ABSTRACT

With the growing power of computers, deep learning has become so strong that it's now easy to make videos where you can't tell if they are real or fake, commonly known as deepfakes. These face-swapped videos could be used in situations like causing political confusion, faking terrorist attacks, creating revenge porn, or blackmailing people. In my project, I'm working on a deep learning method that can tell if a video is AI-generated or real. It focuses on detecting two types of deepfakes: replacement and reenactment. Basically, I'm using AI to fight AI. My system uses a Res-Next Convolutional Neural Network (CNN) to capture features from individual video frames. Then, these features are used to train a Long Short Term Memory (LSTM) Recurrent Neural Network (RNN) to classify if a video has been manipulated or not, meaning if it's a deepfake or a real video. To make the model work better in real-life situations, I've trained it on a large, balanced dataset combining Face-Forensics++, the Deepfake Detection Challenge, and Celeb-DF. My approach is simple but gives competitive results.

INRODUCTION

Our project focuses on detecting deepfake videos, which is a growing concern because they can be misused to create misleading and harmful content. Deepfakes are AI-generated videos where a person's face is swapped with someone else's, making it look like they are doing or saying things they never actually did. This can have serious consequences, especially in situations like political manipulation, blackmail, or spreading fake news.

In this project, we are developing an AI-based system that can distinguish between real videos and deepfakes. The system combines a ResNext Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) to detect small details that may be invisible to the human eye. While creating deepfakes has become relatively simple, detecting them requires a sophisticated analysis of video frames. Our model analyzes the features of each frame and examines changes between frames to identify manipulation.

We trained the system on a large dataset of real and deepfake videos, including data from FaceForensics++, the Deepfake Detection Challenge, and Celeb-DF. The ultimate goal is to build a tool that can classify videos as real or fake in real-time. This will allow users to upload videos for verification in an easy-to-use platform. Given the risk of misinformation spreading in today's digital world, this tool is critical. By using AI to counter AI-generated deepfakes, we are contributing to the fight against the malicious use of technology.

Objectives

- Uncover the hidden truth behind deepfakes.
- Help reduce the misuse of deepfakes and prevent people from being misled online.
- Identify and classify videos as either deepfake or real.
- Create a simple, user-friendly system where people can upload videos to check if they're genuine or fake.

Problem Statement

In recent years, we've seen some super convincing manipulations of digital images and videos thanks to visual effects. But now, with the rise of deep learning, things have gone to a whole new level. Fake content, especially what we call deepfakes, has become way more realistic and super easy to make with AI tools. The scary part is how difficult it is to tell when something's a deepfake. These have been used to stir up political drama, fake terrorist attacks, create revenge porn, and for blackmail. So, it's really important that we figure out how to detect and stop them from spreading all over social media. My project focuses on tackling this issue by using an LSTM-based artificial neural network to detect deepfakes.

Existing System

The current deepfake detection methods use a bunch of different techniques, but all of them come with their own set of issues. One popular method involves looking for face warping artifacts. Basically, it uses a Convolutional Neural Network (CNN) to compare the fake face to the area around it. This works because most deepfake algorithms still create images with low resolution, so they need some tweaking to match the video. But the problem is that this method doesn't look at how things change over time, making it less useful when the fake parts are subtle and consistent throughout the video.

Another method focuses on eye blinking, where the system checks if there's any blinking as a sign of deepfakes. This technique uses something called a Long-term Recurrent Convolution Network (LRCN) to analyze eye movements over time. But, with more advanced deepfake tech now generating realistic eye blinks, this method

isn’t as effective anymore.

Other methods like capsule networks try to detect deepfakes by looking for forged images and videos. The problem is they rely on random noise during training, which makes them less reliable when working with real-time data. Then there are systems that try to pick up biological signals from faces to catch deepfakes, but these methods struggle to keep things consistent and coherent over time.

Proposed System

For our project, we’re using a deep learning-based approach to detect deepfake videos more effectively. The system uses a mix of ResNext Convolutional Neural Networks (CNN) to pull detailed features from each frame and a Long Short-Term Memory (LSTM) based Recurrent Neural Network (RNN) to process the video over time, identifying whether it’s real or fake. ResNext helps break down the visual details of each frame, while LSTM checks the sequence of frames for any inconsistencies that often show up in deepfakes.

To make sure it works well in real-time, we’ve trained the system using a variety of datasets, like FaceForensic++, the Deepfake Detection Challenge (DFDC), and Celeb-DF, which all provide a huge collection of real and fake videos. By combining these datasets, we’re making sure our model can generalize better and perform well across different types of videos.

We’ve also made it easy to use: there’s an interface where users can upload videos, and the system will tell them if the video is real or fake, along with how confident it is. This project has the potential to scale up, with features like browser plugins or integration with platforms like WhatsApp or Facebook for real-time deepfake detection.

LITERATURE SURVEY

[1] Detection of Deepfake Videos through Eye Blinking

- **Author:** Yuezun Li, Pu Sun, Honggang Qi, Xin Yang, Siwei Lyu
- **Date Published:** June 2018
- **Reference:** "Exposing Deepfake Videos by Detecting Eye Blinking" (IEEE)

  - The authors present a method for detecting deepfakes based on the absence or unusual patterns of eye blinking.
  - The Long-term Recurrent Convolution Network (LRCN) is used to capture the temporal behavior of eye blinking in video frames.
  - However, deepfake algorithms have improved to the point where eye blinking alone is not a reliable detection method. Other factors, like facial details, need to be considered.

[2] Face Warping Artifacts for Deepfake Detection

- **Author:** Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Nießner
- **Date Published:** January 2019
- **Reference:** FaceForensics++: Learning to Detect Manipulated Facial Images (arXiv:1901.08971)

  - The approach uses Convolutional Neural Networks (CNNs) to detect face warping artifacts in deepfake videos.
  - The method relies on the observation that deepfake generation tools leave artifacts due to resolution limitations when replacing faces in video frames.
  - However, this method lacks temporal analysis, which is critical for tracking changes between frames.

[3] Biological Signal-Based Detection for Deepfake Videos

- **Author:** Younghyun Park, Mohamed Elgharib, Christian Theobalt
- **Date Published:** September 2020
- **Reference:** "Detecting Deepfake Videos Using Biological Signals"

  - This study extracts biological signals like pulse from facial regions in deepfake videos to detect inconsistencies.
  - The authors developed a probabilistic Support Vector Machine (SVM) and Convolutional Neural Network (CNN) to classify videos based on photoplethysmography (PPG) signals.
  - Although this method showed high accuracy, the challenge lies in formulating a loss function that properly captures biological signals.

[4] Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers

- **Author**: Vrizlynn L. L. Thing
- **Date Published**: April 2023
- **Reference**: Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers (arXiv:2304.03698)

  - The study compares the effectiveness of *Convolutional Neural Networks (CNNs)* and *Transformers* in detecting deepfakes.
  - Experiments were conducted on second and third-generation deepfake datasets, achieving high accuracy rates, with AUC values as high as 99.99% in some cases.
  - The paper highlights the unique strengths of CNNs and Transformers in processing different types of deepfake data.

[5] Spotting Deepfakes with Semantic Analysis and Watermarking Techniques

- **Author**: Rahul Vishwakarma (IEEE Senior Member)
- **Date Published**: 2024

- **Reference**: Spotting Deepfakes with Semantic Analysis and Watermarking Techniques (IEEE Transmitter)

    - This research focuses on *semantic analysis*, examining inconsistencies in object relationships within video frames, rather than relying solely on pixel-level differences.
    - The study introduces *digital watermarking* as a method for marking AI-generated images to help identify deepfakes.
    - It also addresses biases in current training datasets, which are predominantly composed of white male subjects, and how this may affect detection accuracy.

SOFTWARE REQUIREMENTS:

Operating System: Windows 7+
Programming Language: Python
Framework: Django
Cloud Platform: Google Cloud Platform
Libraries: OpenCV, Face-recognition

HARDWARE REQUIREMNETS:

Processor: Intel Xeon E5 2637 3.5 GHz
RAM: 8 GB
Hard Disk: 100 GB
Graphics Card: NVIDIA GeForce 1050 (4 GB RAM)

REFERENCES

[1] IEEE Paper - Y. Li, P. Sun, H. Qi, X. Yang, and S. Lyu, "Exposing Deepfake Videos by Detecting Eye Blinking," in IEEE Conference on Computer Vision and Pattern Recognition, *June 2018.*

[2] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in arXiv:1901.08971, *Jan. 2019.*

[3] Y. Park, M. Elgharib, and C. Theobalt, "Detecting Deepfake Videos Using Biological Signals," in IEEE Transactions on Information Forensics and Security, *Sept. 2020.*

[4] V. L. L. Thing, "Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers," in arXiv:2304.03698, *Apr. 2023.*

[5] IEEE Paper - R. Vishwakarma, "Spotting Deepfakes with Semantic Analysis and Watermarking Techniques," in IEEE Transmitter, *2024.*

Signature of Guide
Ms.Darshini Y
Assistant Professor

Signature of Coordinator
Dr.Pavithra A C
Associate Professor