Fake Video Detection Using CNN- Deep Learning Model

Joyel Albert M
Dept of Computer Science of Enigineering
Knowledge Institute of Technology
Salem,India
mjoelalbert5@gmail.com

Suveetha S
Dept of Computer Science of Engineering
Knowledege Institute of Technology
Salem,India
ssucse@kiot.ac.in

Gowtham S
Dept of Computer Science of Engineering
Knowlegde Institute of Technology
Salem,India
gowthamofficial077@gmail.com

Dhanush A
Dept of Computer Science of Engineering
Knowlegde Institute of Technology
Salem,India
sabaridhanush2004@gmail.com

Dhanuskaarthikeyan S
Dept of Computer Science of Engineering
Knowledge Institute of Technology
Salem,India
dhanushkaarthikeyan27@gmail.com

Abstract— ***The invention of sophisticated video generating technology has facilitated the production of fake videos to become more realistic and achievable with less effort than before. These artificial videos commonly referred to as deepfakes are a major threat to online communication by feeding false information and loss of faith in online media. Consequently, the task of identifying such fabricated material has become a significant study issue in multimedia security and computer vision. This paper presents a counterfeit video detection method, which is based on detecting visual anomalies in facial areas in video frames. The process starts with spilt the video frames as frame by frame to analysis the face details.fake videos would has some small mistakes or unnatural patterns would be there in the video.the process will frame to frame time check which means it would continuous frame check as one frame at a time.This would be done using Convolutional Neural Network(CNN) a deep learning model which is used for image pattern detection.***

Keywords-: Fake Video Detection, Computer Vision, Multimedia Security, Convolutional neural network (CNN), Deepfake Detection, Deep Learning.

Introduction

The videos are the important and familiar way to communication and entertainment etc. The fake videos can be generated using Artificial Intelligence Techniques. The Ai will analysis and learn the face details and it will help generate the fake video.The generate video will looks like a real video.It would be difficult to find the video is ai generated.The Ai video will be generate using some algorithms and machine learning modelse of available. deepfake technology has been used positively in the filmmaking industry, virtual reality, and digital entertainment, it is also a significant threat.Fake videos can affect the person's life by spreading misinformation . It can create political issues among the parties.

Deep fake videos can destroy the respect and repetution of the person.So Ai videos can affect and create direct mental health to the person.Ai video wih false information spreaded it can be easily believed by the public.now a days,Ai video creating technologies are easy to access and eassy to use so anyone can generate any video with the help of models and algorithms.

The videos are one of the main communication medium.but we can use this feature in good manner for education ,entertainment purposes Ai with Deep learning is used to create fake videos.Model will learn the face deatials and then it will replace the face in the in the generating video. fakecan have a negative impact based on misinformation, identity theft, political interference, and computer crimes.

The edited fake videos can convey misinformation about a person and spoil the person's name. There are n number of fake video tools to generate the fake video and it is difficult to identify either video is ai edited or original.The detecting video is ai or original is difficult.

Current available models are used specific dataset so it give result with good accuracy but when the new dataset given it result may be appropriate and accuracy will be bad.The result will be not good because the video quality change.the video compression.The lighting in the video would be different.The facial expressions also would be different.

So these are the factors that affect the accuracy of the model.As a solution we have to design strong and generalization model by using Deep learning.Deep learning is used to detect and manipulate to identify the pattern.We can create model using Advanced Deep Learning techniques.The Deep Learning models are used to find the video was ai generated or original as efficiently as possible

This study aims to help in making more trustable and scalable deepfake detect model which can succeed in detecting manipulate context in a real-time situation.The fake video

detection models are used to ensure the security and privacy of a person is ensured and it would protect the respect and the reputation of a person.

I. LIRTRETURE SURVEY

This paper suggests a deep fake detection algorithm that relies on dual attention convolutional network to enhance the localization of manipulated facial areas. The model combines Spatial Reduction Attention Block (SRAB) and Forgery Feature Attention Module (FFAM) to improve the accuracy of detection. FaceForensics++ and DFDC experiments demonstrate better performance in the context of AUC than using traditional CNN models, but there is still a problem with generalization of the results to unseen manipulations. [1].

An efficacious deepfake detection model that is based on EfficientNet-B0 and Temporal Convolutional Neural Network is suggested to examine spatial and temporal anomalies. The only methods of face alignment and augmentation through MTCNN to enhance robustness are the CutMix and MixUp used in the system. Although the detection performance is good, the computational complexity is a source of concern when used in large scale. [2].

The multi-stream deepfake detection network is suggested to learn spatial, temporal, and structural abnormalities at the same time. This model takes three ResNet-50 encoders of RGB frames, optical flow and edge features and then fuses decision based on fuzzy inference. Experiments on FaceForensics++ and Celeb-DF yield better known faces recognition, but there is still poor generalization to novel manipulation algorithms. [3].

Other methods use a Conv-LSTM hybrid network which processes video frame sequence in order to discover finer facial movement akin to blinking and lip-reading. It is an image-based approach that integrates convolutional feature learning with temporal sequence learning to enhance classification of fake videos, and it needs a large number of computational resources when dealing with large datasets. [4].

A capsule network based deception detector model combines both facial expression and heart rate signals derived out of video frames. To improve the feature representation and the classification, channel-wise attention and supervised contrastive learning are used. Nonetheless, the fact that large real-world datasets are not available is a drawback. [5].

An evidence-based forensic-driven deepfake detector is suggested to be used in court prosecution. The system is a combination of frame selection, confidence threshold, timestamps, and heatmap visualization to help forensic analysts ascertain the authenticity of video. Although it can be interpreted, the system is constantly being updated because of the rapid development of deepfake generation techniques. [6].

A compression-aware hybrid detection model is presented, which can be used with low quality or other highly compressed videos. The methodology incorporates wavelet-based frequency characteristics together with Conv3D spatiotemporal characteristics and a lean ResNet classifier. The experimental findings reveal enhanced strength at varying levels of compression. [7].

The other hybrid system is a system that integrates two features, the features of the facial appearance and the features of upper-body motions to detect irregularities in the manipulated videos.The model has the VGG16 and MoveNet motion features, Conv1D and Transformer layers to learn the time-dependent features.This multimodal technique enhances the performance in detection of advanced deepfake methods. [8].

The wide survey touches upon the deepfake generation and detection techniques, which are presented using CNN, RNN, GAN, and autoencoders.The article categorizes the deepfake manipulations into identity swap expression transfer, attribute editing and full-face synthesis and the benchmark datasets. It points out the fact that most detection models are not able to generalize to unknown manipulations. [9].

The other survey looks at the generation and detection of image, video and audio modalities by utilizing machine learning and deep learning algorithms of deepfake. It compares methods of GAN generation, multimodal detection approaches, and benchmark datasets and highlights the requirement of scalable and supported frameworks of detection in real-life contexts. [10].

II. EXISTING SYSTEM

The existing deepfake detectors mostly use deep learning algorithm in order to detect spoofed face image and videos. Most of the conventional methods involve the use of a convolutional Neural Networks (CNN) in order to detect visual anomalies, including abnormal textures, lighting changes and edges of compatibility created throughout the face manipulation process. Time-related information of every video frames is being used in other methods to detect the unnatural facial movement, blinking pattern of the eyes and lip-synch problems. They are usually trained using large level of data contained databases consists of real and spoofed videos.

The ability of these systems to detect may be a reasonable one when used in controlled environments, but it would be compromised as more deepfake methods are created or new data sets are discovered. and certain of the having models are very tough to calculate in addition to the reason that their centralized training data can be very large to an extent that scalability, processes-time and data-management problem may arise.

The deepfake video is developed used ai technology.The video frame would be selected and as frame by frame and it would be processed to find either the videos was original or fake video.This would be done using some techniques and algorithms to classify the video find the truth and avoid conflicts of spreading misleading information spread among the social media and in the internet by misleading persons.

The fake video would be become thread to make a person to feel uncomfortable and it would be lead to make crimes and insecure among the peoples so we can avoid this by building CNN model to detect deep fake videos.

III. SYSTEM OVERVIEW

The suggested system will identify manipulated or fake videos with the help of a deep learning-based detection system. The system takes video input and extracts the features on the face in each frame in order to ascertain whether the content is genuine or it is an artificially generated content.

The primary goal of the system is to detect visual inconsistencies that have been inserted in the process of creating deepfakes and label the video as either real or fake.The general system has a number of stages such as video frame extraction, face detection, preprocessing, feature extraction, and classification.

The first step involves processing of videos to first capture the input video and then convert it into single frames through video processing. The areas of faces are detected with the help of a previously-trained face detection model on each frame. The purpose of this step is to make sure that only the potential .

The extracted facial images are then preprocessed by using resizing and normalization of the images to maintain consistent input to the dl model. All the processed images are then insert into a CNN which learns spatial attribute and pattern that compared the real time and altered faces.

the trained model classifies binary and gives an estimate on how original or a fake video frames was the video frame that was being analysed. The framework combine frame-based predictions to assist the accuracy of the whole video and displays the results and the scores of the confidence. This strategy will contribute to raising the accuracy of identifying a fake media in the online context.

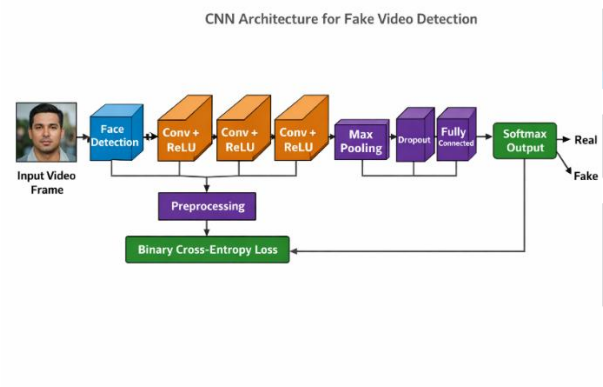

Fig 1.1 CNN Architecture Fake Video Detection.

IV. PROPOSED SYSTEM

The system would be used to identify visual contents that are manipulate by a binary classified model through deep learning and a trained face localization model. The system processes the input video frames or images and finds the regions of faces and analysis them with the help of the cnn, and the content is detect as a genuine one or a manipulated one. Preprocessing is taken to ensure that the performance remains constant regardless of variations in input size and illumination through preprocessing operations such as resizing, normalization and spatial alignment.

A. Deep CNN Model Architecture.

Its fundamental classification element is a progressive Convolutional Neural Network (CNN) that has been designed to derive discriminative features of facial Regions of Interest (ROI). Convolutional layers of (3 times 3) filters are used in order to extract spatial patterns in the input information. The operation of convolution may be mathematically described as:

$$(f * g)(n) = \Sigma_{m=-M}^{M} f(m)\, g(n-m)$$

the f is the input image and g is the convolution kernel.

Each convolutional layer will have a Rectified Linear Unit (ReLU) activation function that provides non-linearity of the network and increases the learning capacity. The ReLU function is defined as:

$$f(x) = \max(0, x)$$

In order to minimize space and calculation costs MaxPooling layers of size (2 2) are used. These layers retain the most important details and reduce the unnecessary information. Furthermore, there are Dropout layers, which are introduced during the training to combat overfitting by chance using a sample of neurons to be inactive in each training step.Once the features are extracted, there are flatteners and fully connected layers which take the resulting feature maps. The last output layer uses Softmax activation function to generate probability scores of how likely each of the classes is

B. Mathematical Model and Optimization.

The Binary Cross-Entropy Loss Function is used to point out the training process of the proposed model as it measures the difference between the predicted outputs and the true labels. The loss functional can be defined as:

$$L(y, \hat{y}) = -(1/N) \Sigma_{i=1}^{N} [\, y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \,]$$

Where (N) is the number of training samples, The ground truth label is presented (yi), and (ŷi) is the predicted probability of the (i th) sample.

Adam optimizer is used in order to optimize the model parameters. Adam varies the learning rate of different parameters depending on the breakdown of first and second moment estimates of gradients. The rule of parameter update will be written as:

$$\theta_{t+1} = \theta_t - \eta\, \hat{m}_t / (\sqrt{\hat{v}_t} + \varepsilon)$$

and where (θ) is the model parameters, (η) is the learning rate, (m̂t), and (v̂t) are the bias-corrected moment estimates.

C. Face Detection using SSD

The system detects faces before the classification step by a Single Shot Detector (SSD) platform on the basis of a ResNet-10 backbone. The SSD model is unlike the traditional sliding window methods that would need to do multiple forward passes through the image to predict bounding boxes and confidence scores.The face detecting model (res10300x300ssditer140000.caffemodel) also downsizes the input frames to (300 (w) x 300 (h) pixels), which allows the face parts to be localized correctly and only then sent to the classification model.

D. System Flow Overview

The general workflow of the system is as follows: we capture video frames or images with the help of OpenCV. The SSD-based detector removes facial bounding boxes out of the

input frames. The identified areas are then preprocessed with the methods of resizing and converting them into grayscale. The CNN-based classifier is fed with the processed facial images, which are binarily classified and the authenticity of the content is predicted. Lastly, the system shows the results of the prediction, in form of bounding boxes and labels of the identified faces, allowing real-time visualization of the prediction output

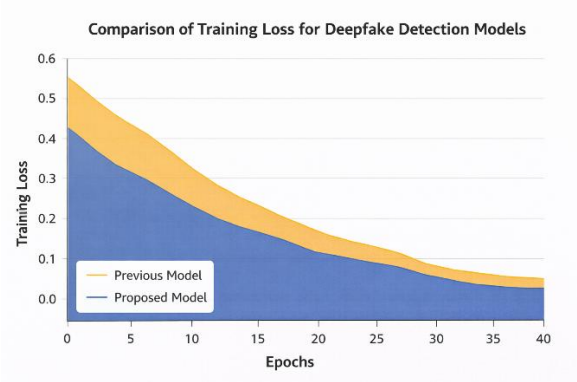

Fig 1.2:Comparison of Training Loss for Deepfake Detection Model.

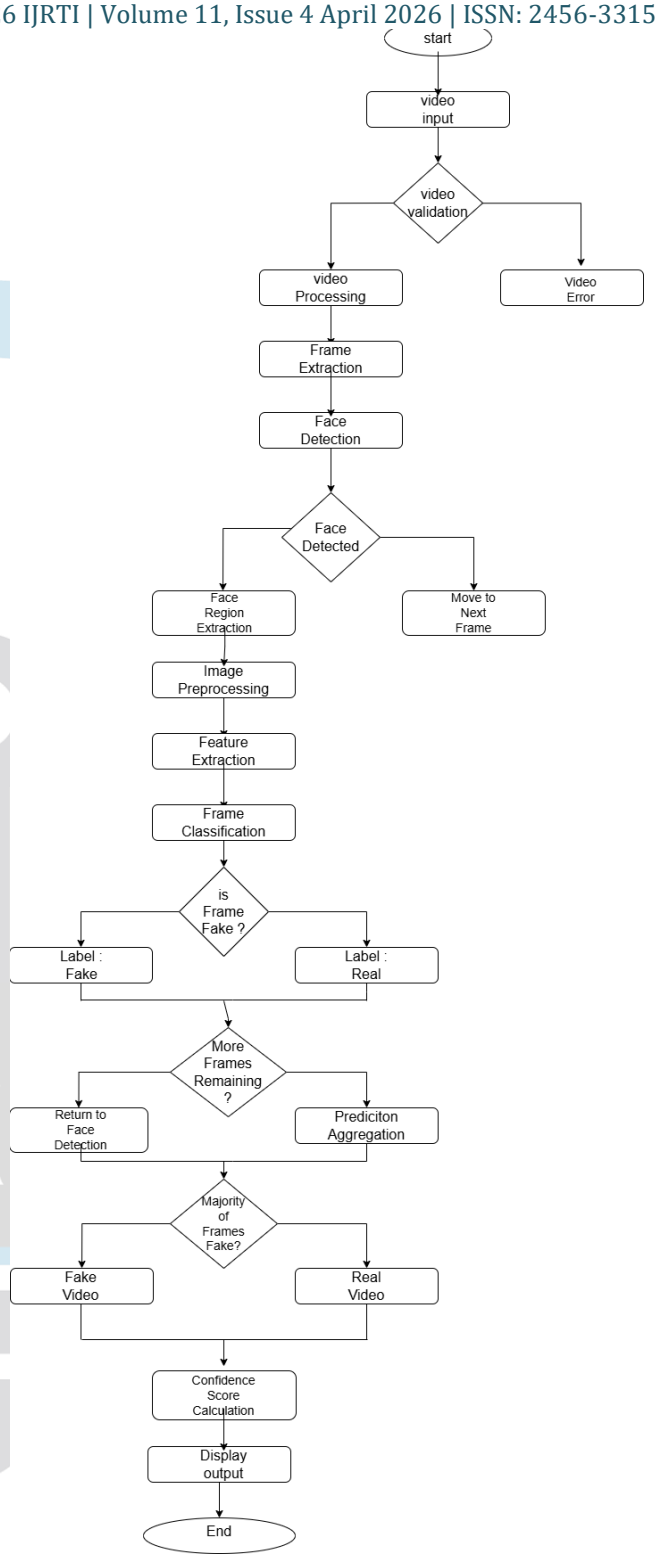

Fig 1.3:System Flow Architecture

RESULTS

The given fake video detection system based on deep learning was tested on a dataset of both samples of real and manipulated videos. The trained CNN model is used to perform the task of classify the video frames, identify the position of the facial part and conduct the classify. The model was trained and could identify visual artifacts and anomalies that were brought about by the technique of deepfake. Based on the findingsof the experiment, the model has been found to be reliable in both detection and classification accuracy. The combination of frame-level predictions was made to derive the final classified of the video. In most cases, the system was able

to detect fair and fake content. the model provides the measure of confidence of all the predictions, this is used to determine the reliability of the detection output. The results confirm the efficiency of the proposed system to perform a successful facial analysis and manipulated video detection, and, that is why, it can be efficiently used in the sphere of digital media verification and controlling of online material.

Conclusion

These research has provided and used a deep learning of a fake video detector. It is a combination of face detection and the classification based on convolutional neural network to analyze the facial features of the video frames. The model has the ability of identifying the real and fake content through the analysis of spatial patterns and irregularities that are inserted during the production of the deepfake.

The experimental analysis shows that the suggested approach offers credible detection characteristics and can be used to the real-life situations where the verification of the authenticity of the digital media is value. There is also the system of confidence-based predictions where the system helps the user to make knowledgeable decisions about the quality of video.To achieve more work in the future, the model may be trained on larger and more diverse datasets, learn temporal representations with one or more successive frames and utilize more complex deep learning architecture to improve detection and resilience

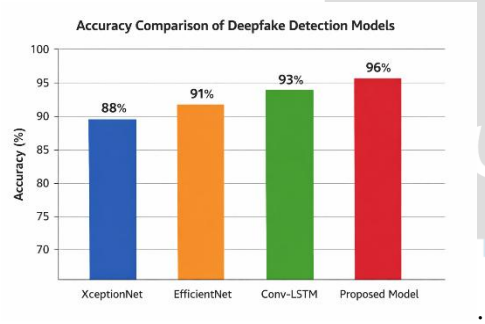

Fig 1.4: Accuracy Comparison of Deepfake Detection Models.

REFERENCES

[1] S. Bhuvaneshwari, M. R. Kishore, and P. Vignesh, "Dual Attention Network Approaches to Face Forgery Video Detection," in *Proceedings of the International Conference on Artificial Intelligence and Data Engineering*, 2024.

[2] A. Sharma, R. Kumar, and S. Gupta, "Hybrid Deep Learning Framework for Deepfake Detection Using Temporal and Spatial Features," *Journal of Artificial Intelligence and Data Science*, vol. 12, no. 3, pp. 45–56, 2024.

[3] M. Hassan, A. Ahmed, and T. Khan, "Deepfake Detection Using Spatio-Temporal-Structural Anomaly Learning and Fuzzy System-Based Decision Fusion," *IEEE Access*, vol. 11, pp. 102345–102356, 2023.

[4] K. Prakash and S. Balasubramanian, "An Exploratory Analysis on Visual Counterfeits Using Conv-LSTM Hybrid Architecture," *International Journal of Computer Vision and Pattern Recognition*, vol. 9, no. 2, pp. 78–88, 2023.

[5] Y. Zhang, L. Wang, and H. Liu, "Video-Based Deception Detection via Capsule Network With Channel-Wise Attention and Supervised Contrastive Learning," *IEEE Transactions on Multimedia*, vol. 25, pp. 1550–1562, 2023.

[6] J. Peterson and M. Clark, "Empirical Assessment of Deepfake Detection: Advancing Judicial Evidence Verification Through Artificial Intelligence," *Digital Forensics Journal*, vol. 18, no. 1, pp. 22–34, 2024.

[7] R. Mehta and D. Patel, "Compression-Aware Hybrid Framework for Deep Fake Detection in Low-Quality Video," *IEEE Access*, vol. 11, pp. 84521–84532, 2023.

[8] K. P. Rahatwal, S. Pundir, M. Wazid, and V. Bhat K., "A Novel Approach to Deepfake Detection: Leveraging Fused Facial and Body Dynamics With a CNN–Transformer Hybrid Network," *IEEE Access*, vol. 13, pp. 197085–197097, 2025.

[9] S. Agarwal, T. El-Gamal, and H. Farid, "DeepFake Detection for Human Face Images and Videos: A Survey," *ACM Computing Surveys*, vol. 55, no. 4, pp. 1–36, 2023.

[10] Y. Mirsky and W. Lee, "Deepfake Generation and Detection: A Survey and Case Study," *ACM Computing Surveys*, vol. 53, no. 4, pp. 1–36, 2021.