DETECTION ENHANCEMENT FOR VARIOUS DEEPFAKE TYPES BASED ON RESIDUAL NOISE AND MANIPULATIONTRACES

DR. M. NAVANEETHA KRISHNAN, M.E., Ph.D., Professor,

Department of Computer Science and Engineering

Ms. MARY FLAVIA L, **M.E** Student of Computer Science and Engineering

St. Joseph College of Engineering, Sriperumbudur, Chennai.

Abstract:

As deepfake techniques become more sophisticated, the demand for fake facial image detection continues to increase. Various deepfake detection techniques have been introduced but detecting all types of deepfake images with a single model remains challenging. We propose a technique for detecting various types of deepfake images using three common traces generated by deepfakes: residual noise, warping artifacts, and blur effects. We adopted a network designed for steganalysis to detect pixel-wise residual-noise traces. We also consider landmarks, which are the primary parts of the face where unnatural deformations often occur in deepfake images, to capture high-level features. Finally, because the effect of a deepfake is similar to that of blurring, we apply features from various image quality measurement tools that can capture traces of blurring. The results demonstrate that each detection strategy is efficient, and that the performance of the proposed network is stable and superior to that of existing detection networks on datasets of various deepfake types.

Introduction:

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. Until recently, the number of fake videos and their degrees of realism has been limited by the lack of sophisticated editing tools, the high demand on domain expertise, and the complex and time-consuming process involved. However, the time of fabrication and manipulation of videos has decreased significantly in recent years, thanks to the accessibility to large-volume training data and high-throughput computing power, but more to the growth of machine learning and computer vision techniques that eliminate the need for manual editing steps.

In particular, a new vein of AI-based fake video generation methods known as Deep Fake has attracted a lot of attention recently. It takes as input a video of a specific individual ('target'), and outputs another video with the target's faces replaced with those of another individual ('source'). The backbone of deep fake is deep neural networks trained on face images to automatically map the facial expressions of the source to the target. With proper post-processing, the resulting videos can achieve a high level of realism.

Objectives:

We propose a technique for detecting various types of deep fake images using three common traces generated by deep fakes: residual noise, warping artifacts, and blur effects. We adopted a network designed for steganalysis to detect pixel-wise residual-noise traces. We also consider landmarks, which are the primary parts of the face where unnatural deformations often occur in deep fake images, to capture high-level features. Finally, we apply features from various image quality measurement tools that can capture traces of blurring.

Literature Survey:

Deep Fake Detection for Human Face Images and Videos: A Survey.

AUTHORS: Asad Malik, Minoru Kuribayashi, Sani M. Abdullahi, Ahmad Neyaz Khan. **YEAR:** 2021.

Deep Fake is a generative deep learning algorithm that creates or modifies face features in a super realistic form, in which it is difficult to distinguish between real and fake features. This technology has greatly advanced and promotes a wide range of applications in TV channels, video game industries, and cinema, such as improving visual effects in movies, as well as a variety of criminal activities, such as misinformation generation by mimicking famous people.

A Large-scale Challenging Dataset for Deep Fake Forensics.

AUTHORS: Yuezun Li1, Xin Yang, Pu Sun, Honggang Qi and Siwei Lyu. **YEAR**: 2021.

The need to develop and evaluate Deep Fake detection algorithms calls for large-scale datasets. We present a large new scale challenging Deep Fake video dataset, Celeb DF, which contains 5, 639 high-quality Deep Fake videos of celebrities generated using improved synthesis process.

Spatial-Phase Shallow Learning: Rethinking Face Forgery Detection in Frequency Domain.

AUTHORS: Honggu Liu Xiaodan Li, Wenbo Zhou Yuefeng, Chen Yuan He Hui Xue, Weiming Zhang Nenghai Yu**. YEAR:** 2021.

According to the property of natural images, the phase spectrum preserves abundant frequency components that provide extra information and complement the loss of the amplitude spectrum.

To this end, we present a novel Spatial-Phase Shallow Learning (SPSL) method, which combines spatial image and phase spectrum to capture the up-sampling artifacts of face forgery to improve the transferability, for face forgery detection.

Face X-ray for More General Face Forgery Detection.

AUTHORS: Lingzhi Li1 Jianmin Bao Ting Zhang Hao Yang Dong Chen Fang Wen Baining Guo. **YEAR:** 2021.

We observe that most existing face manipulation methods share a common step: blending the altered face into an existing background image. For this reason, face X-ray provides an effective way for detecting forgery generated by most existing face manipulation algorithms.

Face X-ray is general in the sense that it only assumes the existence of a blending step and does not rely on any knowledge of the artifacts associated with a specific face manipulation technique.

Fighting Deepfake by Exposing the Convolutional Traces on Images.

AUTHORS: Luca Guarnera, Oliver Giudice, And Sebastiano Battiato. **YEAR:** 2021.

In this paper, a new approach aimed to extract a Deepfake fingerprint from images is proposed. Finally, tests carried out on Deepfakes generated by FACEAPP achieving 93% accuracy in the fake detection task, demonstrated the effectiveness of the proposed technique on a real-case scenario. The results are utterly realistic and extremely easy to exploit even for non-experienced users. This kind of media object took the name of Deepfake and raised a new challenge in the multimedia forensics field.

System Design:

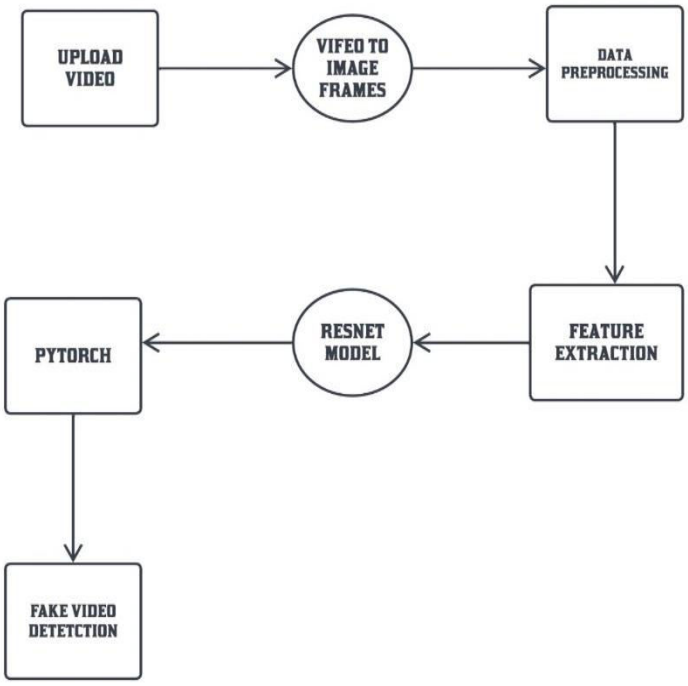
In this proposed application consists following modules,

- Dataset
- Data Preprocessing
- Feature Extraction

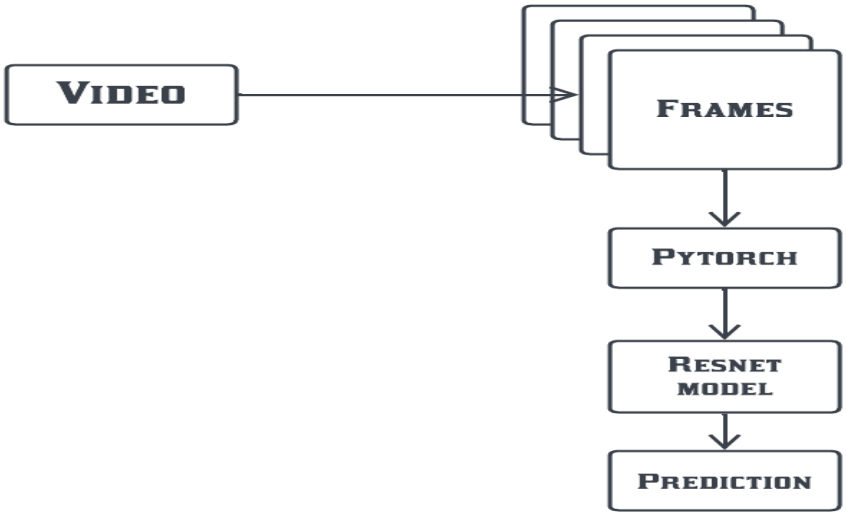
- Experimental Setup
- Evaluation Metrics

DATA FLOW DIAGRAM:

Data flow diagram is made up of a few symbols, which represents system components. Most data flow modelling methods use four kinds ofsymbols. These symbols are used to represent four kinds of system components. Processes, data stores, data flows and external entities. Processes are representedby circles in DFD. A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system. It differs from the flowchart as it shows the data flow instead of the control flow of the program. A data flow diagram can also be used for the visualization of data processing. The DFD is designed to show how a system is divided into smaller portions and to highlight theflow of data between those parts. Data Flow Diagram (DFD) is an important technique for modelling a system's high-level detail by showing how input data is transformed to output results through a sequence of functional transformations. DFDs reveal relationships among and between the various components in a program or system. DFD consists of four major components: entities, processes, data stores and data flow. Data flows are represented by a thin line in the DFD.



ARCHITECTURE DIAGRAM:

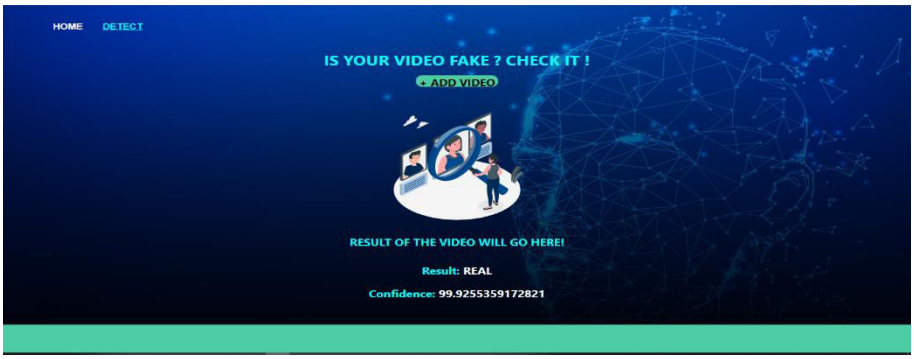


Implementation:

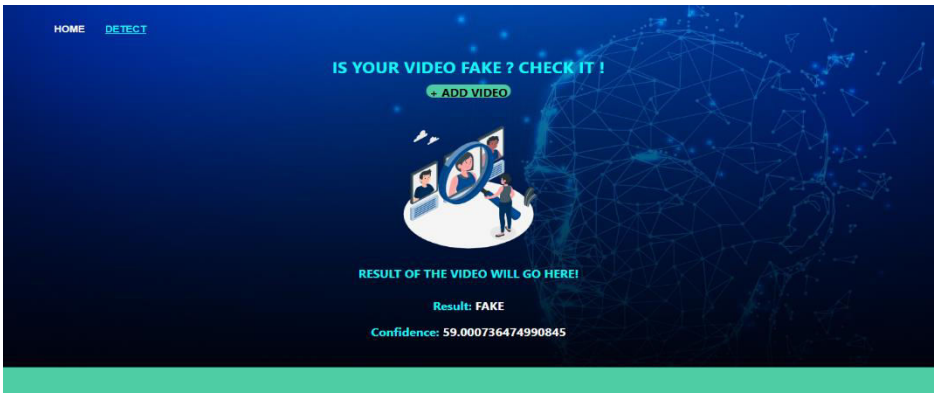
The first step of the system is to input a video. The second phase of the system deals with quality enhancement of the input signals of the video. The third step is Feature extraction involves the analysis of the video. It is considered as an important phase of the system as extraction of relevant and significant features heavily impact on the final recognition.

Here we used (CNN) convolutional Neural Network algorithm, it is used to do image classification and image recognition in neural networks. It is a deep learning algorithm. The convolutional layer preserves the relationship between pixels. The system sees an array of pixels and depends on the resolution of the Frames. It is the main step of the system in which the frames are filtered into Pixels based on the features extracted from the video using Convolutional layer, Pooling Layer, and Fully connected Layer. For example, it will be downscaling the 224*224 pixels to 112*112 pixels and so on till it detects output layer. Another one is Generative Adversarial Network (GAN). A Generative Adversarial Network (GAN) is a deep learning architecture that consists of two neural networks competing against each other in a zero-sum game framework. In GANs, there is a Generator and a Discriminator. The Generator generates fake samples of data (be it an image, videos etc.) and tries to fool the Discriminator. The Discriminator, on the other hand, tries to distinguish between the real and fake sample videos. The Generator and the Discriminator are both Neural Networks and they both run in competition with each other in the training phase. The steps are repeated several times and in this, the Generator and Discriminator get better form of real noises and see if it can correctly predict them as fake and result will be displayed.

Results After Face Detection:



RESULT SCREEN – REAL



RESULT SCREEN - FAKE

Conclusion:

This project is used as a deepfake creation, and detection methods. Deepfake creates forged images or videos that persons cannot differentiate from real images or videos. Deepfakes are created using generative adversarial networks, in which two machine learning models exit. One model train on a dataset and the other model tries to detect the deepfakes. The forger creates fakes until the other model can't detect the forgery. Deepfakes creating fake news, videos, images, and terrorism events that can cause social and financial fraud It is increasing affects religions, organizations, individuals and communities', culture, security, and democracy. When deepfake videos and images increase on social media people will ignore to trust the truth.

Future Enhancement:

- So, deepfake datasets and cross-platform detection techniques need to be developed in the future.
- This needs efficient, reliable, and robust mobile detectors to detect deepfakes in widely used mobile devices.
- Moreover, will improve deepfake detection by integrating deepfake detection and object detection algorithms.

Author 1:

Dr. M. NAVANEETHA KRISHNAN, M.E., Ph.D., Professor,
Department of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu.

Author 2:

Ms. MARY FLAVIA L, M. E student of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu.