Deepfake Detection Using Deep Learning

Sushant Dawane[1], Kamla Vishwakarma[2], Prince Kori[3], Yogita Chavan[4]
[1,2,3,4]*Computer Engineering, New Horizon Institute of technology and management, Thane, India*

Abstract **- Increasing computing power has made deep learning algorithms so powerful that creating a fake video generated by artificial intelligence, popularly called as deep fakes, is very simple. Scenarios where this realistic face has been replaced by deep fakes are used to create political unrest, fake terrorist acts, revenge porn, blackmailing nations can easily be imagined. In this work, a new method is based on deep learning that can effectively distinguish fake videos generated by artificial intelligence from real videos. This method is able to automatically detect replacement and reenactment of deep forgery. Using artificial intelligence (AI) to fight against artificial intelligence (AI). This system uses Res-Next Convolution Neural Network to extract frame-level features and these features and further uses Long-Short-Term Memory (LSTM)-based Recurrent Neural Network (RNN) to classify whether the video is subject to some kind of manipulation or not, i.e.. whether the video is deepfake or real video. System can also achieve competitive results using a very simple and robust approach.**

Keywords – Res-next convolution neural network, RNN, LSTM (Long short-term memory)

I. INTRODUCTION

Deep fake is a technique for human image synthesis based on neural network tools like GAN(Generative Adversarial Network) or Auto Encoders etc. These tools super impose target images onto source videos using a deep learning techniques and create a realistic looking deep fake video. These deep-fake videos are so real that it becomes impossible to spot difference by the naked eye. In this work, to describe a new deep learning-based method that can effectively distinguish AI-generated fake videos from real videos. Using the limitation of the deep fake creation tools as a powerful way to distinguish between the pristine and deep fake videos. During the creation of the deep fake the current deep fake creation tools leaves some distinguishable artifacts in the frames which may not be visible to the human being but the trained neural networks can spot the changes. Deepfake creation tools leave distinctive artifacts in the resulting Deep Fake videos, and can be effectively captured by Res-Next Convolution Neural Networks. The system uses a Res-Next Convolution Neural Networks to extract frame-level features. These features are then used to train a 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. The proposed to evaluate this method against a large set of deep fake videos collected from multiple video websites. This model performs better on real time data. To achieve this model is trained on combination of available data-sets.

II. LITERATURE SURVEY

A. Deepfake Video Detection using Image Processing and Hashing Tools

This paper was published in 2020 by International Research Journal of Engineering and Technology (IRJET), In this paper, System which works over making deepfake video detection an easier and much simpler task by using the image processing and hashing techniques unlike using the RNNs for this purpose. In this paper, a proposed solution is given where images are not compared to entire images ' data of each frame in the video.

B. DeepFake Detection for Human Face Images and Videos

This paper was published in 2020 by Journal of IEEE Access, in this paper, to identify and classify Deepfake, research in Deepfake detection using deep neural networks (DNNs). Generally, Deepfake models are trained on Deepfake datasets and tested with experiments.

C. DeepFake Detection

This paper was published in 2021 by IRE (Iconic Research and Engineering) Journal, in this paper, to identify and classify Deepfakes, and developed a

detection model using convolution neural network (CNN), for face detection and Recurrent neural network (RNN) for video classification and develop the detection techniques by training the neural network in an effective and optimizing way.

D. Deepfake Videos Detection Based on Texture Features

This paper was published in 2021 by Journal of CMC (Computer, Material & Continua), In this paper, a new method is proposed to detect Deepfake videos. Firstly, the texture features are constructed, which are based on the gradient domain, standard deviation, gray level co-occurrence matrix and wavelet transform of the face region. Then, the features are processed by the feature selection method to form a discriminant feature vector, which is finally employed to SVM for classification at the frame level.

III. PROBLEM STATEMENTS, OBJECTIVES AND SCOPE

A. Problem Statement

To design and develop an deep fake detection system to classify the video as fake or real using Deep Learning

B. Objecctives

- To create a deepfake detection model to target AI generated videos.
- To distinguish and classify whether the video is fake or real.

C. Scope

To detect give video is deepfake or not on the basis of some parameters such as:

- ❖ Teeth Enhancement
- ❖ Face Angle
- ❖ Skin tone
- ❖ Facial Expression
- ❖ Lighting
- ❖ Different Pose
- ❖ Hairstyle

- Path of the video is to be uploaded in the trained model on Google Colab for video detection.
- Users of the application will be able detect whether the uploaded video is fake or real with accuracy.
- The User will be able to see the playing video with the output accuracy shown with a square frame on the face.
- The goal of this project is to detect fake videos using trained models.
- The dataset will be divided into training data and test data.
- The deepfake dataset will be a combination of Kaggle, face forensics.

IV. PROPOSED SYSTEM

A. Feasibility Analysis:

Feasibility analysis addresses things like where and how the system will work. It works deep into the details of a system to see if and how it can succeed and serves as a valuable tool for developing a winning business plan. The feasibility study is further divided into four categories as described below.

1)Economic Feasibility: The application is economically feasible as it uses a lightweight Android application in low-cost handheld devices.

2) Technical Feasibility: Operating System – Windows 10 or higher, 64 bit or 32 bit , Processor – Intel Core i5 10th GEN , Ram – 8GB RAM , Hard Drive – minimum 16GB hard drive , Technology, Used – Python Language, Tools Used - Google Chrome, Google colab

3) Feasibility Plan: This assessment is most important for the success of the project because the project will fail if it is not completed on time. In feasibility planning, the organization estimates how much time it will take to complete the project. This app takes 5 and a half months to develop, so it's practically doable.

4) Operational Feasibility: It is defined as the process of evaluating the extent to which a proposed system solves business problems or exploits business opportunities. The solved problems and benefits of this system are listed in the documentation.

B. Methodology

The system is trained on the PyTorch deepfake detection model on an equal number of real and fake videos to avoid biasing the model. In the development phase, it took the dataset, pre-processes the dataset, and created a new processed dataset that contains only the cropped videos.

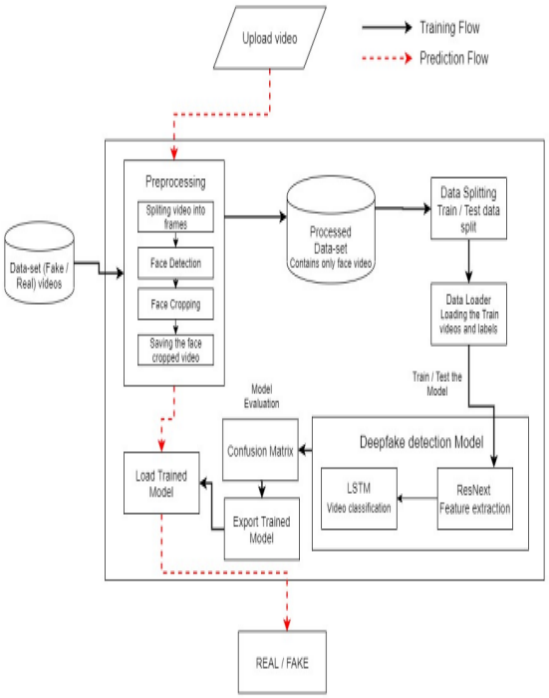

Fig. 1 Proposed Architecture

- Module 1: Dataset Collection

To make the model effective for real-time prediction. The data is collected from various datasets available such as FaceForensic.

- Module 2: Preprocessing

In this step, the videos are pre-processed and all unnecessary and noise is removed from them. Only the desired part of the video, i.e. the face, is detected and cropped. The first steps in video preprocessing is to divide the video into frames. After dividing the video into frames, a face is detected in each frame and the frame is cropped along the face. Later, the cropped frame is re-converted to a new video by joining each frame of the video. For each video, a process is followed that leads to the creation of a processed dataset containing videos with only faces. A frame that does not contain a face is ignored in preprocessing.

- Module 3: Data set partitioning

The dataset is divided into a train and a test dataset with a ratio of 70% train videos and 30% test videos. The split of train and test is balanced, i.e. 50% real and 50% fake videos in each split.

- Module 4: Module Architecture

The model is a combination of CNN and RNN. A pre-trained ResNext CNN model is used to extract features at the frame level, and based on the extracted features, an LSTM network is trained to classify the video as deepfake or pristine. Using the Data Loader on the training video partition, the video labels are loaded and fed into the model for training.

ResNext:
Instead of writing code from scratch, a pre-trained ResNext model is used for feature extraction. ResNext is a residual CNN optimized for high performance on deeper neural networks.

- Module 5: Tuning Hyperparameters

It is the process of selecting the perfect hyperparameters to achieve maximum accuracy.

C. System Analysis Models:

The System Analysis Model includes the Unified Modelling Language diagrams. A UML diagram is a diagram based on the UML (Unified Modelling Language) with the purpose of visually representing a system along with its main actors, roles, actions, or classes, in order to better understand, alter, maintain, or document information about the system. A UML diagram consists of Use Case Diagram, Activity Diagram, Class Diagram etc. some of the UML Diagram are given below.

a) Use Case Diagram:

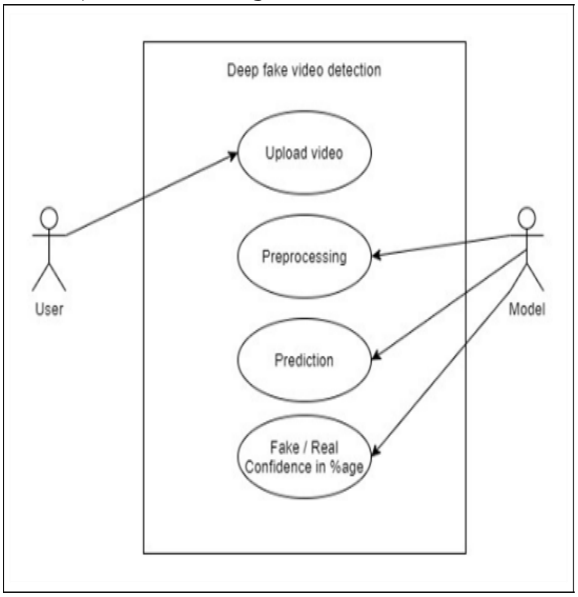

Fig. 2 Use case diagram

V. RESULT AND DISCUSSION

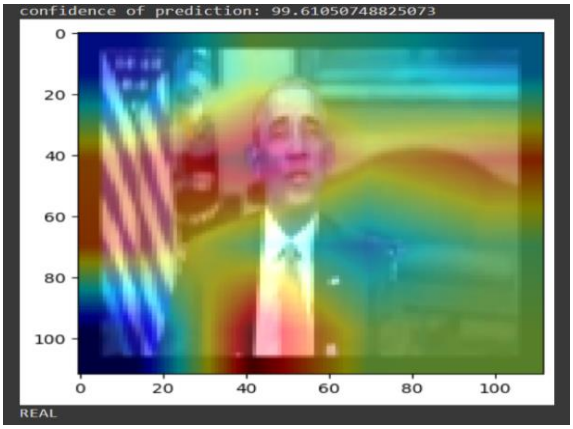

Fig. 3 Real video detected

A fake video has been given as input and it has detected that video is Real.

VI. CONCLUSION

The system presents a neural network-based approach to classify the video as deep fake or real, along with the confidence of proposed model. The system is capable of predicting the output by processing 1 second of video (10 frames per second) with a good accuracy. The implement the model by using pre-trained ResNext CNN model to extract the frame level features and LSTM for temporal sequence processing to spot the changes between the t and t-1 frame. The model can process the video in the frame sequence of 10,20,40,60,80,100.

VII. FUTURE SCOPE

There is always a scope for enhancements in any developed system, especially when the project build using latest trending technology and has a good scope in future

- Web based platform can be upscaled to a browser plugin for ease of access to the user.
- Currently only Face Deep Fakes are being detected by the algorithm, but the algorithm can be enhanced in detecting full body deep fakes.

REFERENCES

[1] M. Tarasiou and S. Zafeiriou, "Extracting Deep Local Features to Detect Manipulated Images of Human Faces," 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 1821-1825, doi: 10.1109/ ICIP 40778. 2020.9190714.

[2] Divya Babu and Uppala Santosh Kumar "Deepfake Video Detection using Image Processing and Hashing Tools" 2020 International Research Journal of Engineering and Technology (IRJET) Mar 2020, | ISSN 2395-0056

[3] Bozhi Xu, Jiarui Liu, Jifan Liang, Wei Lu1, "DeepFake Videos Detection Based on Texture Features",Computers, Materials & Continua (CMC) February 2021

[4] Abishek N , Shilpa H L, "Detection of the AI Generated Deepfake Video by using multi-task cascaded convolution neural network", JETIR August 2021, | ISSN-2349-516