Malnad College of Engineering, Hassan

(An Autonomous Institution affiliated to VTU, Belgavi)

A Project Report

On

"Deep Fake Detection System"

Submitted in partial fulfillment of
the requirements for the award of the degree of

Bachelor of Engineering
in
Computer Science and Engineering

Submitted by

Gurukiran D P	4MC20CS050
Joywin Monteiro	4MC20CS056
Karan Anjan	4MC20CS057
Kushal C	4MC20CS065

Under the guidance of
Mrs. Nivyashree R
Assistant Professor

Department of Computer Science and Engineering

2023-2024

Malnad College of Engineering

Department of Computer Science and Engineering
Hassan - 573201, Karnataka, India

Certificate

This is to certify that mini project work entitled **"Deep Fake Detection System "** is a bonafide work carried out by **Gurukiran D P (4MC20CS050), Joywin Monteiro (4MC20CS056), Karan Anjan (4MC20CS057) and Kushal C (4MC20CS065)** in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering of the Visvesvaraya Technological University, Belgavi during the year 2023-2024. The project report has been approved as it satisfies the academic requirements in respect of mini project work prescribed for the Bachelor of Engineering Degree.

Signature of the Guide	Signature of the HOD	Signature of the Principal
Mrs. Nivyashree R	Dr. Geetha Kiran A	Dr. A J Krishnaiah
Assistant Professor	Prof. & HOD	Principal
Dept. of CSE, MCE	Dept. of CSE, MCE	MCE

Examiners

Name of the Examiner	Signature of the Examiner
1.	
2.	

ABSTRACT

Deep Fake Detection System employs deep learning techniques to identify manipulated media content known as deepfakes. Deepfakes are synthetic media generated using advanced AI algorithms, posing challenges to media authenticity. Our project aims to develop a robust system capable of distinguishing between genuine and manipulated content to combat the spread of misinformation and ensure media integrity. Through a comprehensive assessment of deep learning-based detection methods, we aim to enhance the accuracy and efficiency of deepfake detection. By leveraging machine learning algorithms and advanced neural networks, our system seeks to provide real-time detection capabilities, contributing to the ongoing efforts in addressing the proliferation of deepfakes in digital media.

Keywords: Deep Fake Detection, Deep Learning, Media Integrity, Misinformation, Neural Networks

ACKNOWLEDGEMENTS

We present with immense pleasure this work titled **"Deep Fake Detection System"**. An endeavour over a long period can be successful with the advice and support of many well wishers.

We take this opportunity to express our gratitude and appreciation to all of them. The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without mentioning the people who made it possible. So, with gratitude we acknowledge all those whose guidance and encouragement made to successfully complete this project.

We would like to express sincere thanks to our Principal **Dr. A J Krishnaiah**, Malnad College of Engineering for his encouragement made to successfully completion of the project work. We wish to express our gratitude to **Dr. Geetha Kiran A**, Professor and Head, Department of Computer Science & Engineering for providing a good working environment and for her constant support and encouragement.

It gives great pleasure in placing on record a deep sense of gratitude to our guide **Mrs. Nivyashree R**, Assistant Professor, Department of Computer Science & Engineering for her daily evaluation of the work and for providing us constant encouragement with her unflinching support and valuable guidance throughout this project.

We would also like to thank all the staff of Computer Science and Engineering department who have directly or indirectly helped us in the completion of the project work. At last we would hereby acknowledge and thank our parents who have been a source of inspiration and also instrumental in the successful completion of the project work.

Gurukiran D P

Joywin Monteiro

Karan Anjan

Kushal C

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

Deep fake is a technique for human image synthesis based on neural network tools 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. In the world of ever growing Social media platforms, Deepfakes are con- sidered as the major threat of the AI. There are many Scenarios where these realistic face swapped deepfakes are used to create political distress, fake terrorism events, revenge porn, blackmail peoples are easily envisioned.Some of the examples are Brad Pitt, Angelina Jolie videos. It becomes very important to spot the difference between the deepfake and pristine video. We are using AI to fight AI.Deepfakes are created using tools like Face App and Face Swap , which using pretrained neural networks like GAN or Auto encoders for these deepfakes creation. Our method uses a LSTM based artificial neural network to process the sequential temporal analysis of the video frames and pre-trained Res-Next CNN to extract the frame level features. Convolution neural network extracts the frame level features and these features are further used to train the Long Short Term Memory based artificial Recurrent Neural Network to classify the video as Deepfake or real. To emulate the real time scenarios and make the model perform better on real time data, we trained our method with large amount of balanced and combination of various available dataset like FaceForensic++, Deepfake detection challenge, and Celeb-DF. Further to make the ready to use for the customers, we have developed a front end application where the user the user will upload the video. The video will be processed by the model and the output will be rendered back to the user with the classification of the video as deepfake or real and confidence of the model.

1.1 Overview or background and motivation

The increasing sophistication of mobile camera technology and the ever growing reach of social media and media sharing portals have made the creation and propagation of digital videos more convenient than ever before. Deep learning has given rise to technologies that would have been thought impossible only a handful of years ago.

Modern generative models are one example of these, capable of synthesizing hyper realistic images, speech, music, and even video. These models have found use in a wide variety of applications, including making the world more accessible through text-to-speech, and helping generate training data for medical imaging. Like any trans-formative technology, this has created new challenges. So called "deep fakes" produced by deep generative models that can manipulate video and audio clips. Since their first appearance in late 2017, many open-source deep fake generation methods and tools have emerged now, leading to a growing number of synthesized media clips. While many are likely intended to be humorous, others could be harmful to individuals and society. Until recently, the number of fake videos and their degrees of realism has been increasing due to availability of the editing tools, the high demand on domain expertise. Spreading of the Deep fakes over the social media platforms have become very common leading to spamming and peculating wrong information over the platform. Just imagine a deep fake of our prime minister declaring war against neighboring countries, or a Deep fake of reputed celebrity abusing the fans. These types of the deep fakes will be terrible, and lead to threatening, misleading of common people. To overcome such a situation, Deep fake detection is very important. So, we describe a new deep learning-based method that can effectively distinguish AI- generated fake videos (Deep Fake Videos) from real videos. It's incredibly important to develop technology that can spot fakes, so that the deep fakes can be identified and prevented from spreading over the internet.

1.1.1 Problem Statement

Convincing manipulations of digital images and videos have been demonstrated for several decades through the use of visual effects, recent advances in deep learning have led to a dramatic increase in the realism of fake content and the accessibility in which it can be created. These so called AI-synthesized media (popularly referred to as deep fakes).Creating the Deep Fakes using the Artificial intelligent tools are simple task. But, when it comes to detection of these Deep Fakes, it is major challenge. Already in the history there are many examples where the deepfakes are used as powerful way to create political tension, fake terrorism events, blackmail peoples etc.So it becomes very important to detect these deepfake and avoid the percolation of deepfake through social media platforms.

1.1.2 Research Objectives

- Our project aims at discovering the distorted truth of the deep fakes.
- Our project will reduce the Abuses' and misleading of the common people on the world wide web.
- Our project will distinguish and classify the video as deepfake or pristine.

- Provide a easy to use system for used to upload the video and distinguish whether the video is real or fake.

1.1.3 Research Significance

- **Enhanced Security:** Deep face detection systems play a pivotal role in bolstering security measures in various domains such as law enforcement, border control, and access control systems .
- **Advancements in AI:** Research in deep face detection contributes to the evolution of artificial intelligence (AI) and deep learning techniques, facilitating more accurate and efficient face recognition algorithms .
- **Real-world Applications:** These systems find practical applications in diverse fields, including surveillance, biometric authentication, forensic analysis, and human-computer interaction, thereby bridging the gap between theoretical advancements and real-world utility .
- **Challenges Addressed:** Researchers tackle challenges such as variations in pose, lighting conditions, occlusions, and facial expressions, leading to the development of robust algorithms capable of accurate detection under diverse circumstances .
- **Ethical Considerations:** As deep face detection systems become more pervasive, the research also delves into ethical considerations regarding privacy, bias mitigation, and potential misuse, fostering discussions on responsible deployment and regulation .
- **Future Prospects:** Continued research in this domain holds promise for further innovation, potentially enabling novel applications such as emotion recognition, medical diagnostics, and augmented reality interfaces, thereby shaping the future of human-machine interaction

1.1.4 Research Outcomes

There are many tools available for creating the deep fakes, but for deep fake detection there is hardly any tool available. Our approach for detecting the deep fakes will be great contribution in avoiding the percolation of the deep fakes over the world wide web. We will be providing a web-based platform for the user to upload the video and classify it as fake or real. This project can be scaled up from developing a web-based platform to a browser plugin for automatic deep fake detection's. Even big application like WhatsApp, Facebook can integrate this project with their application for easy pre detection of deep fakes before sending to another user.

Chapter 2

Literature Survey

Anjali Mahantesh Mudakavi and et al.,[1]A slew of machine learning based software has appeared to facilitate the seamless creation of convincingly altered facial features in videos, a phenomenon often referred to as deepfake Video. These manipulations are so sophisticated that they often leave little trace of tampering, giving rise to concerns about possible misuse in various scenarios, such as inciting political unrest, orchestrating blackmail schemes or orchestrating fake terrorist activity. The purpose of this article is to provide a comprehensive overview of recent research projects focused on the comprehensive detection of fraudulent content using advanced deep learning methods.They aim to build research excellence through methodical research of different categories in the field of fake content detection. Many studies have already delved into the development of detection methods aimed at mitigating the potential harmful effects associated with widespread deep forgery. The application of neural networks and deep learning stands out as a central approach to reformulate these sentences and deal with the multifaceted challenges presented by the prevalence of deep falsification.

Saima Waseem and et al.,[2] Advancements in facial manipulation technology have resulted in highly realistic and indistinguishable face and expression swap videos. However, this has also raised concerns regarding the security risks associated with deepfakes. In the field of multimedia forensics, the detection and precise localization of image forgery has become essential tasks. Current deepfake detectors perform well with high quality faces within specific datasets, but often struggle to maintain their performance when evaluated across different datasets. To this end, we propose an attention-based multi task approach to improve feature maps for classification and localization tasks. The encoder and the attention-based decoder of our net- work generate localized maps that highlight regions with information about the type of manipulation. These localized features are shared with the classification network, improving its performance. Instead of using encoded spatial features, attention based localized features from the decoder's first layer are combined with frequency domain features to create a discriminative representation for deepfake detection. Through extensive experiments on face and expression swap datasets, we demonstrate that our method achieves competitive performance in comparison to state of the art deepfake detection approaches in both in-dataset and cross-dataset scenarios

Mohamad Nur Nobi and et al.,[3]AI, machine learning, and deep learning has resulted in new techniques and various tools for manipulating multimedia. Though the technology has been mostly used in legitimate applications such as for entertainment and education, etc., malicious users have also exploited them for unlawful or nefarious purposes. For example, high-quality and realistic fake videos, images, or audios have been created to spread misinformation and propaganda, foment political discord and hate, or even harass and blackmail people. The manipulated, high quality and realistic videos have become known recently as Deepfake. Various approaches have since been described in the literature to deal with the problems raised by Deepfake. To provide an updated overview of the research works in Deepfake detection, we conduct a systematic literature review (SLR) in this paper, summarizing 112 relevant articles from 2018 to 2020 that presented a variety of methodologies. We analyze them by grouping them into four different categories: deep learning-based techniques, classical machine learning based methods, statistical techniques, and blockchain based techniques. We also evaluate the performance of the detection capability of the various methods with respect to different datasets and conclude that the deep learning based methods outperform other methods in Deepfake detection.

Yuezun and et al.,[4]AI-synthesized face-swapping videos, commonly known as DeepFakes, is an emerging problem threatening the trustworthiness of online information. The need to develop and evaluate DeepFake detection algorithms calls for large scale datasets. However, current DeepFake datasets suffer from low visual quality and do not resemble Deep Fake videos circulated on the Internet. We present a new large scale challenging DeepFake video dataset, Celeb- DF, which contains 5, 639 high quality DeepFake videos of celebrities generated using improved synthesis process. We conduct a comprehensive evaluation of DeepFake detection methods and datasets to demonstrate the escalated level of challenges posed by Celeb-DF.

Davide Cozzolino and et al.,[5]The rapid progress in synthetic image generation and manipulation has now come to a point where it raises significant concerns for the implications towards society. At best, this leads to a loss of trust in digital content, but could potentially cause further harm by spreading false information or fake news. This paper examines the realism of state of the art image manipulations, and how difficult it is to detect them, either automatically or by humans.

Chapter 3

Project Design

3.1 Deep Fake Generation

In essence, this system is trying to learn what real and fake videos look like by analyzing a large dataset of videos. Once it has learned these patterns, it can then apply them to new videos to determine whether or not they are likely to be fake. To detect the deepfake videos it is very important to understand the creation process of the deepfake. Majority of the tools including the GAN and autoencoders takes a source image and target video as input. These tools split the video into frames , detect the face in the video and replace the source face with target face on each frame. Then the replaced frames are then combined using different pre-trained models. These models also enhance the quality of video my removing the left-over traces by the deepfake creation model. Which result in creation of a deepfake looks realistic in nature. We have also used the same approach to detect the deepfakes. Deepfakes created using the pretrained neural networks models are very realistic that it is almost impossible to spot the difference by the naked eyes. But in reality, the deepfakes creation tools leaves some of the traces or artifacts in the video which may not be noticeable by the naked eyes. The motive of this paper to identify these unnoticeable traces and distinguishable artifacts of these videos and classified it as deepfake or real video. Video manipulation detection is a challenging task, and there is no foolproof way to do it. However, systems like the one in the image can be effective at identifying some types of video manipulation. As video manipulation techniques become more sophisticated, so too will video manipulation detection systems. It is an ongoing arms race between the creators of deepfakes and the creators of deepfake detection algorithms.

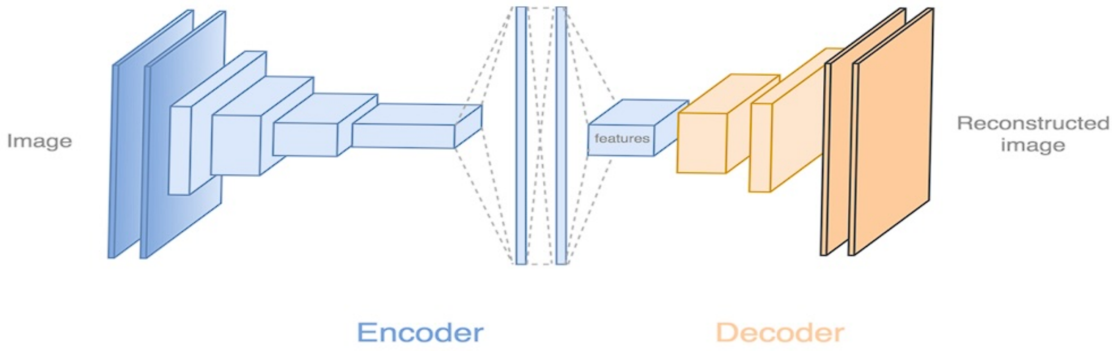

Figure 3.1: Encoder-Decoder Process

Figure 3.2: Detection Process

3.2 Proposed Methodology

Figure 3.1 Shows the Data flow Diagram of the System

- **Upload Video:**In this initial step, a video file is uploaded to the system.
- **Preprocessing:**The uploaded video undergoes preprocessing to prepare it for facial recognition. This stage can involve several steps including:
  - **Splitting the video into individual frames:**A video is made up of a sequence of images called frames. Splitting the video into frames separates the video into its constituent images, making it easier to analyze each frame independently for the presence of faces.
  - **Face detection:**Each frame is scanned to identify the presence and location of faces within the frame.

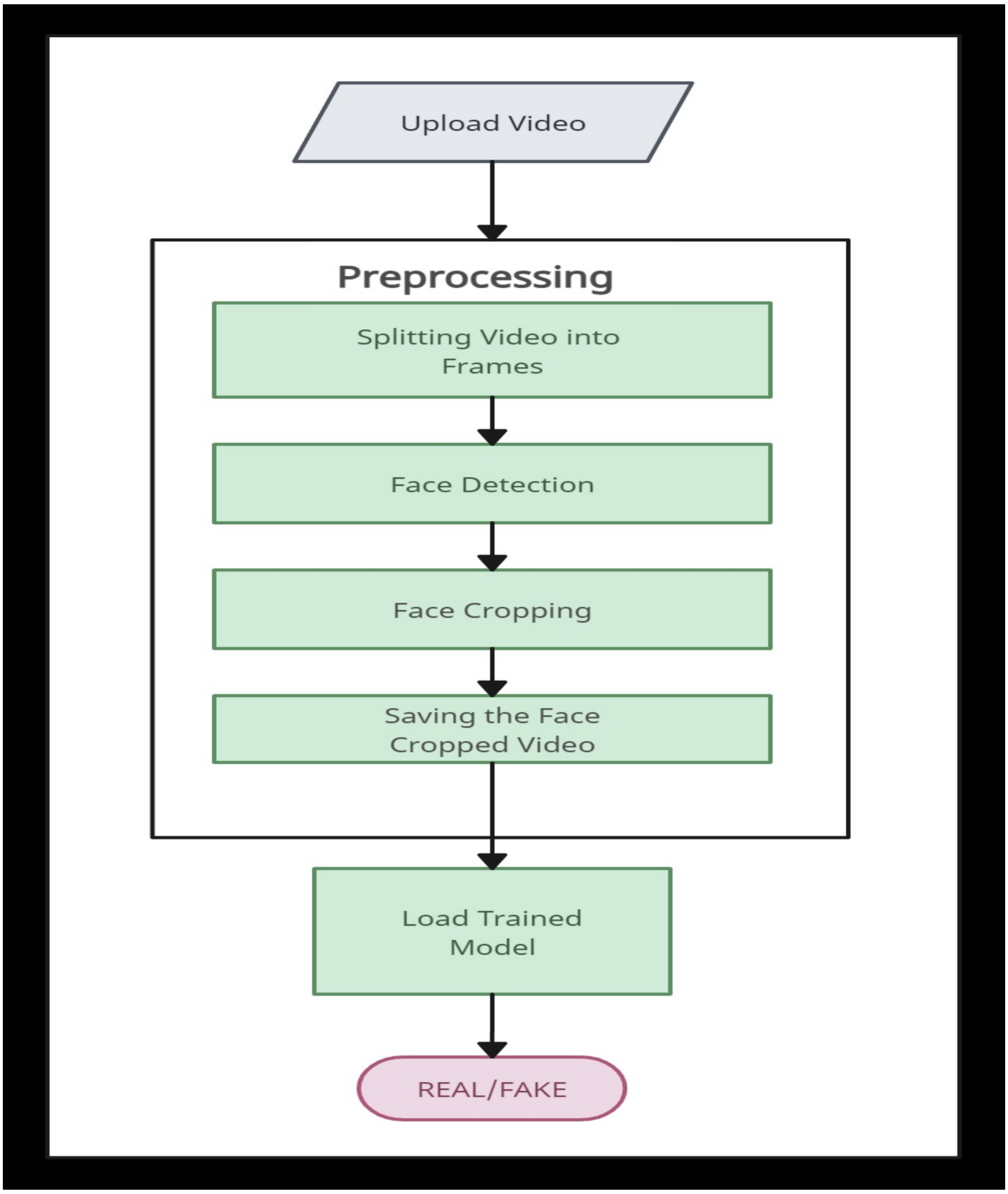

Figure 3.3: Data flow Diagram

- **Face cropping:**Once a face is detected in a frame, the system crops or extracts the facial region from the frame, discarding irrelevant background data.
- **Saving the face cropped video:**The cropped faces are then saved into a new video file.
- **Load Trained Model:**A facial recognition model, which has been trained on a large dataset of facial images and their corresponding identities, is loaded into the system.

- **Real/Fake:**The system compares the faces in the preprocessed video with the faces in the trained model to determine if they are real or fake.

3.3 System Architecture

- **Input Video:**The first step is to input the video into the computer system.
- **Pre-processing:**The video is then converted into a format that can be analyzed by the manipulation detection algorithm. This may involve resizing the video or converting it into a different color space.
- **Frame Extraction:**The video is then broken down into individual frames. Each frame is a single image that makes up the video.
- **Encoder:** Here, each frame is passed through an encoder. An encoder is a type of neural network that is used to compress information. In this case, the encoder is used to compress the video frame into a lower-dimensional representation.
- **Face Detection, Cropping and Alignment:**The system then detects any faces in the frame. Once a face is detected, it is cropped from the frame and aligned in a specific way. This is done because the manipulation detection algorithm is often more effective at detecting manipulations in faces than in other parts of the image.
- **CNN (Convolutional Neural Network):**The cropped and aligned face image is then passed through a convolutional neural network (CNN). A CNN is a type of neural network that is specifically designed for image recognition. The CNN is used to extract features from the image that can be used to determine whether or not the image has been manipulated.
- **RNN (Recurrent Neural Network):** The output of the CNN is then passed through a recurrent neural network (RNN). An RNN is a type of neural network that is able to process sequential data. In this case, the RNN is used to process the sequence of features extracted from the video frames.
- **Output:** Fake Video Percentage: The RNN then outputs a percentage score that indicates the likelihood that the video is a fake. A higher score indicates that the video is more likely to be a fake.

3.4 Front End

3.4.1 Flask

Flask is a Python web framework for building web applications. Lightweight and beginner friendly, it lets you define routes for URLs and write Python functions to handle them. Templates generate dynamic HTML content. Flask manages form

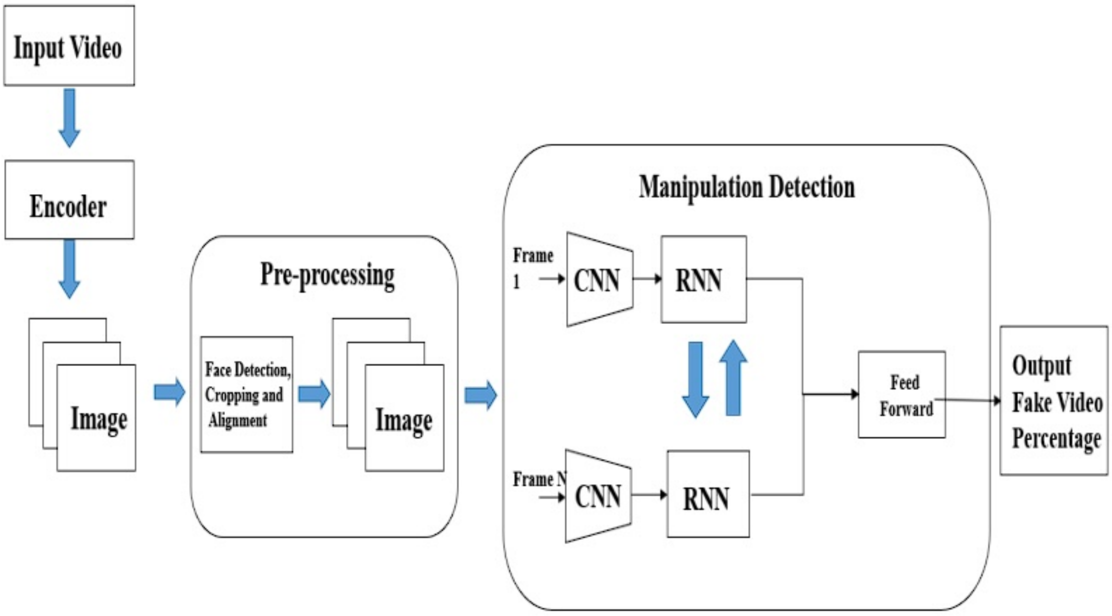

Figure 3.4: System Architecture

submissions and data, while extensions add features like databases or user logins. Ideal for smaller web apps or prototypes, Flask is even used by companies like Netflix for specific functionalities. Explore the Flask documentation to unleash its potential.

- **Micro framework:** Flask is a lightweight and minimalistic web framework for Python, offering essential tools to build web applications without unnecessary features.
- **Routing and URL Handling:** Flask uses decorators to define routes and map them to view functions, making it easy to handle different URLs and HTTP methods.
- **Modular and Extensible:** Flask follows a modular design and supports extensions for integrating additional features like database interactions, form handling, and authentication.
- **Built-in Development Server:** Flask includes a built-in development server for quick testing and debugging of applications, enabling rapid iteration during development.

Chapter 4

System Requirements

4.1 Hardware Reuirements

- Intel Xeon E5 2637- 3.5 GHz
- RAM -8 GB
- Hard Disk -100 GB
- Graphic card- NVIDIA GeForce
- GTX Titan (12 GB RAM)

4.2 Software Requirements

- Operating System: Windows 7+
- Programming Language : Python 3.0
- Framework: PyTorch 1.4
- Libraries : OpenCV, Face- recognition

Chapter 5

Results

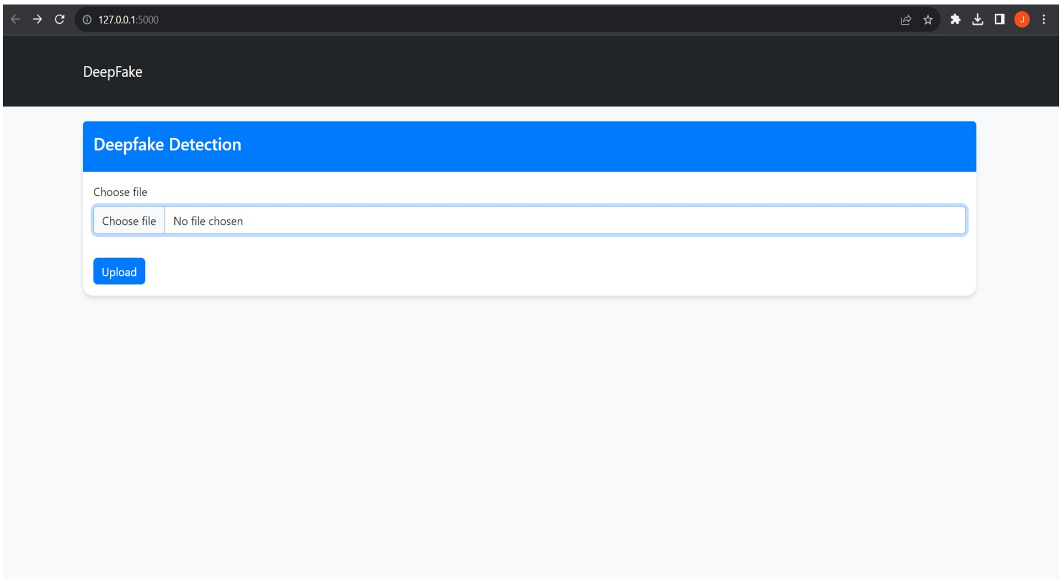

Figure 5.1: User Interface

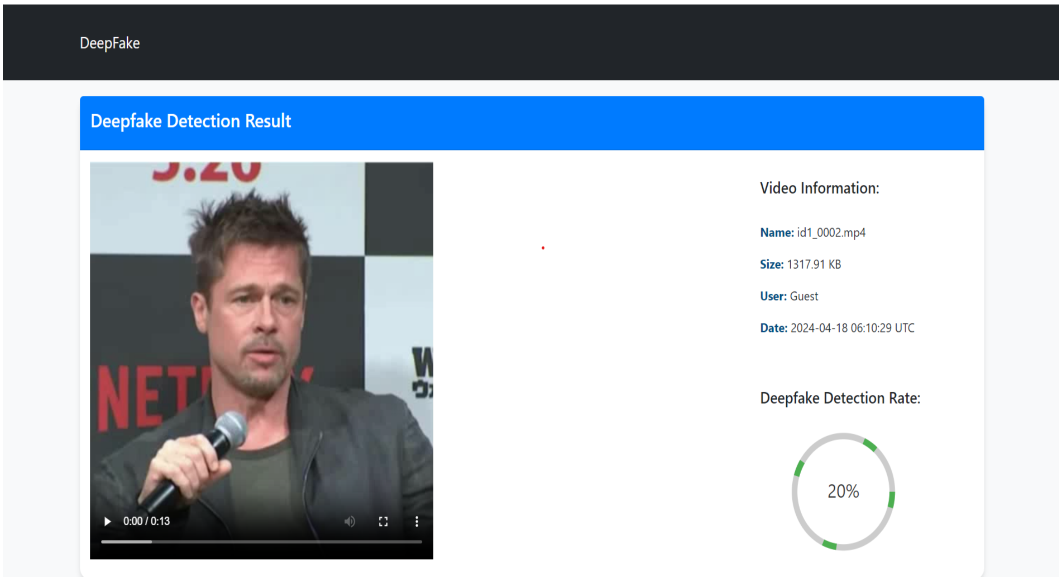

Figure 5.2: Original Video detected

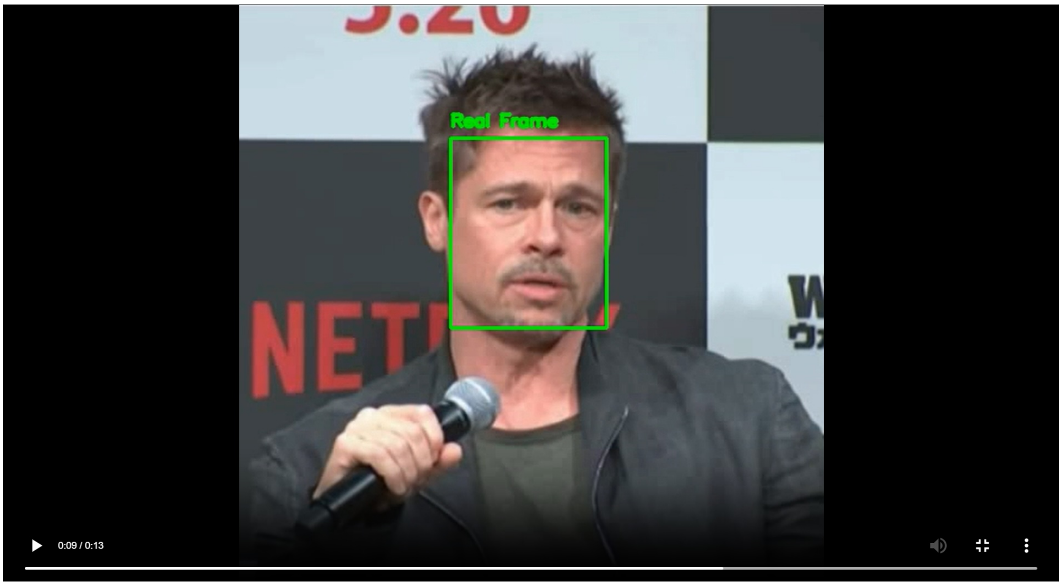

Figure 5.3: Original Video

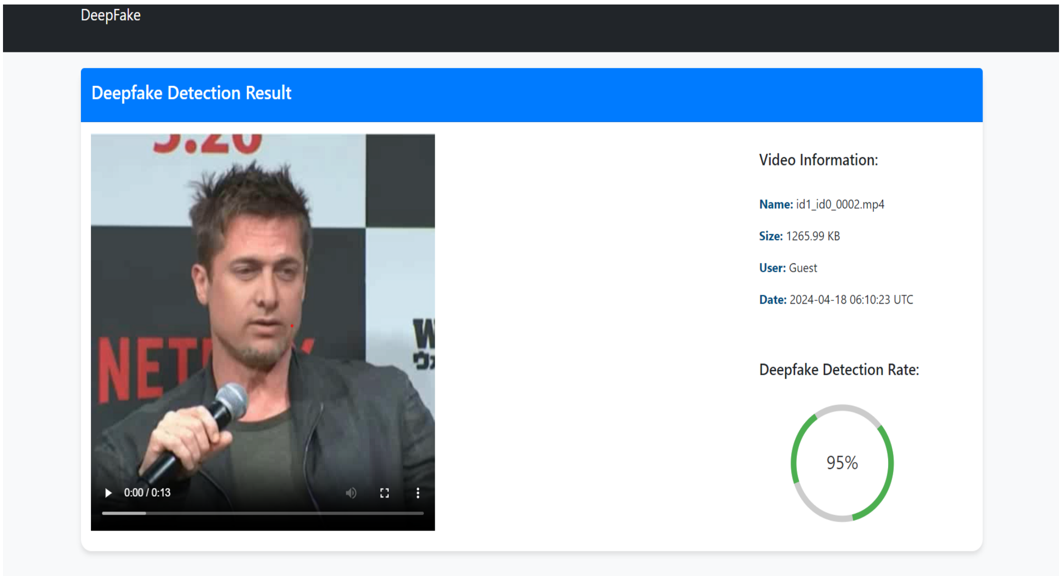

Figure 5.4: Deepfake Video Detected

Figure 5.5: Fake Video

Chapter 6

Conclusion

We introduced a neural network-driven strategy for distinguishing between genuine and deep fake videos, providing confidence scores for the model's predictions. Our approach demonstrates proficiency in analyzing one second of video data, comprising 10 frames per second, with high accuracy. Leveraging a pre-trained MTCNN (Multi-Task Cascaded Convolutional Neural Network), we extract facial features from each frame and employ an LSTM (Long Short-Term Memory) network for temporal sequence analysis to discern alterations between consecutive frames (t and t-1). Our model operates effectively on video sequences with frame intervals of 10, 20, 40, 60, 80, and 100 frames, ensuring robust detection capabilities across various scenarios. In addition to its robust performance across various frame intervals, our model showcases versatility in real-world applications, offering reliable detection of deep fake videos while providing interpretable confidence scores. By leveraging the pre-trained MTCNN for efficient facial feature extraction and LSTM for temporal sequence processing, our approach exhibits scalability and adaptability to different video lengths and complexities. Furthermore, our method contributes to the ongoing efforts in combating the proliferation of synthetic media by providing an effective tool for identifying manipulated content with high precision and confidence. Through continuous refinement and integration with existing video analysis frameworks, our model holds promise for enhancing the security and integrity of digital media platforms in the face of evolving threats posed by deep fake technologies.

APPENDIX A

Program

```
from flask import Flask, render_template, request, redirect, url_for
import os
from datetime import datetime
import json
from time import time as current_time
import importlib

app = Flask(__name__)

UPLOAD_FOLDER = 'static/videos'
app.config['UPLOAD_FOLDER'] = UPLOAD_FOLDER

@app.route('/')
def index():
    return render_template('index.html')

# Handle file upload and redirect to the result page
@app.route('/upload', methods=['POST'])
def upload_file():
    if 'file' not in request.files:
        return redirect(request.url)

    file = request.files['file']

    if file.filename == '':
        return redirect(request.url)

    if file:
        timestamp = int(current_time())
        filename = f"uploaded_video_{timestamp}.mp4"
```

```
        video_path = os.path.join(app.config['UPLOAD_FOLDER'], filename)
        file.save(video_path)

        video_path2 = os.path.join(app.config['UPLOAD_FOLDER']

        module = importlib.import_module("deepfake_detector")
        function = getattr(module, "run")

        result_from_det = function(video_path ,video_path2)
        print(result_from_det)

        # Get video information
        video_info = {
            'name': file.filename,
            'size': f"{os.path.getsize(video_path) / (1024):.2f} KB",
            'user': 'Guest',
            'source': datetime.utcnow().strftime('%Y-%m-%d %H:%M:%S UTC'),
            'per': result_from_det
        }

        video_info_json = json.dumps(video_info)



@app.route('/result')
def result():
    video_info_json = request.args.get('video_info')
    video_path2 = request.args.get('video_path2')
    print(video_path2)

    video_info = json.loads(video_info_json)
    print(video_info['name'])

    return render_template('result.html', video_url=video_path2)

if __name__ == '__main__':
    app.run(debug=True)
import cv2
import numpy as np
from facenet_pytorch import MTCNN, InceptionResnetV1
from torchvision.transforms import functional as F
```

```
import time

def run(video_path , video_path2):

    start_time = time.time()

    # Equivalents for deepfake detection
    threshold_face_similarity = 0.99
    threshold_frames_for_deepfake = 15

    mtcnn = MTCNN()
    facenet_model = InceptionResnetV1(pretrained='vggface2').eval()
    cap = cv2.VideoCapture(video_path)
    frame_count = 0
    fps = int(cap.get(cv2.CAP_PROP_FPS))
    width = int(cap.get(cv2.CAP_PROP_FRAME_WIDTH))
    height = int(cap.get(cv2.CAP_PROP_FRAME_HEIGHT))
    fourcc = cv2.VideoWriter_fourcc(*'H264')
    out = cv2.VideoWriter(video_path2, fourcc, fps, (width, height))

    deepfake_count = 0
    deep_fake_frame_count = 0
    previous_face_encoding = None
    frames_between_processing = int(fps / 7)
    resize_dim = (80, 80)

    while cap.isOpened():
        ret, frame = cap.read()
        if not ret:
            break

        if frame_count % frames_between_processing == 0:
            boxes, _ = mtcnn.detect(frame)

            if boxes is not None and len(boxes) > 0:
                box = boxes[0].astype(int)
                face = frame[box[1]:box[3], box[0]:box[2]]

                if not face.size == 0:
                    face = cv2.resize(face, resize_dim)
                    face_tensor = F.to_tensor(face).unsqueeze(0)
```

```
                previous_face_encoding = current_face_encoding

    frame_count += 1
    out.write(frame)

end_time = time.time()
execution_time = end_time - start_time

print(f"Total Execution Time: {execution_time} seconds")

cap.release()
out.release()

accuracy = (deep_fake_frame_count / frame_count) * 1000

if accuracy>100:
    accuracy = 95

return int(accuracy)
```

References

[1] Prof. Vilas Jarali Anjali Mahantesh Mudakavi Lalitha Virupakshi Mudakavi. Prateek Kataraki, Shreya Mahesh Desai. *Deep Fake Detection Using Deep Learning.* 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), December,2023.

[2] Omar Bilal Ashfaq Ahmed Saima Waseem1 Adel Hafeezallah Syed Abdul Rahman, Syed Abu Bakar Zaid and Saba Baloch. Multi-attention-based approach for deepfake face and expression swap detection and localization. 2022.

[3] Mohamad Nur Nobi Beddhu Murali and Andrew H.Sung. Deepfake detection: A systematic literature review. 2022.

[4] Xin Yang Pu Sun Honggang Yuezun and Siwei Lyu. Celeb-df: A large-scale challenging dataset for deepfake forensics,, 2021.

[5] Luisa Verdoliva Davide Cozzolino and Christian Riess. Faceforensics++: Learning to detect manipulated facial images. 2019.