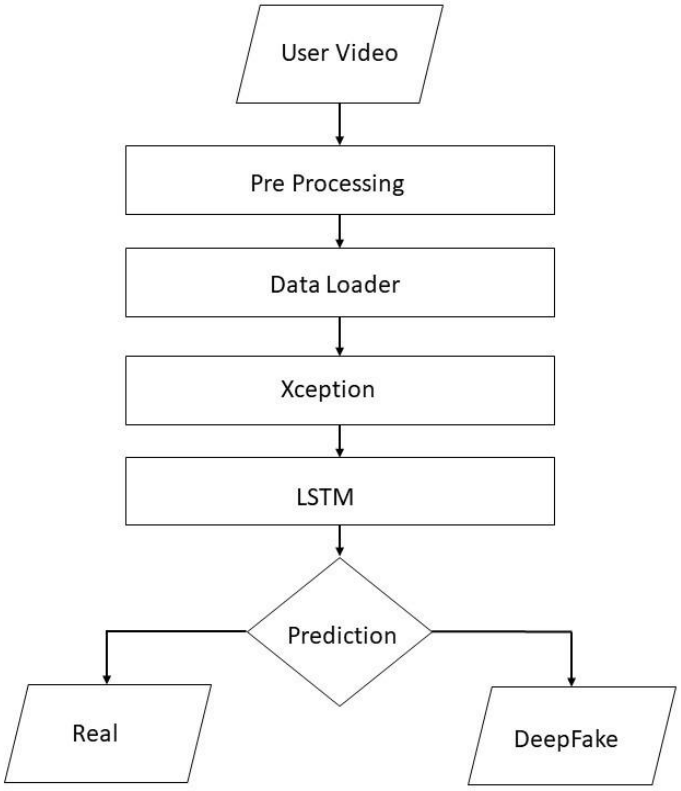

Figure 4: Prediction Flow

III. RESULTS AND DISCUSSION

The output of the model is whether the uploaded video is deepfake or a real video. After all the pre-processing steps, the face-cropped frames are given input into Xception and LSTM neural networks. These cropped frames are used for evaluating whether a video is fake or real. Training test evaluation is shown in Fig 5

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

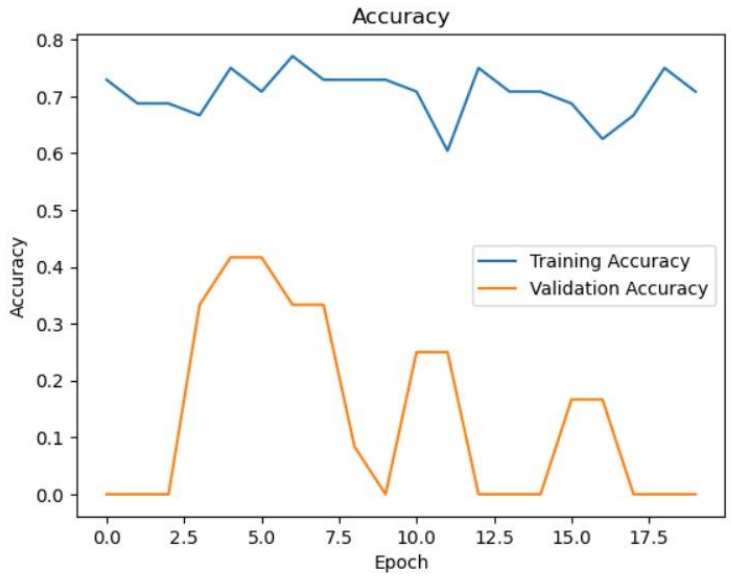

Figure 5: Training Validation Accuracy.

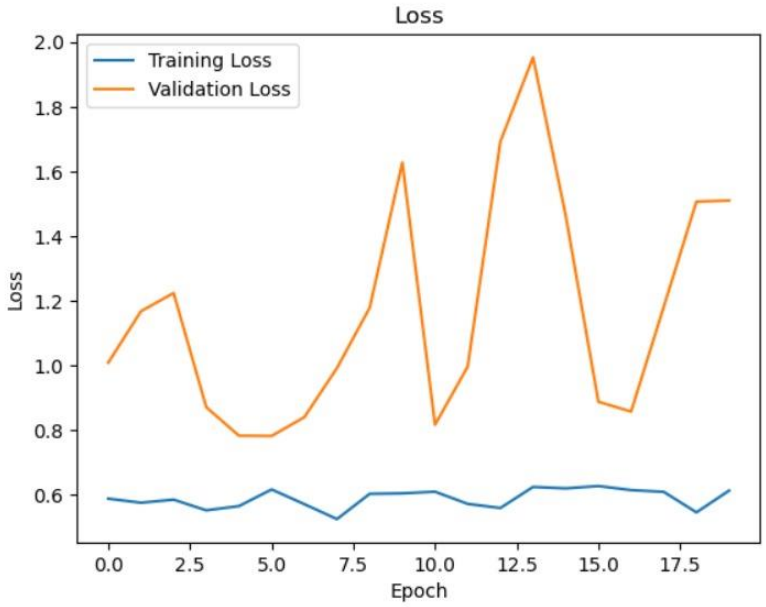

Figure 6: Training Validation Loss.

IV. CONCLUSION

We presented a neural network-based approach to classify the video as deep fake or real. Our method does the frame level detection using Xception CNN and video classification using RNN along with LSTM. The proposed method is capable of detecting the video as a deep fake or real. We believe that it will provide an average accuracy on real-time data with less computational effort. Overall, the combination of Xception CNN and LSTM RNN is a promising approach for deepfake detection that can provide high accuracy and robustness. However, further research is needed to explore the performance of this approach on more diverse datasets and in different real-world scenarios.

V. FUTURE SCOPE

The field of deepfake detection is an area of active research, and there are several avenues for future exploration and development. Firstly, while our study showed promising results using Xception CNN and LSTM RNN, there is still room for improvement in the accuracy and efficiency of deepfake detection models.

Secondly, the availability of larger and more diverse datasets can further improve the accuracy of deepfake detection models. The development of more advanced generative models for creating fake videos will also require the creation of new datasets for training and testing the models.

In conclusion, the future of deepfake detection research is promising, and there are several areas of exploration and development. The use of more advanced deep learning architectures, larger and more diverse datasets, integration with video hosting platforms, and explainable models are some of the avenues for future research.

VI. REFERENCES

[1] Roberto Caldelli et. al., 2021, "Optical Flow based CNN for detection of unlearnt deepfake manipulations", Pattern Recognition Letters vol. 146, pp. 31-37

[2] X. Wu, R. Liu, H. Yang and Z. Chen, 2020 "An Xception Based Convolutional Neural Network for Scene ImageClassification with Transfer Learning", 2nd International Conference on Information Technology and Computer Application (ITCA), pp. 26

[3] Sowmya Arukala, Vishanth Reddy Battula & M.Bhavya Sri, 2022, "Deepfake detection using LSTM and ResNext", Journal of Engineering Sciences, vol 13

[4] Y. Li, M. Chang, and S. Lyu, 2018, In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking, IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1-7

[5] Ruben Tolosana et. al., 2022, "DeepFakes detection across generations: Analysis of facial regions, fusion, and performance evaluation", Engineering Applications of Artificial Intelligence vol.110.

[6] David Guera and Edward J, 2018, "Deepfake video detection using recurrent neural networks", 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)

[7] Alankrita Aggarwal et. al., 2021, "Generative adversarial network: An overview of theory and applications", International Journal of Information Management Data Insights vol. 1

[8] B. Dolhansky et. al., 2020, The DeepFake detection challenge (DFDC) dataset.

[9] Xu et al., 2021, "DeepFake Video Detection Based on Xception and Spatial-Temporal Attention", IEEE Transactions on Circuits and Systems for Video Technology, 31(11), 4607-4619

[10] M. Hossain et al., 2019, "Deepfake Video Detection using Xception and Local Binary Patterns", 10th International Conference on Intelligent Systems, Modelling, and Simulation.