FORGERY & DEEPFAKE DETECTION USING COMPUTER VISION & DEEP LEARNING

1 Dr. Mageshwari M
Department of Computer Science and Engineering
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology
Avadi, Chennai, India–600062
drmageshwarim@veltech.edu.in

2 Dr. Kausalya K
Department of Computer Science and Engineering
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology
Avadi, Chennai, India–600062
drkausalyak@veltech.edu.in

3 Sedhu Prasanna M
Department of Computer Science and Engineering
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology
Avadi, Chennai, India–600062
vtu24141@veltech.edu.in

4 Nithish S
Department of Computer Science and Engineering
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology
Avadi, Chennai, India–600062
vtu22587@veltech.edu.in

***Abstract*—Deepfake technology has progressed rapidly because of the availability of very powerful deep learning algorithms, and it has become very easy to manipulate faces, voices, and videos realistically. Although these technologies have opened up many creative avenues, they also pose very serious threats in terms of misinformation, identity theft, political manipulation, and cybercrime. This paper proposes an efficient deepfake and forg- ery detection system based on Convolutional Neural Networks (CNN), Vision Transformers (ViT), and feature-level inconsis- tencies such as facial landmarks, eye blink patterns, texture anomalies, and artifacts in the frequency domain. The system combines location and timing information to enhance the accuracy of deepfake detection in images and videos. The system has high precision in tests and works well against state-of-the-art deepfake methods such as GANs, autoencoders, and diffusion models. Its applications include digital forensics, image verification on social media platforms, and preventing cybercrime.**

***Index Terms*—Deepfake Detection, Computer Vision, Forgery Detection, CNN, Vision Transformer, GAN, Digital Forensics.**

I. Introduction

With the rapid evolution of artificial intelligence and generative models, the creation of counterfeit digital media is becoming increasingly easier and realistic. The current state of deepfake technology allows for realistic manipulation of facial expressions, emotions, voices, and identities in images and videos, making it difficult to distinguish between what is real and what is not. The increasing popularity of social media platforms accelerates the dissemination of these deepfakes, which are used for spreading misinformation, identity theft, political manipulation, and cybercrime.

In the past, image and video forgery detection was based on manually designed forensic features such as artifacts, compression patterns, and pixel anomalies. However, the emergence of advanced deepfake technology based on GANs and diffusion models has made it possible to create high-resolution images with minimal forensic features, thus highlighting the inefficacies of traditional approaches. This has led to the development of deep learning-based forgery detectors that can automatically learn from large datasets to detect deepfakes. These models are capable of detecting minute anomalies in space, time, and frequency domains that go unnoticed by human observers. There is an increasing recognition that combining features from multiple domains can improve the accuracy and robustness of forgery detection.

More sophisticated forgery detectors analyze facial landmarks, blink patterns, head movements, texture anomalies, and spectral anomalies to detect forgeries. When implemented on digital platforms, these tools can help control the dissemination of misleading images and maintain digital trust. Intelligence-driven deepfake detection is becoming increasingly imperative for securing the integrity of information on online platforms.

II. OVERVIEW

A. Brief Overview

The Deepfake Detection System is an intelligent digital forensic assistant that applies AI-powered computer vision algorithms to authenticate the authenticity of images and videos. The system enables users to upload media files to authenticate whether the files are authentic or deepfakes. The system applies the concept of deep learning models to analyze the input data and extract features in the spatial, temporal, and frequency domains. The hybrid model of the system ensures that it is highly accurate, scalable, and able to perform real-time processing, making it applicable in the field of cybersecurity and social media.

B. Problem Statement

The pace at which the development of deepfake technology has reached has led to serious problems in the process of authenticating digital media. The current techniques for authenticating digital media are time-consuming and prone to errors. Most of the current techniques are not robust against new generative models, video compression, and low-quality media. There is a need for an intelligent system that can detect deepfakes effectively.

C. Project Goal

The aim of this project is to design a system that is capable of detecting deepfakes and forgeries using computer vision and deep learning. The system will be able to analyze the spatial textures, temporal anomalies, and frequency domain artifacts in order to detect forgeries. The ultimate aim of this project is to design an accurate and scalable digital forensics system.

D. Limitations of Existing Detection Systems

Most existing deepfake detection methods are associated with particular datasets and fail to generalize to novel manipulation methods. They primarily focus on spatial information while ignoring the temporal and spectral domains. Moreover, their performance degrades when dealing with compressed videos, videos with reduced resolutions, and noisy videos.

III. Literature Review

A. Deepfake Generation Techniques

The history of deepfakes starts with face swaps done using autoencoders. GANs later emerged, taking the technology to the next level and allowing the development of applications such as FaceSwap, DeepFaceLab, and StyleGAN that were increasingly realistic. Diffusion models have now entered the scene, taking realism to the next level.

B. Traditional Forgery Detection Methods

In the early stages of forgery detection, the methods were based on handcrafted features like chromatic aberration, noise patterns, and JPEG compression artifacts. Although these methods are computationally efficient, they are not robust against current deepfake attacks.

C. CNN-Based Detection Approaches

CNNs have gained popularity in the area of deepfake detection because of their ability to learn spatial features automatically. VGG, ResNet, DenseNet, and XceptionNet models have proved to be promising. However, CNNs are mostly local feature detectors and fail to capture global dependencies.

D. Temporal Analysis Methods

Video-based detection methods employ the analysis of temporal inconsistencies through the use of recurrent neural networks like LSTM and GRU. The dynamics of eye blink rate, lip sync, and head movement are some of the features that can be used for the detection of manipulated videos.

E. Transformer-Based and Frequency-Domain Methods

Vision Transformers utilize self-attention mechanisms to capture global context and long-range dependencies. Frequency domain analysis using FFT and wavelet transforms helps to identify GAN-specific artifacts that are not visible in the spatial domain.

F. Identified Research Gaps

Although considerable progress has been made, the current state-of-the-art models are challenged by cross-dataset generalization, robustness to compression attacks, and diffusion-based deepfake adaptability. These challenges create a need for a hybrid model that combines multiple feature domains.

IV. Abbreviations and Acronyms

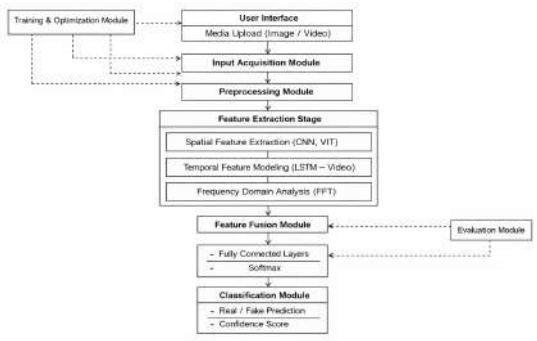
- AI – Artificial Intelligence
- DL – Deep Learning
- CV – Computer Vision
- CNN – Convolutional Neural Network
- GAN – Generative Adversarial Network
- ViT – Vision Transformer
- LSTM – Long Short-Term Memory
- RNN – Recurrent Neural Network
- FFT – Fast Fourier Transform
- DCT – Discrete Cosine Transform
- DFDC – Deepfake Detection Challenge
- API – Application Programming Interface
- ROC – Receiver Operating Characteristic
- AUC – Area Under Curve
- TPR – True Positive Rate
- FPR – False Positive Rate
- RGB – Red Green Blue
- FPS – Frames Per Second
- JPEG – Joint Photographic Experts Group
- GPU – Graphics Processing Unit

V. PROPOSED METHODOLOGY

A. Requirement Analysis

The system requirements include the ability to correctly identify forged images and videos, resistance to compression, scalability, and real-time processing. The non-functional requirements include data security, reliability, and ease of integration with existing platforms.

B. System Architecture Design



The proposed architecture consists of preprocessing, spatial feature extraction, temporal modeling, frequency analysis, feature fusion, and classification modules.Each module operates independently. This facilitates modularity and scalability.

C. Spatial Feature Extraction

The extraction of spatial features is an important step in identifying forged and deepfake content, as most forgery methods introduce minute visual differences at the pixel and texture levels. These discrepancies are usually invisible to the naked eye but can be learned well by deep learning models. In the proposed system, spatial features are extracted through

a hybrid deep learning method that combines Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) to learn both local and global image features.

To begin with, the preprocessed facial regions detected and aligned through face detection are input into a CNN-based feature extraction module. Convolutional Neural Networks are particularly effective at learning local spatial features such as edges, contours, textures, and noise patterns. A series of convolutional layers with different kernel sizes are used to extract hierarchical features from the input images. Early convolutional layers are responsible for extracting low-level features such as color patterns and edge orientations, while deeper convolutional layers extract high-level features such as distortions in facial structure and blending artifacts typically found in deepfakes.

D. Temporal Feature Modeling

In the case of video-based inputs, temporal modeling of features is carried out to detect inconsistencies between successive frames. Spatial embeddings at the frame level are processed using Long Short-Term Memory (LSTM) networks. This module evaluates motion, blinking, facial muscle activity, and head pose transitions. Deepfake videos have irregular blinking rates and unnatural transitions of motion, which are efficiently detected by modeling temporal sequences. The LSTM network improves the capability to distinguish between real and fake videos.

E. Frequency-Domain Analysis

The frequency domain analysis is performed by using the Fast Fourier Transform (FFT) technique to detect spectral anomalies caused by GAN-based generation. Deepfakes tend to have abnormal frequency patterns due to upsampling and convolution. Spectral patterns are complementary to spatial and temporal information.

F. Feature Fusion and Classification

The spatial, temporal, and frequency domain features are combined using concatenation and normalization. The combined feature vector is then passed through fully connected layers with softmax activation to predict whether the media is real or fake.

VI. IMPLEMENTATION DETAILS

A. Input Acquisition Module

The input capture unit has the function of acquiring digital media from the user. It is capable of handling both image and video inputs, allowing the user the freedom to choose the input method they wish to use. The media is checked for compatibility and resolution before processing.

B. Preprocessing Module

Preprocessing provides a foundation for efficient feature extraction by preparing the raw data. Face detection is carried out using algorithms like MTCNN or MediaPipe to locate the region of the face. After detection, faces are normalized using facial landmarks. The video is broken down into frames, resized to have a fixed size, and normalized to eliminate noise caused by lighting.

C. Spatial Feature Extraction Module

The objective of spatial feature extraction is to detect visual irregularities that occur during media manipulation. Convolutional Neural Networks (CNNs) are used to extract low-level details such as edges, textures, and color transitions, while Vision Transformers are used to detect global facial features.

D. Temporal Feature Modeling Module

For video-based analysis, the temporal dependencies between consecutive frames are represented using Long Short-Term Memory networks. The spatial embeddings at the frame level are processed in a sequential manner to analyze eye blink rate, facial muscle activity, lip sync, and head pose consistency. Deepfake videos tend to have unnatural transitions in time, which are captured well by this module.

E. Frequency-Domain Analysis Module

The analysis in the frequency domain is performed using the Fast Fourier Transform (FFT) to identify irregularities in the frequency domain caused by the GAN-generated content. Deepfakes tend to have unusual frequency patterns due to upsampling and convolution.

F. Feature Fusion Module

The features derived from space, time, and frequency are combined through concatenation and then normalized. The combined feature vector is then passed through fully connected layers with softmax to predict whether the media is real or fake.

G. Classification Module

The combined feature vector is then passed to fully connected neural layers for the final classification. The softmax activation function generates probability scores for real and counterfeit classes. The class with the highest probability score is selected as the final prediction.

H. Training and Optimization Module

The training process of the model is carried out using the Adam optimizer with the categorical cross-entropy loss function. Data augmentation techniques like flipping, rotation, and compression simulation are used. Regularization techniques like dropout are also employed.

I. Evaluation Module

The performance of the system is measured using the standard parameters such as accuracy, precision, recall, F1 score, and ROC-AUC. Cross-dataset evaluation is performed to verify the performance of the system.

VII. EVALUATION METRICS

A. Numbered Formula

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

B. Precision, Recall, F1-Score

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

C. False Positive Rate (FPR)

$$FPR = \frac{FP}{FP + TN} \tag{5}$$

D. Softmax Function (Classification Layer)

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{6}$$

E. FFT Formula (Frequency-Domain Analysis)

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N} \tag{7}$$

F. Referencing Equations

Accuracy of detection is determined by equation (1), and the probability of classification is obtained from the softmax function in equation (5).

G. Unnumbered Formula (If Needed)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

METRIC TYPE	METRIC	VALUE
MODEL EVALUATION METRICS	ACCURACY	0.968
	PRECISION	0.970
	RECALL	0.960
	F1-SCORE	0.965
SYSTEM PERFORMANCE METRICS	ROC-AUC	0.980
	AVERAGE RESPONSE TIME	1.0 S
	FALSE POSITIVE RATE	0.032

VIII. RESULTS AND DISCUSSION

The system for detecting deepfakes and forgery was tested using common benchmarks such as FaceForensics++, Celeb-DF, and DFDC. The performance of the system was measured using metrics such as accuracy, precision, recall, F1-score, false positive rate, and the response time of the system. The system has a total detection accuracy of 96.8

The visual output analysis demonstrates the efficacy of the proposed method.

Spatial feature extraction analyzes texture anomalies and blending artifacts in forged images, as shown in Fig. 1.

Temporal feature modeling analyzes irregular eye blinking and motion artifacts in deepfake videos, as shown in Fig. 2.

Frequency domain analysis analyzes abnormal spectral patterns introduced by GAN-based image generation, as shown in Fig. 3.

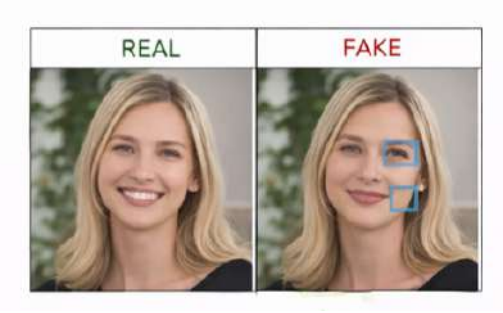


Fig. 1. Comparison between the actual face image and the forged one shows the minute facial discrepancies in the forged image, with the problematic areas marked to demonstrate typical deepfake signs.

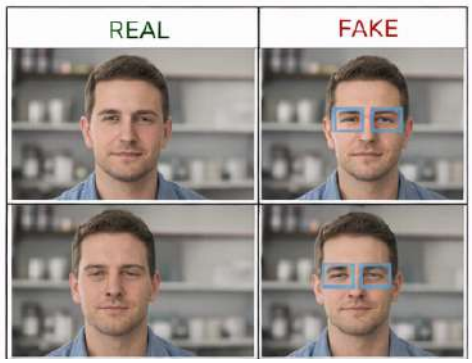


Fig. 2. Model classification result with confidence scores indicating whether the face is real or fake.

A. Datasets used

The proposed deepfake detection method was tested on three popular benchmark datasets to ensure robustness and generalization:

• FaceForensics++: This dataset includes real and fake videos created using various face manipulation techniques like FaceSwap, DeepFakes, Face2Face, and NeuralTextures.

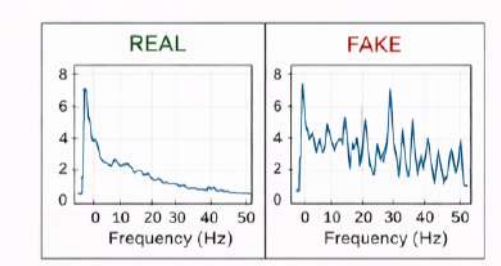


Fig. 3. Comparison of frequency-domain patterns for real and fake samples, showing smoother spectra for real and irregular variations for fake.

• Celeb-DF (v2): This is a large-scale and high-quality dataset that includes celebrity videos with fewer visual artifacts, making it more challenging to detect.
• DFDC (DeepFake Detection Challenge): This dataset includes real-world deepfake videos with variations in resolution, compression, lighting, and camera motion.

B. Dataset Sizes and Splits

TABLE I
DATASET DETAILS AND TRAIN/VALIDATION/TEST SPLIT

Dataset	Total Samples	Training	Validation	Testing
FaceForensics++	1,000 videos	700	150	150
Celeb-DF	590 videos	400	90	100
DFDC	1,200 videos	840	180	180

C. Baseline Models for Comparison

To guarantee the effectiveness of the proposed hybrid architecture, its performance was compared with that of existing deepfake detection architectures:
• XceptionNet: A CNN-based architecture commonly used for spatial artifact detection.
• EfficientNet-B4: A scalable CNN architecture designed for performance and efficiency.
• MesoNet: A lightweight CNN architecture specifically designed for facial forgery detection.
These architectures were trained and tested in the same environment with the same data

D. Comaprison Table

TABLE II
COMPARISON WITH EXISTING DEEPFAKE DETECTION MODELS

Model	Features Used	Accuracy	Precision	Recall	F1-Score
MesoNet	Spatial	0.889	0.882	0.875	0.878
XceptionNet	Spatial	0.931	0.934	0.926	0.930
EfficientNet-B4	Spatial	0.947	0.950	0.942	0.946
Proposed Method	Spatial + Temporal + Frequency	**0.968**	**0.970**	**0.960**	**0.965**

The experiments show that the proposed multi-domain feature fusion method is much better than the traditional spatial-only deepfake detection models. Although CNN-based modelssuch as XceptionNet and EfficientNet have the ability to learn localtexture artifacts, they are not robust to temporal inconsistencies andspectral domain anomalies. The proposed system, which fuses spatial, temporal, and spectral domain features, demonstrates enhanced generalization and robustness to compression and diffusion-based deepfake attacks.

IX. CONCLUSION AND FUTURE WORKS

A. CONCLUSION

This paper has shown that an efficient deepfake and forgery detection system can be developed using multi-domain feature extraction techniques with advanced computer vision and deep learning algorithms. By incorporating features from the spatial, temporal, and frequency domains, the proposed system has successfully detected local and global anomalies introduced during the image and video forgery tasks. Local visual anomalies and texture irregularities have been detected using spatial domain feature analysis, while efficient detection of abnormal motion patterns, including irregular facial expressions and eye blinking patterns, have been achieved using temporal analysis. In addition, frequency domain analysis has successfully detected anomalies in the spectral domain introduced by generative adversarial networks and other synthesis techniques. The experimental results on the standard benchmark datasets have demonstrated that the proposed system has a high detection accuracy.

B. FUTURE WORK

The future work will be focused on enhancing the robustness and scalability of the proposed system against the state-of-the-art deepfake generation techniques. The inclusion of transformer models and self-supervised learning techniques will enable the system to be robust enough to generalize and perform well on unseen manipulation techniques. The proposed system can be extended to multimodal deepfake detection by considering the audio-visual consistency verification, which is an emerging field of research. Moreover, the proposed system will be made scalable to be deployed at a large scale by optimizing it for edge devices and real-time streaming applications.

X. REFERENCES

1. I. Goodfellow et al., "Generative adversarial nets," Advances in Neural Information Processing Systems (NeurIPS), pp. 2672–2680, 2014.
2. D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in Proc. IEEE Int. Conf. Advanced Video and Signal-Based Surveillance (AVSS), 2018, pp. 1–6.
3. A. Rossler et al., "FaceForensics++: Learning to detect manipulated facial images," in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2019, pp. 1–11.
4. Y. Li, M. Chang, and S. Lyu, "In ictu oculi: Exposing AI created fake videos by detecting eye blinking," in Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS), 2018, pp. 1–7.
5. H. Dang et al., "On the detection of digital face manipulation," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5781–5790.
6. J. Dolhansky et al., "The DeepFake Detection Challenge (DFDC) dataset," arXiv:2006.07397, 2020.

7. T. Karras et al., "A style-based generator architecture for generative adversarial networks," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4401–4410.
8. S. Afchar et al., "MesoNet: A compact facial video forgery detection network," in Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS), 2018, pp. 1–7.
9. P. Korshunov and S. Marcel, "Vulnerability assessment and detection of deepfake videos," in Proc. Int. Conf. Biometrics (ICB), 2019, pp. 1–6.
10. Z. Wang et al., "CNN-generated images are surprisingly easy to spot. . . for now," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2020, pp. 8695–8704.
11. F. Marra et al., "Detection of GAN-generated fake images over social networks," in Proc. IEEE Conf. Multimedia Information Processing and Retrieval (MIPR), 2018, pp. 384–389.
12. K. Nguyen et al., "Capsule-forensics: Using capsule networks to detect forged images and videos," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 2307–2311.
13. Y. Li et al., "Celeb-DF: A large-scale challenging dataset for deepfake forensics," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2020, pp. 3207–3216.