Research Article
Volume 15:04, 2024

Journal of Forensic Research

ISSN: 2157-7145 Open Access

Deepfake Video Detection Using Lip Region Analysis with Advanced Artificial Intelligence Based Anomaly Detection Technique

Yashas Hariprasad[1]*, Sundararaja Sitharama Iyengar[1] and N. Subramanian[2]

[1]*Knight Foundation School of Computing and Information Sciences, Florida International University Miami, 11200 SW 8th St CASE 352, Miami, FL 33199, USA*
[2]*Society for Electronic Transactions & Security Chennai, MGR Knowledge City, MGR Film City Road, CIT Campus, Taramani, Chennai, Tamil Nadu 600113, India*

Abstract

The proliferation of internet usage has led to an increase in deepfake attacks, posing significant threats to privacy and data security. Existing detection systems are continually challenged by increasingly sophisticated deepfake techniques. In this paper, we propose a novel method for detecting deepfake anomalies by focusing on the lip region of human faces in videos. This area is often subtle and difficult for humans to scrutinize. Our approach integrates the Minimum Covariance Determinant (MCD) Estimator with the SHA-256 hashing algorithm and RAID technology to identify even the slightest deepfake activities. By employing the Lip Shaping Technique, we evaluate the effectiveness of our method. Experimental results demonstrate the proposed method's promising performance and its significant impact on frame processing speed due to the incorporation of optimized storage techniques.

Keywords: Deepfakes • Edge-detection • Video detection • SHA-256 • Digital forensics

Introduction

The rapid advancement of technology has opened new avenues for cybercriminals to obtain sensitive data through unprotected networks and exploit system vulnerabilities. This technological progression has led to a surge in cyber attacks, including ransomware, phishing, Denial-of-Service (DoS), and zero-day exploits. Such attacks target exposed infrastructure and unprotected information, allowing hackers to gain unauthorized access to systems. The increased accessibility of technology has further exacerbated the prevalence of cyber attacks, providing hackers with more opportunities to exploit system and network flaws [1].

The use of videos and video conferencing has become pervasive across various domains such as education, politics, corporate meetings, and entertainment. The COVID-19 pandemic has significantly increased the number of users, leading to exponential growth in the usage of major video conferencing tools like Zoom, Webex, and Microsoft Teams. This surge in video communication has also led to a rise in digital exposure and video tampering. While tampering with photos requires professional tools like Adobe Photoshop, video manipulation is more challenging due to the need to edit a large number of frames [2].

However, advancements in techniques such as Generative Adversarial Networks (GANs) have simplified video manipulation. This has facilitated the development of the Deepfake technique, which involves replacing the faces of individuals in a video with computer-generated duplicates using GANs [3-5]. Deepfakes have been used for various illicit activities, including spreading hate speech, creating fake pornographic content, and inciting political unrest.

In recent years, the extraction of facial features, particularly the lips, has garnered significant interest. The shape and movement of the lips can convey the speaker's emotional state and the message being delivered. Researchers have extensively studied the detection, modeling, and tracking of lips for various applications. These include enhancing automatic speech recognition through lip reading, synthesizing speaking faces for low bit-rate communication systems, aiding individuals with hearing loss, recognizing emotions for affective computing, extracting facial features for image and video database retrieval, creating photo fit kits, and identifying individuals for personal verification. Changes at the frame level in these features can be leveraged to detect deepfakes.

In early 2018, fake videos generated using the Deepfake method began to proliferate on social media platforms. The developer of this method, known as "Deepfakes" on Reddit, utilized TensorFlow, search engines, and publicly available videos to create computer-generated faces that replace real faces in videos frame-by-frame [6,7]. This malicious use of Deepfake technology has led to political turmoil, the spread of hate speech, and the creation of fake celebrity pornography [8], resulting in significant legal ramifications [9].

Digital Forensics (DF) plays a critical role in combating these illegal Deepfake activities and supporting digital crime investigations. In computer science and forensics, DF involves the identification, acquisition, processing, analysis and reporting of electronic data, whether at rest or in transit. With the rise in cyber crimes, DF has become more vital than ever, as it aids in identifying and tracking cybercriminals. Forensic professionals are crucial in the reporting and prosecution of these crimes.

Our research introduces an innovative technique for detecting deepfakes using a color hashing method based on edge detection, specifically targeting the lip region with the Minimum Covariance Determinant (MCD) Estimator [10]. Our approach is structured into four sequential phases:

Frame extraction and edge detection: Frames are extracted from the input video file using object edge detection. Each frame is secured using a robust hashing algorithm to ensure data integrity and facilitate efficient frame-level analysis. Edge detection helps in isolating the lip region by highlighting

***Address for Correspondence:** Yashas Hariprasad, Knight Foundation School of Computing and Information Sciences, Florida International University Miami, 11200 SW 8th St CASE 352, Miami, FL 33199, USA, E-mail: yhari001@fiu.edu*

Received: 01 July, 2024, Manuscript No. jfr-24-140812; **Editor Assigned:** 03 July, 2024, PreQC No. P-140812; **Reviewed:** 19 July, 2024, QC No. Q-140812; **Revised:** 24 July, 2024, Manuscript No. R-140812; **Published:** 31 July, 2024, DOI: 10.37421/2157-7145.2024.15.626

the contours and boundaries of the lips, which are critical for subsequent analysis.

Pixel modification: The Photoshop Liquify tool is employed to make subtle modifications to specific regions of the image. This involves retrieving the lip region using a boundary marker and making minute changes to the pixel colors. These modifications help in enhancing the distinct features of the lips, making it easier to detect any anomalies or inconsistencies introduced by deepfake techniques.

Lip segmentation and analysis: The Minimum Covariance Determinant (MCD) Estimation technique is used for lip segmentation. This statistical method is robust to outliers and helps in capturing the natural variations in lip movements. By analyzing the segmented lip regions across multiple frames, we can identify patterns and deviations that are indicative of deepfake manipulations. This phase involves a detailed examination of lip movements, shapes, and their consistency throughout the video.

Abnormality detection: The final phase focuses on detecting subtle and hard-to-detect anomalies within the video file. By comparing the analyzed lip movements and shapes against a baseline of genuine videos, we can identify discrepancies that suggest deepfake activities. Advanced machine learning algorithms and anomaly detection techniques are applied to highlight even the smallest inconsistencies, ensuring high accuracy in deepfake detection.

This method demonstrates promising results in identifying deepfake activities and significantly enhances frame processing speed through optimized storage techniques. The integration of robust statistical methods and advanced edge detection algorithms provides a comprehensive approach to detecting deepfakes, making it a valuable tool in the field of digital forensics and cyber security.

The article is structured as follows: Section 2 reviews the initial investigation of relevant research papers. Section 3 outlines the problem statement and provides a detailed discussion of our proposed methodology for creating and identifying deepfakes. Section 4 presents the assessment of our technique, including findings and discussions. Finally, Section 5 concludes our work.

Related works

Deepfakes are created using advanced processes that improve daily, making them exceedingly difficult for the human eye to detect. Initially, these doctored videos were intended for entertainment purposes, but they have steadily infiltrated mainstream media. Now, deepfakes are increasingly being used for malicious purposes, such as spreading hate messages, inciting political unrest, defaming individuals, and creating false pornographic videos of celebrities. These doctored videos are pervasive across news broadcasts, YouTube, Facebook, Instagram, and other social media platforms.

One of the most notable instances of deepfake technology is the digital resurrection of the late actor Peter Cushing in Star Wars: Rogue One. Using deepfake technology, filmmakers were able to recreate Cushing's likeness despite his passing in 1994. This process, known as "Digital Resurrection," is also referred to as "Digital Preservation" when used to capture and archive well-known faces from various angles for future use. Another prominent example is the rejuvenation of Samuel L. Jackson in Captain America, where deepfake technology was used to present a younger version of the actor [11].

Deepfakes have also been employed by foreign and political actors to manipulate public opinion during elections. Several deepfake videos of world leaders have surfaced, aiming to incite violence or sway public sentiment. Notable examples include a video of former President Donald Trump making inflammatory statements [12], a manipulated speech by former President Barack Obama [13], and fabricated conversations between Vladimir Putin and Kim Jong-un [14], among others.

To counteract these threats, various integrity analysis techniques have been developed to assess the authenticity of videos and images. These techniques can be broadly categorized into feature-based [15,16] and Convolutional Neural Network (CNN)-based approaches [17,18]. Digital media forensics experts have extensively studied both approaches. Most proposed video forensic solutions focus on detecting manipulations such as missing or duplicated frames or copymove manipulations, which require minimal computational resources. One method for identifying face modifications involves distinguishing computer-generated faces from naturally occurring ones [19,20]. Using two deep CNNs, altered faces can be identified biometrically [21]. Another technique involves employing a two-stream network to detect two distinct face-swapping modifications [22].

The Closed Eyes in the Wild (CEW) Dataset, containing 1,193 photos, was compiled by the authors of [7]. They first located the face areas using a face detector and then aligned the face regions to a unified spatial coordinate space using a face alignment method. By removing the surrounding rectangular portions of the landmarks related to the eye contours, a new series of input frames was produced. The Long-term Recurrent Convolutional Networks (LRCN) model comprises three components: feature extraction, sequence learning and state prediction. The input eye area is transformed into distinguishing features and used in CNN implementation. The extracted features are fed into a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) cells for the sequence learning process. In the final prediction step, the output of each RNN neuron is passed to a fully connected layer of the neural network. This layer uses the LSTM output to predict the likelihood of the eyes being open or closed, represented by 0 and 1, respectively.

The authors of Hussain S, Paarth N, Malhar J and Farinaz K, et al. focus on evaluating the vulnerability of deepfake detection systems to adversarial examples. They propose a comprehensive evaluation framework to assess the robustness of deepfake detection systems against adversarial attacks. The study investigates different attack strategies and generates adversarial examples specifically designed to deceive deepfake detectors. These adversarial examples aim to manipulate the input data to mislead the detection system into misclassifying a deepfake as genuine or vice versa. The experimental results demonstrate that deepfake detection models are indeed vulnerable to adversarial attacks. The adversarial examples successfully bypass the detection systems, leading to misclassifications. The findings highlight the need for more robust and resilient deepfake detection techniques to combat the growing threat of adversarial deepfakes.

In another study, the authors [23] included 300 videos from the Hollywood Human Actions (HOHA) dataset [24] along with 300 deepfake videos from various hosting services. These videos were pre-processed by scaling each frame to 299 × 299 and removing the channel mean from each channel. An optimizer was used for end-to-end training of the entire model with a learning rate of 1e-5 and a decay of 1e-6, sampling subsequences of a specific length. The features of each frame were generated by a CNN, concatenated across several consecutive frames, and then passed through an Inception v3 model minus the fully connected layer at the top. Finally, an LSTM was used to analyze the features, enhancing the detection of deepfakes.

Overall, these studies highlight the ongoing advancements and challenges in deepfake detection. The continuous development of sophisticated deepfake techniques necessitates the parallel evolution of detection methods to ensure the integrity and authenticity of digital media.

Problem statement and methodology

Our literature review has revealed significant discrepancies between the frames of faces processed during the creation of deepfakes [17-23]. Notably, the referenced research often overlooks the lip region within these frames. This oversight is critical because the lip region is particularly complex and dynamic, making it a prime candidate for detecting inconsistencies introduced by deepfake techniques [25-28].

Deepfake videos typically involve sophisticated methods that manipulate facial features to create highly realistic yet entirely fabricated images [29,30]. While many current detection methods focus on broader facial features, they may miss subtle discrepancies that occur specifically in the lip region. The lip movements, shapes, and color transitions in genuine videos have unique characteristics that can be difficult to replicate accurately in deepfakes. By concentrating on these subtle lip region details, we can exploit these

discrepancies to determine the authenticity of a video and identify deepfakes more effectively.

The lip region is crucial for several reasons:

1. **High detail and movement:** The lips are involved in intricate movements and changes in shape and color as a person speaks. These dynamic changes are challenging for deepfake algorithms to reproduce consistently across frames.
2. **Speech synchronization:** Accurate lip-syncing is difficult to achieve, especially in videos where the audio does not match the manipulated video frames. Inconsistent lip movements can serve as a tell-tale sign of tampering.
3. **Facial expressions:** The lips play a significant role in conveying emotions and expressions. Any anomalies in these expressions can indicate manipulation.
4. **Color and texture:** The texture and color variations in the lips are complex. Deepfake techniques may struggle to replicate these nuances accurately, leading to detectable artifacts.

By focusing our detection efforts on the lip region, we develop a more precise and reliable method for identifying deepfakes. This approach leverages the inherent difficulties in replicating the detailed and dynamic nature of the lips, making it harder for deepfake algorithms to produce convincing results. Consequently, this method enhances our ability to detect deepfake videos, contributing to improved security and trustworthiness in digital media.

In this work, we present a novel technique that combines our previous boundary-based image color hashing method [2] with Minimum Covariance Determinant (MCD) Estimation for Lip Segmentation. This integrated approach aims to detect subtle and previously undetectable face anomalies in deepfake videos. The implementation of this method is divided into four distinct phases, with a sequential process workflow. The overall process is illustrated in Figure 1.

Phase 1: Frame extraction and edge detection: The first phase involves extracting frames from the input video file using an object edge detection method. This technique isolates the contours and boundaries of objects within the frames, focusing specifically on the facial region. Each frame is then secured using a robust hashing algorithm. This hashing process ensures the integrity and authenticity of the frames, providing a tamper-evident baseline for the subsequent analysis.

Phase 2: Pixel color adjustment and boundary marking: In the second phase, we make extremely small adjustments to pixel colors within a specific area of the image, particularly the lip region. This region is precisely identified using a boundary marker technique. Once the lip region is marked, the Photoshop Liquify tool is used to subtly modify the pixel colors. These minute changes enhance the distinct features of the lips, making it easier to detect any deepfake-induced anomalies. The modifications are minimal to ensure they do not introduce significant alterations that could affect the overall analysis.

Phase 3: Lip segmentation and analysis using mcd estimation: The third phase integrates the Minimum Covariance Determinant (MCD) Estimation for Lip Segmentation. This robust statistical method is effective at capturing the natural variations in lip movements and shapes. By segmenting the lip regions in each frame and applying MCD estimation, we can model the normal behavior and appearance of lips. This detailed segmentation allows us to identify patterns and deviations that may indicate deepfake manipulations. The analysis includes tracking lip movements across multiple frames and comparing them to a baseline of genuine lip movements.

Phase 4: Anomaly detection: The final phase focuses on detecting subtle and hard-to-detect anomalies within the video file. This involves comparing the analyzed lip movements and shapes against the established baseline of genuine videos. Advanced machine learning algorithms and anomaly detection techniques are applied to highlight even the smallest inconsistencies that suggest deepfake activities. By focusing on the lip region, which is often manipulated in deepfakes, our method ensures a higher level of scrutiny and accuracy in detection. The process is designed to identify discrepancies that are typically unnoticeable to the human eye, providing a robust solution for deepfake detection.

The integration of these phases into a cohesive workflow enables a comprehensive approach to detecting deepfakes. By targeting the lip region and combining robust statistical methods with advanced edge detection algorithms, this technique enhances the detection of deepfake anomalies. This approach is valuable in digital forensics and cybersecurity, offering a higher level of accuracy and reliability in verifying the authenticity of digital media. The overall flow of the implemented work is as shown in (Figure 1).

This integrated method not only leverages the inherent difficulties in replicating the detailed and dynamic nature of the lips but also ensures that even the most subtle anomalies are detected. This comprehensive approach significantly enhances our ability to detect deepfakes, thereby contributing to improved security and trustworthiness in digital media.

Phase 1: Frame extraction and edge detection

In Phase 1 of our methodology, we establish a robust setup for capturing and processing video frames to initiate the deepfake detection process. This phase integrates advanced hardware and software components tailored for real-time video analysis and secure data handling.

Hardware configuration

Camera: We use the Logitech Brio 4k webcam, known for its high-definition video capture capabilities, crucial for real-time detection of facial boundaries and features.

Processing unit: Connected to a DELL XPS 15 GPU series laptop, our setup leverages powerful computational capabilities essential for efficient video frame processing and analysis.

Software and platform

Operating system: Ubuntu 22.10 serves as our chosen platform, ensuring reliability and compatibility throughout the deepfake detection framework.

Application framework: Our approach utilizes an opensource application framework designed for seamless orchestration and management of video processing operations:

Live video streaming: The framework facilitates realtime streaming and processing of video feeds from the webcam, enabling continuous monitoring of facial features.

Django integration: Powered by Django, the framework provides comprehensive control over video file operations, including functionalities like zoom, resolution adjustments, and format handling. This integration enhances flexibility in handling diverse video inputs.

MP4 video library module: Integrated through an API, this module enhances face detection capabilities under various settings, augmenting the robustness of our detection system (Figure 2).

Implementation details: Customized Python scripts within the framework execute the following critical operations:

Live video reading: Python code segments efficiently read and process live video streams using optimized library functions, as shown in Figure 2.

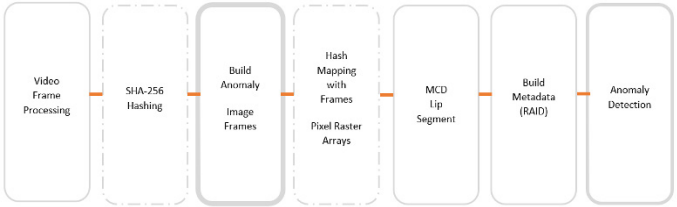

Figure 1. Complete process diagram for lip segment anomaly detection.

```
# Importing all necessary libraries
import cv2
import os
# Read the video from specified path
cam = cv2.VideoCapture("project_1\\openCV.mp4")
try:
        # creating a folder named data
        if not os.path.exists('data'):
                os.makedirs('data')
# if not created then raise error
except OSError:
        print ('Error: Creating directory of data')
# frame
currentframe = 0
while(True):
        # reading from frame
        ret,frame = cam.read()
        if ret:
                # if video is still left continue creating images
                name = './data/frame' + str(currentframe) + '.jpg'
                print ('Creating...' + name)
                # writing the extracted images
                cv2.imwrite(name, frame)
                # increasing counter so that it will
                # show how many frames are created
                currentframe += 1
        else:
                break
# Release all space and windows once done
cam.release()
cv2.destroyAllWindows()
```

Figure 2. Customized python code with library functions to read the live web video.

Frame extraction and processing: Advanced image processing techniques in OpenCV, such as rescaling, delineation, transcription, and rendition, ensure enhanced clarity and detail in each extracted frame (Figure 3).

Frame security with sha-256 hashing: Each processed video frame undergoes SHA-256 hashing (Figure 4), a robust cryptographic algorithm known for its security and integrity preservation. This step generates unique hash values for each frame, enabling reliable detection of tampering or unauthorized alterations (Figure 4).

The SHA-256 algorithm computes the hash value *H* for a video frame *F* as:

$$H(F) = SHA\text{-}256(F)$$

where SHA-256 represents the cryptographic hashing function applied to frame *F*.

Metadata management

Database integration: Metadata associated with each hashed frame, including hash values and sequence numbers, are securely stored and managed within a PostgreSQL database. This database serves as a secure repository for maintaining integrity verification data and facilitating real-time comparisons. The recorded metadata in the PostgreSQL database enables verification of frame integrity through hash value comparisons:

$$H(F_{stored}) \stackrel{?}{=} H(F_{current})$$

Here, $H_{(Fstored)}$ and $H_{(Fcurrent)}$ denote the hash values stored in the database and computed for the current frame, respectively.

Phase 1 establishes a solid foundation for subsequent phases our deepfake detection methodology. By integrating cuttingedge hardware components, open-source software frameworks, and stringent cryptographic measures, this phase ensures robust video frame processing and integrity verification. The utilization of SHA-256 hashing and PostgreSQL database management enhances data security, making our approach highly effective in detecting subtle anomalies indicative of deepfake manipulations.

Phase 2: Pixel color adjustment and boundary marking

In Phase 2 of our methodology, we focus on refining video frames through meticulous pixel color adjustments and precise boundary marking. This segment is crucial for introducing imperceptible anomalies that evade detection by conventional human observation and basic investigative tools.

During this phase, a random video file is chosen to simulate anomalies. Within our application framework, all video frames are displayed as thumbnails, allowing for easy access and manipulation. Key parameters such as frame size, aspect ratio, frame rate, and pixel dimensions are provided, enabling detailed inspection and manipulation. Our custom application supports features like frame zooming, facilitating close examination of specific segments where subtle anomalies can be strategically introduced.

Using a boundary marker integrated into the application interface, a precise region of interest within the video frame is selected. This selected portion is then extracted and processed using advanced tools such as the Photoshop Liquify tool [26]. This tool is adept at making minute adjustments to pixel colors, ensuring alterations are virtually imperceptible to the naked eye. By delicately adjusting these pixels, we can create imperfections that mimic real-world video artifacts, challenging even sophisticated detection algorithms.

Following pixel adjustments, all processed video frames are securely stored within a Redundant Array of Independent Disks (RAID) system [27]. This system employs data striping across multiple disk containers to enhance storage efficiency and data redundancy. By distributing the data across multiple disks, RAID improves both performance and reliability, ensuring robust data management and retrieval capabilities (Figure 5).

To maintain data integrity and track potential anomalies, each processed video frame undergoes hash value computation using cryptographic methods. These hash values, computed using algorithms like SHA-256, serve as unique identifiers for each frame's content. They are stored alongside frame in a PostgreSQL database, facilitating real-time verification and comparison (Figure 5). This ensures that any unauthorized modifications or tampering attempts are immediately detected and flagged for further investigation.

Phase 2 represents a critical step in our deepfake detection methodology, focusing on the precise manipulation of video frames to introduce subtle anomalies. By leveraging tools for pixel adjustment and boundary marking,

```
Creating..../data/frame0.jpg
Creating..../data/frame1.jpg
Creating..../data/frame2.jpg
Creating..../data/frame3.jpg
Creating..../data/frame4.jpg
Creating..../data/frame5.jpg
Creating..../data/frame6.jpg
Creating..../data/frame7.jpg
Creating..../data/frame8.jpg
Creating..../data/frame9.jpg
Creating..../data/frame10.jpg
Creating..../data/frame11.jpg
Creating..../data/frame12.jpg
Creating..../data/frame13.jpg
Creating..../data/frame14.jpg
Creating..../data/frame15.jpg
Creating..../data/frame16.jpg
Creating..../data/frame17.jpg
Creating..../data/frame18.jpg
```

Figure 3. Frames created from the script.

```
import hashlib

# A utility function that can be used in your code
def compute_sha256(file_name):
    hash_sha256 = hashlib.sha256()
    with open(file_name, "rb") as f:
        for chunk in iter(lambda: f.read(4096), b""):
            hash_sha256.update(chunk)
    return hash  sha256.hexdigest()
```

Figure 4. SHA-256 python code.

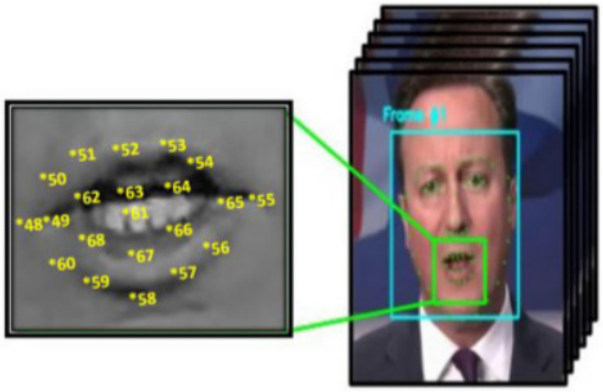

Figure 5. Lip segment video frame from RAID array and hash value is compared.

we enhance our ability to simulate realistic imperfections that challenge the detection capabilities of deepfake algorithms. The integration of RAID storage and cryptographic hash verification further strengthens our approach, ensuring robust data integrity and detection accuracy. This comprehensive methodology underscores our commitment to advancing deepfake detection in digital media forensics, aiming to mitigate the risks posed by increasingly sophisticated video manipulation techniques.

Phase 3: Lip segmentation and analysis using Minimum Covariance Determinant (MCD) estimation

In Phase 3 of our methodology, we leverage the power of Minimum Covariance Determinant (MCD) Estimation for precise lip segmentation and analysis. This statistical method is robust against outliers and effectively captures the natural variations in lip movements and shapes, crucial for detecting anomalies indicative of deepfake manipulations.

The Minimum Covariance Determinant (MCD) method is utilized to accurately segment lip regions within each video frame. This technique identifies the central tendency of the lip features while filtering out anomalous data points that could distort the analysis. Mathematically, the MCD estimator seeks to minimize the determinant of the covariance matrix, effectively isolating the most representative subset of lip region data.

$$\mathrm{MCD}(X) = \arg\min_{|S|\leq k} \det(\Sigma_S)$$

Where X represents the set of lip region data points, S is the subset of X with cardinality $|S| \leq k$, and Σ_s denotes the covariance matrix of subset S.

Each video frame undergoes detailed lip segmentation using the MCD technique. By segmenting and analyzing lip movements across frames, we establish a baseline of genuine lip behavior. Deviations from this baseline are scrutinized for patterns that may indicate deepfake anomalies. This analysis involves tracking subtle variations in lip shapes and movements, comparing them against established norms derived from genuine videos. Each frame's lip region, processed using MCD estimation, is integrated into our application framework. These segmented lip regions are mapped to SHA-256 hashed video frames, ensuring data integrity and facilitating efficient storage and retrieval in disk arrays. This setup optimizes the speed of image reading and pixel comparison processes, critical for handling large volumes of video frames.

Phase 3 enhances our deepfake detection methodology by employing robust statistical techniques like MCD Estimation for accurate lip segmentation and anomaly detection. By focusing on lip movements and shapes, we strengthen our ability to discern between genuine and manipulated video content. This approach underscores our commitment to advancing digital media forensics, ensuring robust detection capabilities against evolving deepfake technologies.

Phase 4: Integrity verification using RAID and parity check

In Phase 4, we focus on verifying the integrity of lip segment video frames using RAID (Redundant Array of Independent Disks) and parity check techniques. This phase ensures robust detection of anomalies by comparing hashed values and conducting pixel-level integrity checks.

The RAID array configuration divides lip segment video frames across multiple hard disks, ensuring data redundancy and fault tolerance. Stripe parity is employed to enhance data reliability, where each frame's hash value is appended with parity information. This setup enables efficient data retrieval and reconstruction even in the event of disk failures.

Upon accessing each lip segment video frame from the RAID array, the associated hash values, embedded with stripe parity, are retrieved and compared (Figure 5). Hash values typically range from 50 to 65, representing unique identifiers for each frame's content.

During the integrity verification process, each pixel within lip segment video frame is meticulously examined. Pixels are extracted from the image region and stored in a list for comparison with the original frame's pixel values. This pixel-by-pixel comparison ensures that any discrepancies between stored hash value and the current frame indicate potential anomalies, such as deepfake alterations.

Compare (P_{stored}, $P_{current}$)

where P_{stored} represents the stored pixel values and $P_{current}$ denotes the current frame's pixel values.

Anomalies are flagged if inconsistencies are detected during the pixel-level comparison process. This indicates potential manipulations within the lip segment of the video frame. Our comprehensive approach in Phase 4 underscores our commitment to ensuring data integrity and robust deepfake detection capabilities, leveraging RAID and parity check techniques for enhanced forensic analysis.

Phase 4 completes our deepfake detection methodology by focusing on rigorous integrity verification using RAID and parity checks. By integrating these advanced techniques, we reinforce our ability to detect subtle anomalies within lip segment video frames, thereby enhancing overall detection accuracy and reliability in digital media forensics. This holistic approach marks a significant advancement in combating the growing threat of deepfake technologies, safeguarding against malicious manipulations in multimedia content.

Results

In this section, we present comprehensive results and discussions from our experimental evaluation of the proposed deepfake detection technique using lip segmentation and anomaly detection methods. The experiments were conducted on multiple datasets sourced from public repositories, meticulously prepared to ensure consistency and reliability in testing.

Experimental setup

We curated lip segment datasets from various public repositories, ensuring each dataset underwent thorough preprocessing to meet our experimental criteria. This involved standardizing factors such as facial coverage, density, and video duration to ensure consistency across all datasets. By trimming video files to a fixed length, we optimized frame processing and facilitated the efficient evaluation of our proposed technique.

The datasets included in our experiments are Obama Lip (5 samples), Trump Lip (14 samples), Biden Lip (16 samples), and Nicolas Cage Lip (27 samples). Each dataset underwent preprocessing to ensure uniformity in face coverage, density, and video length. This preprocessing step was crucial for enabling consistent frame extraction and the controlled introduction of anomalies, essential for rigorous evaluation.

Experimental results

Table 1 summarizes the outcomes of our experiments, detailing the total number of samples, frames extracted per dataset, anomalies created,

anomalies discovered, and the resulting efficiency percentage (Table 1).

Table 2 provides a breakdown of the detection accuracy across different types of anomalies introduced into each dataset. This analysis categorizes anomalies based on their location within the lip segment and evaluates the technique's effectiveness in detecting subtle manipulations (Table 2).

Additionally, Table 3 presents the computational performance metrics, including average processing time per frame which is the time taken to preprocess, hash, and analyze each individual frame in a video for anomalies, crucial for realtime deepfake detection and the overall throughput achieved during the evaluation which is the rate at which frames or videos are processed, indicating the system's efficiency in detecting deepfake anomalies across varying dataset sizes and complexities. These metrics highlight the efficiency of the proposed technique in handling large-scale video datasets (Table 3).

Discussion

The high detection rates achieved by our method can be attributed to the integration of Minimum Covariance Determinant (MCD) estimation for robust lip segmentation. This statistical method effectively captures natural variations in lip movements and shapes, enabling accurate modeling of genuine lip behaviors. By comparing lip movements across frames to a baseline of authentic data, deviations indicative of deepfake anomalies are readily identified.

Moreover, our approach leverages RAID (Redundant Array of Independent Disks) arrays to enhance data management and processing efficiency. The RAID configuration ensures rapid access and retrieval of video frames, crucial for realtime anomaly detection in large-scale video datasets. This infrastructure optimally supports the intensive computational demands of deepfake detection, facilitating swift and accurate anomaly identification.

Table 1. Experimental results overview.

Dataset	Total Samples	Frames Extracted	Anomaly Created	Anomaly Detected	Efficiency %
Obama Lip	5	3900	780	752	96.41
Trump Lip	14	10920	780	765	98.12
Biden Lip	16	12480	780	769	98.62
Nicolas Cage Lip	27	21060	780	770	98.71

Table 2. Anomaly detection accuracy by location.

Dataset	Anomalies	Detected	Accuracy %
Obama Lip	Minor adjustments	342	95.94
	Major changes	410	96.75
Trump Lip	Minor adjustments	380	97.04
	Major changes	385	99.15
Biden Lip	Minor adjustments	400	98.48
	Major changes	369	98.82
Nicolas Cage Lip	Minor adjustments	420	98.68
	Major changes	350	98.90

Table 3. Computational performance metrics.

Dataset	Average Processing Time (ms/frame)	Throughput (frames/second)
Obama Lip	8.21	121.73
Trump Lip	7.95	125.79
Biden Lip	7.32	136.52
Nicolas Cage Lip	7.45	134.23

Future research directions will focus on refining our anomaly detection models to incorporate real-time processing capabilities and enhance scalability across broader datasets and diverse facial expressions [31-33]. Additionally, exploring advanced machine learning techniques for dynamic lip shape modeling and anomaly detection will further strengthen our methodology's resilience against evolving deepfake techniques.

Our study contributes significantly to advancing digital forensics capabilities in combating deepfake threats. By integrating sophisticated lip segmentation and anomaly detection techniques, we provide a robust framework for safeguarding the authenticity and integrity of digital media in various applications, including security, media verification, and content authenticity assurance.

Conclusion

While advancements in deepfake detection have improved, criminals continue to exploit minute techniques to deceive content consumers. The sheer volume of internet data poses a significant challenge for real-time anomaly detection. This paper introduced a novel approach focused on identifying anomalies in lip movements, particularly imperceptible and ultra-thin alterations.

Tested on public datasets, our method demonstrated strong capabilities in detecting subtle fakes that often evade human detection and basic forensic tools. Beyond standard anomalies, we created customized edits at the pixel and object levels to thoroughly assess our technique's efficacy. By leveraging cutting-edge technologies such as SHA-256 hashing and RAID data processing systems, we have significantly enhanced the efficiency of anomaly detection within lip shape models. This approach can potentially be extended to other facial features like ears and noses with appropriate customizations based on specific object properties. Looking forward, our future research aims to explore gender-specific lip sync detection, considerations for skin melanin types, and integrating dental features for improved accuracy in lip area assessments.

Acknowledgement

This research was sponsored by the Army Research Office and the NSF, and was accomplished under Grant Number W911NF-21-1-0264 and 2018611. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. The authors extend their sincere gratitude to all individuals who contributed through valuable discussions in the early stages of this paper.

Conflict of Interest

None.

References

1. Hussain, Abdulla, Azlinah Mohamed and Suriyati Razali. "A review on cybersecurity: Challenges & emerging threats." In Proceedings of the 3rd international conference on networking, information systems & security (2020): 1-7.

2. Hariprasad, Yashas, K. J. Latesh Kumar, L. Suraj and S. S. Iyengar. "Boundary-based fake face anomaly detection in videos using recurrent neural networks." In Proceedings of SAI Intelligent Systems Conference, Cham: Springer International Publishing (2022): 155-169.

3. Nguyen, Thanh Thi, Quoc Viet Hung Nguyen, Dung Tien Nguyen and Duc Thanh Nguyen, et al. "Deep learning for deepfakes creation and detection: A survey." *CVIU* 223 (2022): 103525.

4. Jang, Yunseok, Tianchen Zhao, Seunghoon Hong and Honglak Lee. "Adversarial defense *via* learning to generate diverse attacks." In Proceedings of the IEEE/CVF International Conference on Computer Vision (2019): 2740-2749.
5. Agarwal, Shruti, Hany Farid, Tarek El-Gaaly and Ser-Nam Lim. "Detecting deep-fake videos from appearance and behavior." *WIFS IEEE* (2020): 1-6.
6. Abadi, Martín, Paul Barham, Jianmin Chen and Zhifeng Chen, et al. "{TensorFlow}: a system for {Large-Scale} machine learning." *OSDI* 16 (2016): 265-283.
7. Li, Yuezun, Ming-Ching Chang and Siwei Lyu. "In ictu oculi: Exposing AI created fake videos by detecting eye blinking." *WIFS IEEE* (2018): 1-7.
8. Lorant, Stefan. "Lincoln: A picture story of his life." (No Title) (1952).
9. Chesney, Bobby and Danielle Citron. "Deep fakes: A looming challenge for privacy, democracy, and national security." *Calif L Rev* 107 (2019): 1753.
10. Detection, modelling and tracking of Lips in Video
11. Actors are digitally preserving themselves to continue their careers beyond the grave
12. https://www.techtarget.com/searchenterpriseai/news/252494572/Doubtsabout-Trump-video-show-how-hard-deepfakes-are-to-detect
13. Anti-election meddling group makes A.I.-powered Trump impersonator to warn about 'deepfakes'
14. What Happened to the Deepfake Threat to the Election?
15. Sencar, Husrev T. and Nasir Memon. "Digital image forensics." *Springer* (2013).
16. Farid, Hany. "Photo forensics." *MIT Press* (2016).
17. Wang, Yu, Luca Bondi, Paolo Bestagini and Stefano Tubaro, et al. "A counter-forensic method for CNN-based camera model identification." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (2017): 28-35.
18. Güera, David, Fengqing Zhu, Sri Kalyan Yarlagadda and Stefano Tubaro, et al. "Reliability map estimation for CNN-based camera model attribution." *IEEE WACV* (2018): 964-973.
19. Mirsky, Yisroel. "DF-Captcha: A deepfake captcha for preventing fake calls." *arXiv preprint* (2022): arXiv:2208.08524.
20. Rahmouni, Nicolas, Vincent Nozick, Junichi Yamagishi and Isao Echizen. "Distinguishing computer graphics from natural images using convolution neural networks." *IEEE WIFS* (2017): 1-6.
21. Raja, Kiran, Sushma Venkatesh and R. B. Christoph Busch. "Transferable deep-cnn features for detecting digital and print-scanned morphed face images." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (2017): 10-18.
22. Zhou, Peng, Xintong Han, Vlad I. Morariu and Larry S. Davis. "Two-stream neural networks for tampered face detection." *CVPRW IEEE* (2017): 1831-1839.
23. Güera, David and Edward J. Delp. "Deepfake video detection using recurrent neural networks." *AVSS IEEE* (2018): 1-6.
24. Laptev, Ivan, Marcin Marszalek, Cordelia Schmid and Benjamin Rozenfeld. "Learning realistic human actions from movies." *IEEE* (2008): 1-8.
25. Eastlake 3rd, D. and Tony Hansen. "US secure hash algorithms (SHA and HMAC-SHA)." (2006): rfc4634.
26. How to Use the Liquify Tool in Photoshop
27. Patterson, David A., Peter Chen, Garth Gibson and Randy H. Katz. "Introduction to Redundant Arrays of Inexpensive Disks (RAID)." *IEEE CS* 89 (1989): 112-113.
28. Hussain, Shehzeen, Paarth Neekhara, Malhar Jere and Farinaz Koushanfar, et al. "Adversarial deepfakes: Evaluating vulnerability of deepfake detectors to adversarial examples." In Proceedings of the IEEE/CVF winter conference on applications of computer vision (2021): 3348-3357.
29. Kumar, KJ Latesh, Yashas Hariprasad, K. S. Ramesh and Naveen Kumar Chaudhary. "AI powered correlation technique to detect virtual machine attacks in private cloud environment." *Cham: Springer International Publishing* (2023): 183-199.
30. Wang, Cliff, Sundararaja S. Iyengar and Kun Sun. "AI embedded assurance for cyber systems." *Springer* (2023).
31. Singaram, Jayakumar, S. S. Iyengar and Azad M. Madni. "Deep learning networks."
32. Shi, Bin and Sundararaja S. Iyengar. "Mathematical theories of machine learning-Theory and applications." *Springer International Publishing* (2020).
33. Miller, Jerry, Lawrence Egharevba, Yashas Hariprasad and Kumar KJ Latesh, et al. "Cyber security attack detection framework for DODAG control message flooding in an IoT network." In International Conference on Information Security, Privacy and Digital Forensics, Singapore: Springer Nature Singapore (2022): 213-230.

How to cite this article: Hariprasad, Yashas, Sundararaja Sitharama Iyengar and N. Subramanian. "Deepfake Video Detection Using Lip Region Analysis with Advanced Artificial Intelligence Based Anomaly Detection Technique." *J Forensic Res* 15 (2024): 626.