Deep-Fake Detection

Sakthivel E [1], Dr. SPAnandaraj [2], Amreen Khanum [3], Dr. Sukruth Gowda [4]
[1]Syed Asad, [2]Syed Abdul Azeem, [3]Abdul Rishad kp, [4]Shaik Subhan Basha and [5]A S Mohammad Shareef
Department of Computer Science and Engineering
Presidency University, Bangalore.

***Abstract*—** The world of digital media has been changed by the development of deep learning techniques, especially with the rise of deepfakes. These computer-generated videos, which combine existing footage or alter it to show different content, have caused worries in various areas such as politics, entertainment, and personal privacy. This review combines information from many research studies to provide a clear overview of the latest techniques and findings in detecting deepfake videos. It explains how deepfakes have evolved, the problems they create, and the new solutions proposed by researchers around the world.

Keywords: Generative Adversarial Networks (GANs),Artificial intelligence (AI),Long Short-Term Memory (LSTM),Convolutional Neural Networks (CNNs),ResNext

I. INTRODUCTION

Deepfakes are computer-generated videos that use artificial intelligence to manipulate footage and make it appear real. They have become a major problem in digital media, as bad actors can create convincing fake videos to spread misinformation. This has led to a lot of research into effective ways to detect deepfakes.Deepfakes are primarily made using a technique called Generative Adversarial Networks (GANs). As they have become more advanced, it has become harder for traditional video verification methods to spot them. Undetected deepfakes can have serious consequences for media, politics, and people's privacy, so ongoing research in this area is very important. With the growth of digital platforms and AI tools becoming more widely available, it is now easier for anyone to create and share deepfakes. These manipulated videos can look almost real to most people, making them a powerful tool for spreading false information and propaganda. It is crucial for technology experts and researchers to develop reliable ways to detect deepfakes in order to combat their negative impacts.

1.1 Methods of Creating Deepfakes

There are many sophisticated techniques for creating deepfake content in digital media:

Identity Swap: Replaces one person's face in an image or video with another person's face, often using Convolutional Neural Networks (CNNs).

Face Reenactment: Substitutes the facial expressions in an image or video of one person with those of another, commonly called expression swapping. Real-time face reenactment techniques are frequently used for this.

Attribute Manipulation: Modifies facial attributes like hair or skin color, gender, age, or adds accessories like glasses. Methods based on Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and 3D face models are employed for attribute manipulation.

Entire Face Synthesis: Generates entirely new face images, using approaches such as GANs and U-Net.

Voice Cloning: Replicates a person's voice, allowing for the creation of audio content that mimics someone else's speech patterns and tone, using techniques like text-to-speech (TTS) models and voice synthesis systems.

Full-Body Deepfakes: Extend beyond facial manipulation and replace the entire body of a person in an image or video, often employing pose estimation and GANs.

Object Manipulation: Alters or adds objects within an image or video, including inserting or removing items, changing their appearance, or manipulating their positions. Techniques like image segmentation and object detection are used.

Lip-Sync Deepfakes: Focus on accurately synchronizing the lip movements and speech of a target person in a video, achieved by mapping the phonetic content of the audio to the lip movements of the target using deep learning methods.

Background Replacement: Replaces the background of an image or video with a different setting or scene, using techniques like image segmentation, background removal, and compositing.

II. LITERATURE REVIEW

Deepfake technology, powered by Generative Adversarial Networks (GANs) and artificial intelligence (AI), has become a powerful tool for manipulating media, particularly in altering facial expressions and lip-sync effects. While creating deepfakes has become more accessible, detecting them remains challenging. Current detection tools often lag behind their creation counterparts in accessibility and simplicity.

Some researchers have proposed efficient detection methods:

Ahmed et al. leverage the distinctive blur and noise at the edges of imposed faces, using the Laplacian operator for edge detection.

Waseem et al. introduce an attention-based multi-task approach that enhances feature maps for deepfake classification and localization tasks, demonstrating competitive performance across diverse datasets.

Pathade et al. emphasize the societal impact of deepfakes and the importance of countermeasures such as legislation, regulation, corporate policies, education, and technology development.

Previous research has explored various aspects of deepfake detection:

Nguyen et al. provide a comprehensive survey of deepfake creation algorithms and detection methods, underscoring the necessity of technologies to automatically detect and assess the integrity of digital visual media.

Shet et al. introduce an approach for Deepfake detection utilizing ResNext and Long Short-Term Memory (LSTM), achieving 91% accuracy.

Yuezun Li et al. address the threat posed by open-source deepfake creation tools and present the DeepFake-o-meter platform for user-friendly deepfake detection.

Rafique et al. propose an automated method employing Deep Learning and Machine Learning methodologies, reaching 89.5% accuracy by combining Error Level Analysis, CNN, Support Vector Machines, and K-Nearest Neighbors.

Wodajo et al. present a Convolutional Vision Transformer for Deepfake detection, achieving 91.5% accuracy on the DeepFake Detection Challenge Dataset (DFDC).

Shad et al. focus on the significance of distinguishing between real and deepfake content, utilizing eye blinking as a detection feature to attain high accuracy rates.

Sanil M et al. offer a novel deep learning strategy, incorporating ResNext CNN and LSTM, for the identification of fraudulent videos, demonstrating competitive performance on various datasets.

Murali et al. conduct a systematic literature review, categorizing Deepfake detection methodologies into deep learning-based, classical machine learning-based, statistical, and blockchain-based techniques, with deep learning-based methods outperforming others.

These comprehensive summaries encompass the diverse landscape of research and technological advancements in the domain of deepfake detection, which is crucial in safeguarding the integrity of digital visual media amidst the proliferation of deepfakes.

III. VARIOUS APPROACHES

Deepfake technology uses advanced deep learning models to create fake media, especially videos that manipulate facial expressions and lip movements. While making deepfakes has become easier, detecting them is still quite challenging. Here are some key techniques used for deepfake detection:

Techniques for Deepfake Detection

Facial Landmark Analysis: This technique looks at specific points on a face to find inconsistencies that may indicate manipulation.

Lighting Inconsistencies: Detecting differences in lighting can reveal deepfakes, as they often do not match the natural lighting of the scene.

Audio-Video Mismatches: This method checks for any discrepancies between the audio and the video, which can be a sign of tampering.

Advanced Neural Networks: These networks are trained to identify subtle patterns that may not be obvious to the human eye.

Artifacts Detection: Many detection methods focus on artifacts—small errors or irregularities introduced during the deepfake creation process.

Temporal Consistency: This approach examines how video frames change over time, looking for unnatural movements that suggest manipulation.

Specific Detection Methods

Temporal ID Network: This network targets inconsistencies that occur over time in deepfake videos, as they often show unusual changes from frame to frame.

3D Morphable Model (3DMM) Generative Network: This method uses 3D facial models to compare generated faces with real ones, identifying differences in lighting and texture.

Face Extraction: This step involves isolating the face from a video to analyze it for inconsistencies in texture and lighting.

CViT (Convolutional Vision Transformers): This hybrid model combines CNNs and transformers to effectively detect both local and broader inconsistencies in videos.

Optical Flow based CNN: This technique analyzes motion patterns between video frames to identify unnatural movements typical of deepfakes.

Dataset Collection and Preprocessing: Gathering a diverse set of genuine and manipulated videos is crucial for training detection models effectively.

EfficientNet-V2 Network: This advanced neural network is designed for fast processing of high-definition videos while maintaining accuracy in detection.

Visual Artifacts Detection: This technique looks for irregularities in individual video frames that may indicate manipulation.

Temporal Features Analysis: This method examines patterns across consecutive frames to identify inconsistencies in facial movements.

Adaptive Manipulation Traces Extraction Network (AMTEN): This network detects subtle traces of manipulation in images and videos.

Feature Extraction Using CNN: CNNs analyze video frames to extract features that can indicate whether a video has been altered.

Error Level Analysis (ELA): This technique detects digital manipulations by analyzing compression artifacts in images.

Support Vector Machines (SVM): SVMs classify videos based on features extracted from them, helping to distinguish real from fake.

K-Nearest Neighbors (K-NN): This algorithm classifies videos by comparing their features to those of known real and fake videos.

These methods represent a variety of approaches to improving the accuracy and reliability of deepfake detection, which is crucial for maintaining trust in digital media.

IV. IMPLEMENTATION

Deepfake technology involves creating fake videos or images that look real, and it relies on various techniques for both creation and detection. Here’s a simplified overview of how deepfakes are made and how they can be detected:

Creation of Deepfakes

Facial Landmark Detection:Dlib is a library that helps detect specific points on a person's face, known as facial landmarks. These landmarks include the eyes, nose, mouth, and jawline.Detecting these points accurately is crucial for creating realistic deepfakes, as it allows one face to be mapped onto another while maintaining natural expressions and movements.Dlib uses a type of machine learning model called Convolutional Neural Networks (CNNs) to identify these landmarks. The model is trained on a large dataset of facial images, enabling it to work well under different poses, expressions, and lighting conditions.The process starts with detecting faces in images or videos. Once a face is found, Dlib identifies the facial landmarks and generates coordinates for these points. These coordinates help align and morph one face onto another, ensuring the deepfake looks realistic.Using Dlib’s facial landmark detection shows how powerful CNN models can be in handling complex image tasks, making it easier to create deepfakes that are hard to tell apart from real media.

Detecting Deepfakes

EfficientNet for Detection: EfficientNet is a modern CNN architecture known for its efficiency and accuracy in detecting deepfakes. It improves performance by scaling three key aspects: depth (number of layers), width (number of units in a layer), and resolution (size of the input image) in a balanced way. Steps in Using EfficientNet for Deepfake Detection: Preprocessing: Videos are broken down into frames and adjusted to a uniform size for optimal processing.

Face Detection: Although EfficientNet doesn’t perform face detection directly, this step is necessary to isolate the parts of the video that need to be analyzed.

Feature Extraction: EfficientNet processes the frames to extract features, capturing everything from simple textures to complex patterns related to facial expressions.

Analysis: The extracted features are examined for inconsistencies or signs of manipulation, such as unnatural facial expressions or lighting issues.

Classification: Each frame is classified as either genuine or manipulated, and the results are combined to assess the authenticity of the entire video.

Mathematical Foundations: EfficientNet uses a compound scaling method to balance the scaling of width, depth, and resolution. This ensures that the network remains efficient while improving its ability to process complex images. By using this balanced approach, EfficientNet achieves high accuracy with fewer parameters, making it faster and less resource-intensive, which is essential for quickly analyzing large amounts of video data.

VII. CONCLUSION

Digital media has been greatly changed by deep learning techniques, especially the rise of deepfakes. These computer-generated videos have caused worries in many areas, showing the need for good ways to detect them. Our review looked at how deepfake creation methods have become more advanced at making deceptive media. It also explored the many techniques and new approaches being used to detect deepfakes, from analyzing video over time to using complex neural networks. These methods offer hope for identifying manipulated content. As deepfakes keep getting better, it is crucial to keep working to prevent their negative effects.

REFERENCES

[1] Thanh Thi Nguyen,et al “Deep Learning for Deepfakes Creation and Detection”.arXiv:1909.11573v5 [cs.CV] 11 Aug 2022

[2] Liwei Deng, Hongfei Suo, Dongjie Li, "Deepfake Video Detection Based on EfficientNet-V2 Network", Computational Intelligence and Neuroscience, vol. 2022, Article ID 3441549, 13 https://doi.org/10.1155/2022/3441549 pages, 2022.

[3] Sukanya S. Shet, et al “Deepfake detection in digital media forensics” International Conference on Intelligent Engineering Approach (ICIEA- 2022)

[4] Yuezun Li, et al.”DeepFake-o-meter: An Open Platform for Detection”https://engineering.buffalo.edu DeepFake

[5] Zhiqing Guo, et al “Fake face detection via adaptive manipulation traces extraction network” sciencedirect.com/science/article/pii/S107731422100014 X,

March 2021
[6] Peipeng Yu, et al “A Survey on Deepfake Video Detection” ietresearch.onlinelibrary.wiley 09 April 2021
[7] Deressa Wodajo, Solomon Atnafu “Deepfake Video Detection Using Convolutional Vision Transformer” arxiv.org ,22 Feb 2021
[8] Delaney Conrad “A Machine Learning Approach to Deepfake Detection” cornerstone.lib, feb 2023
[9] Shraddha Suratkar “Deep Fake Video Detection Using Transfer Learning Approach”, link.springer, 11 October 2022
[10] Duha A. Sultan,”A Comprehensive Survey on Deepfake Detection Techniques ”International journal of intelligence, feb 2022
[11] Hasin Shahed Shad, et al “Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network”
,16 Dec 2021
[12] Shuya Fang, et al “DeepFake Video Detection through Facial Sparse Optical Flow based Light CNN”, Journal of Physics, jan 2022
[13] Sarfraj Ahmed” Fast and Effective Deepfake Detection Method Using Frame Comparison Analysis”, researchsquare,
June 9th, 2023
[14] Akul Mehra “Deepfake Detection using Capsule Networks with Long Short-Term Memory Networks”, essay.utwent, aug 2020
[15] Saima Waseem, et al “Multi-attention-based approach for deepfake face and expression swap detection and localization”,springer open,march 2023
[16] Shashi Rekha G,et al “Deepfake: Creation and Detection using Deep Learning ”Ijraset,2023-05-21
[17] Jayesh S. Pathade, et al “Deepfake Detection through Deep Learning”ijarsct,November 2022
[18] Matthew Groh,et al”Deepfake detection by human crowds, machines, and
crowds”,pnas,December 28, 2021 machine-informed
[19] Md Shohel Rana,et al “Deepfake Detection: A Systematic Literature Review ”Iee.org 2022
[20] Thanh Thi Nguyen, et al “Deep learning for deepfakes creation and detection: A survey”, Science direct, October 2022,