

DEEPPFAKE DETECTION USING RESNEXT101 AND LSTM

Major Project Presentation by -

Shresth Gupta (RA2011026010091)

Utkarsh Srivastava (RA2011026010104)

PART ONE

ABSTRACT

ABSTRACT

- A new era in multimedia manipulation has actually been brought about by the widespread use of AI, machine learning, and deep learning technologies. This has produced both encouraging developments and unsettling concerns. Although there are many legal uses for these technologies, as in entertainment and education, their misuse has led to a number of problems, with deepfakes emerging as a particularly prominent concern.
- Deepfakes are powerful instruments for disseminating propaganda, creating political unrest, and disseminating false information. They are distinguished by realistic and high-quality video, picture, or audio modifications. These features are used by malicious individuals to produce content that looks legitimate, making it difficult to distinguish between fact and fiction.
- Researchers and engineers have investigated a number of strategies to deal with the problems caused by deepfakes. Creating sophisticated detection methods to recognize modified content is one approach. Deep neural networks and machine learning algorithms are used to examine patterns, discrepancies, and artifacts in multimedia files in an effort to discern real from fake information. .

PART TWO

INTRODUCTION

INTRODUCTION

- The term "deepfakes" refers to artificial media that is generated using deep learning algorithms. These algorithms have the ability to convincingly alter the audio and visual components of digital video, or image and they also frequently have malicious intentions.
- Deep learning, a type of machine learning that makes use of artificial neural networks with multiple layers, has shown promising results in a variety of computer vision applications. These applications include object recognition, image categorization, and natural language processing, among others.
- Deep learning approaches have the ability to differentiate between real and fake media in the context of deepfake detection. This is accomplished by recognizing intricate patterns and characteristics that are present in the data. This work also tackles the open research problems and challenges in the field of deepfake detection. These include the demand for rigorous assessment criteria, large-scale datasets, and the capacity to apply to manipulation techniques that have not yet been encountered.

PART THREE

MOTIVATION

MOTIVATION

- Since 2016 after the discovery of deepfakes, a lot of internet users have been suffering from cybercrimes that are done using fake content.
- For a long time the only method of detection of such content was after it had reached the mass media and after various complaints, the content was taken down.
- The techniques that were used were remotely accurate and were not available for devices with lower computational power thus not providing access to the multitude.
- Big companies such as Google, Facebook and Microsoft came together to collaborate their collected data and issue a tech-based innovation challenge called the Deepfake detection challenge which focused on developing an efficient method to detect deepfake content on social media.

PART FOUR

INNOVATION

INNOVATION

- Using ResNext model along LSTM architecture to reduce the computational complexity and thus giving an increased scalability.
- The algorithm used will be trained and tested against a dataset of lower complexity as the goal is to obtain a simple classifier algorithm that is easily deployable.
- The algorithm once tested will be incorporated into a web application that can be deployed by devices of lower computational power providing ease of access.

PART FIVE

SPRINT PLANNING

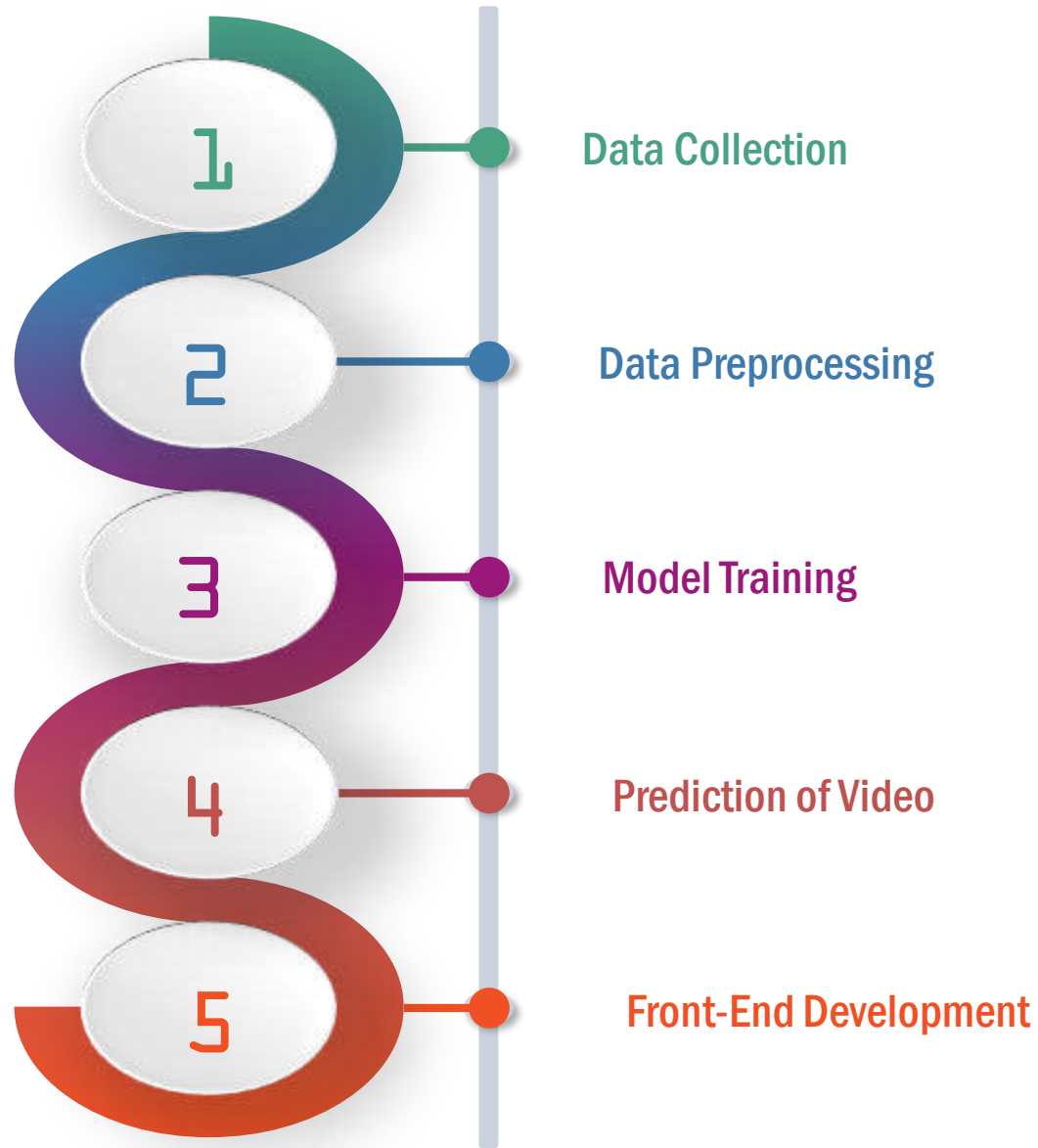
SPRINT PLANNING

- SPRINT 1: Data Collection and Preprocessing
 - Research existing deepfake datasets
 - Gather labeled authentic and deepfake videos
 - Preprocess collected data
 - Create pipeline for data loading and preprocessing
- SPRINT 2 : Model Creation and Training
 - Research deep learning architectures
 - Experiment with different models
 - Train selected model on preprocessed data
 - Evaluate model performance
 - Fine-tune models for better accuracy
- SPRINT 3 : Prediction and Front-End Development
 - Develop prediction pipeline integrating trained model
 - Implement front-end for user interaction
 - Integrate prediction functionality with front-end
 - Test end-to-end functionality

PART SIX

PROCESS WORKFLOW

PROCESS WORKFLOW



DATA COLLECTION

For our proposed deepfake detection system, we are using a combination of two large datasets which are widely used:

Celeb-DF: In the field of deepfake detection research, the Celeb-DF dataset is a valuable resource that offers an extensive compilation of videos that showcase celebrity faces that have been altered by deepfake methods.

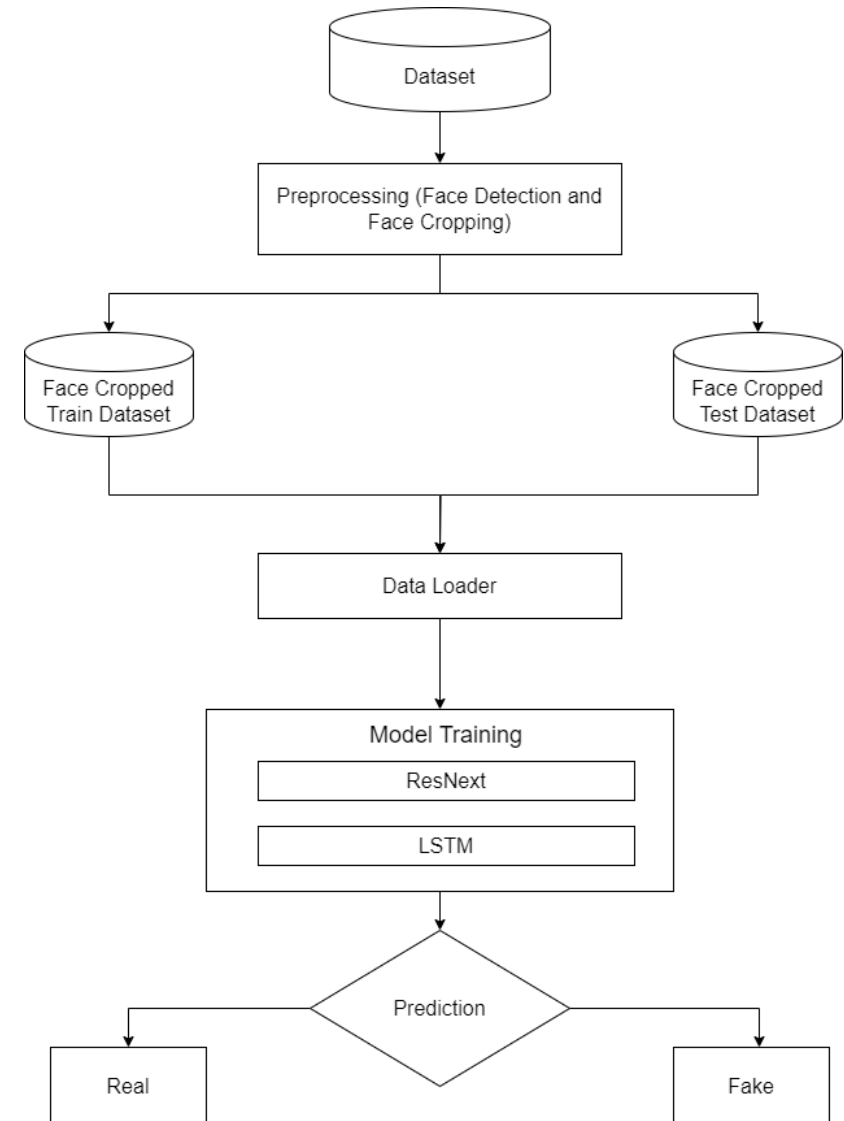
DeepFake Detection Challenge: A crucial tool in the field of deepfake detection is the Deepfake Detection Challenge (DFDC) dataset, which was created to encourage study and creativity in order to stop the spread of artificial intelligence (AI)- generated altered media. Launched as part of a collaboration between industry leaders and academic institutions, the DFDC dataset offers a diverse and comprehensive collection of deepfake videos and corresponding real videos, creating a benchmark for evaluating deepfake detection algorithms.

DATA PREPROCESSING

- In order to prepare the videos for deepfake detection analysis, a comprehensive preprocessing pipeline has been developed. This pipeline is designed to extract facial regions of interest (ROIs) from raw video data, a crucial step in ensuring the accuracy and efficiency of subsequent deepfake detection algorithms.
- The process initiates with the extraction of frames from the input video utilizing the OpenCV library.
- Each frame is then subjected to a face detection algorithm provided by the face_recognition library. This algorithm accurately identifies and localizes faces within the frames, essential for isolating the relevant facial regions for analysis.
- After cropping of faces within each frame, these face cropped frames are converted back into a video

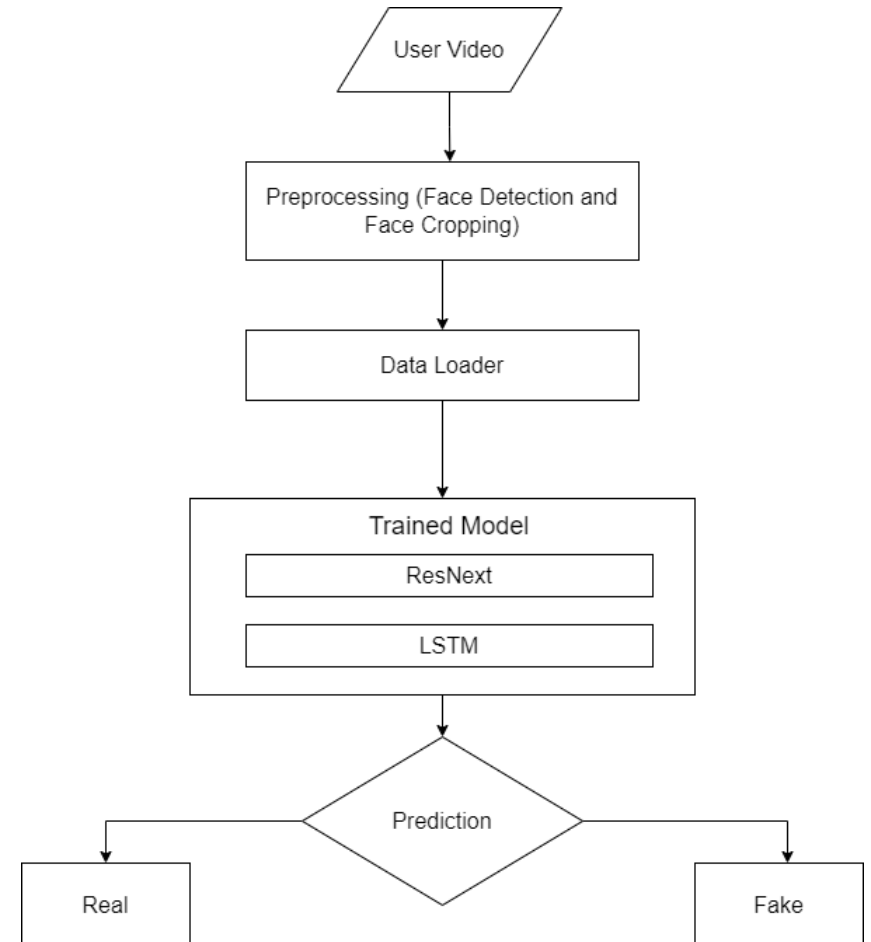
MODEL TRAINING

- Leveraging a powerful Residual Network Convolutional Neural Network (CNN) backbone, specifically the ResNeXt-101 64x4d model pretrained on large-scale image datasets, the model exhibits a robust capacity to extract high-level features from input videos.
- We also take advantage of Long Short Term Memory (LSTM), which has the ability to process sequential data used to capture temporal dependencies and nuances in the videos.



PREDICTION

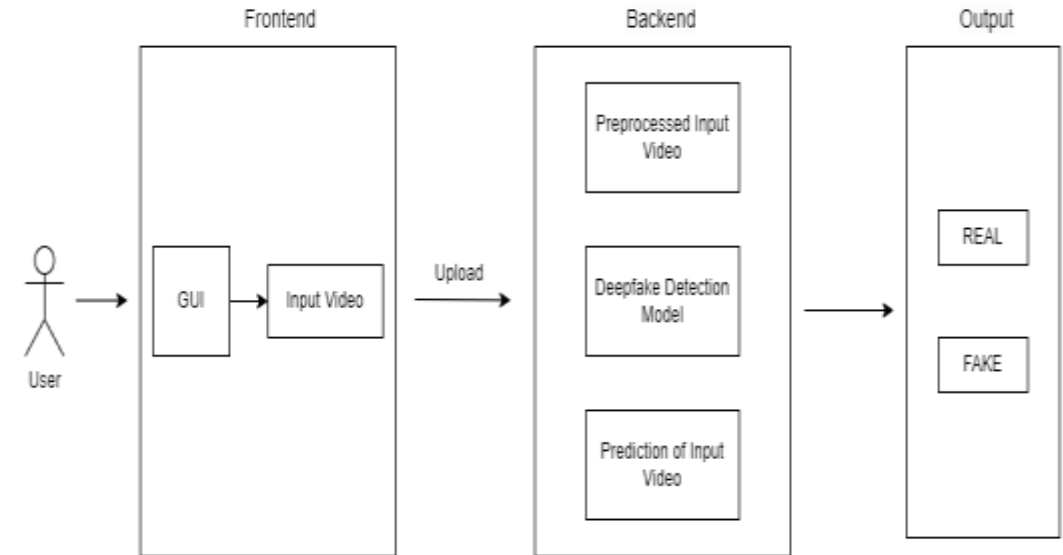
- The trained deepfake detection model is applied to a fresh video in order to make predictions.
- The new video is preprocessed to conform to the format that the trained model expects before it is predicted. The model and the preprocessed video frames are the inputs used to invoke the predict function at first. The anticipated output is then determined by taking the class with the highest probability and assigning it a confidence score that represents the model's level of conviction.
- This prediction process enables the detection of deepfake manipulations in the input video, providing valuable insights into the authenticity of visual media.



FRONT-END DEVELOPMENT

The application is divided into three main components: Frontend, Backend, and Output.

- Frontend: Represents the user interface where the user interacts with a graphical user interface (GUI) to upload an input video.
- Backend: The input video is then uploaded to the backend where it undergoes preprocessing. After preprocessing, the video is passed through a deepfake detection model, which predicts whether the input video is real or fake.
- Output: The final prediction is then presented to the user, with the output clearly labeled as either 'REAL' or 'FAKE'.



PART SEVEN

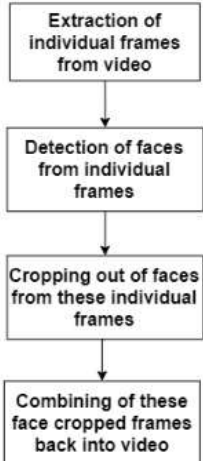
ARCHITECTURE DIAGRAM

ARCHITECTURE DIAGRAM

VIDEO DATASET



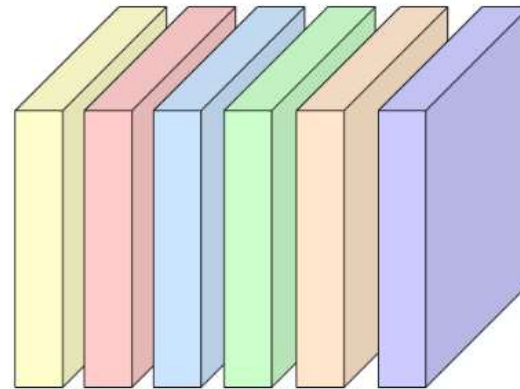
PREPROCESSING









Face Cropped Frames
of the Dataset



DEEFAKE DETECTION MODEL



- | | |
|--|---|
|  ResNext101 layer |  Dropout layer |
|  LSTM layer |  Linear layer |
|  LeakyRELU layer |  Pooling layer |

SOFTMAX LAYER



REAL

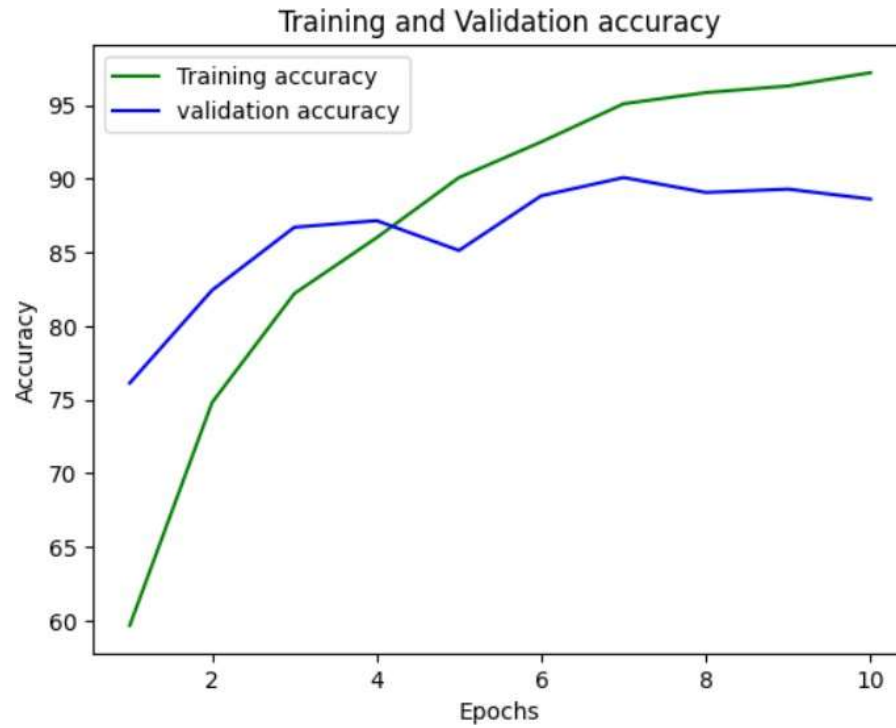


FAKE

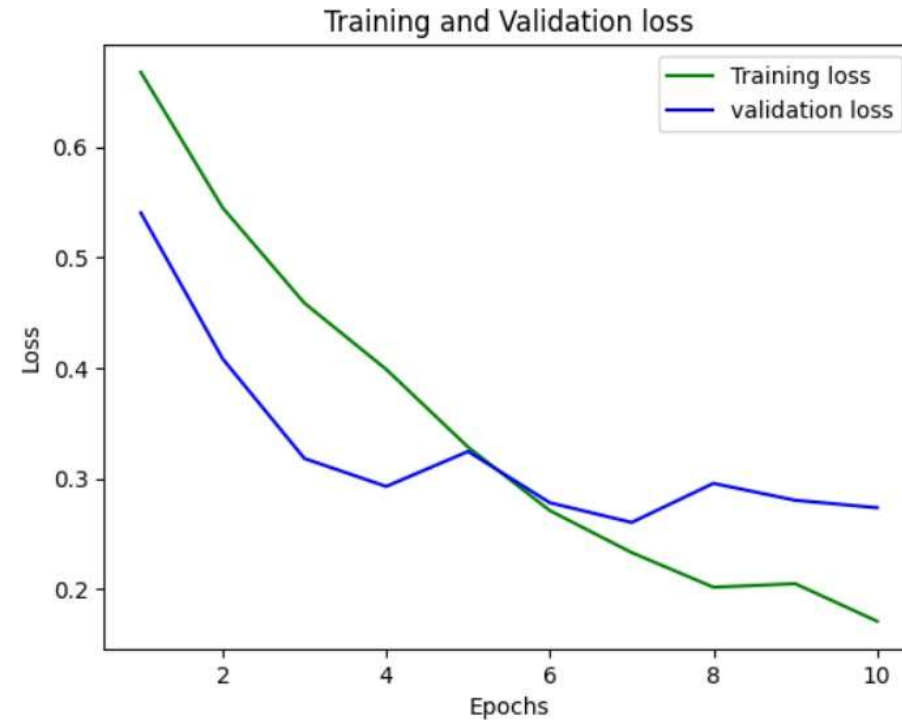
PART EIGHT

RESULTS AND DISCUSSION

RESULTS AND DISCUSSIONS



Training and Validation Accuracy



Training and Validation Loss

PART NINE

OUTPUT SCREENSHOT

OUTPUT SCREENSHOT



The user interface where the user provides an input video for Deepfake Video Detection. After the video is successfully loaded, the user clicks on “Upload” button, to look for whether the video is real or fake



The desired output for the given video. It displays the face cropped images of each of the 40 frames extracted from the video along with the prediction whether it is “REAL” or “FAKE” and the confidence of the model.