WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

DeepFake Video Detection-A Survey

Hiranmayee Panda[1], Vyda Varshita[2], Rayavarapu Siri Pranathi [3], Nadukuditi Sai Sridhar[4]

[1]Hiranmayee Panda, Information Technology, GMRIT, Rajam, India
[2]Vyda Varshita, Information Technology, GMRIT, Rajam, India
[3]Rayavarapu Siri Pranathi, Information Technology, GMRIT, Rajam, India
[4]Nadukuditi Sai Sridhar Information Technology, GMRIT, Rajam, India

ABSTRACT -

Deepfake technology, which leverages deep learning to manipulate facial features and expressions in videos, has emerged as a major concern for digital media integrity, privacy, and security. As these forgeries grow increasingly convincing, there is a critical need for reliable detection mechanisms. This survey paper reviews and analyzes existing research focused on the use of lightweight convolutional neural network (CNN) architectures—MobileNet and ResNet—in the domain of deepfake detection. By examining previous studies, we highlight the strengths, limitations, and application areas of each model. MobileNet is frequently chosen for its computational efficiency and suitability for deployment on mobile and edge devices, while ResNet's deep residual learning structure enables higher accuracy in classifying real and fake content. The analysis includes performance comparisons based on benchmark datasets such as FaceForensics++ and DFDC, emphasizing the trade-offs between model complexity, accuracy, and inference speed. This paper aims to provide a comprehensive understanding of how these architectures have been applied in past research, helping guide future work in selecting appropriate models based on hardware constraints and application needs. The survey concludes by identifying gaps in the literature and suggesting directions for further study to improve the robustness and scalability of deepfake detection systems.

Key Words: *Deepfake detection, MobileNet, ResNet, convolutional neural networks, media forensics, FaceForensics++, DFDC, lightweight models, video manipulation, deep learning.*

1. INTRODUCTION

The rapid advancement of deep learning techniques has led to significant breakthroughs in computer vision, including facial recognition, image synthesis, and video manipulation. Among these, deepfake technology has gained considerable attention for its ability to generate highly realistic fake videos by altering facial features and expressions. While this innovation has potential applications in entertainment, gaming, and accessibility, it also poses severe threats to digital media integrity, privacy, and public trust. Deepfakes have been increasingly used to spread misinformation, commit fraud, and manipulate public opinion, making it essential to develop reliable detection mechanisms. The realism of modern deepfakes makes them challenging to identify with the naked eye, thus demanding the use of automated, intelligent systems that can accurately distinguish between genuine and manipulated content. As a result, researchers have turned to convolutional neural networks (CNNs), which have demonstrated impressive capabilities in image and video classification tasks.

This survey paper focuses on two widely adopted CNN architectures—MobileNet and ResNet—that have been explored in various studies for deepfake detection. MobileNet is known for its lightweight structure and efficient computation, making it ideal for real-time detection on mobile and edge devices. In contrast, ResNet introduces residual learning to train deeper networks effectively, often achieving higher accuracy in complex classification problems. We analyze previous research that applies these architectures to benchmark datasets such as FaceForensics++ and the Deepfake Detection Challenge (DFDC), highlighting their performance, limitations, and application contexts. Through this analysis, we aim to provide a comprehensive understanding of how these models perform in practical deepfake detection scenarios. The goal is to help researchers and practitioners choose appropriate models based on factors such as available hardware, required accuracy, and real-time performance needs. By identifying trends, challenges, and gaps in the existing literature, this survey also outlines potential directions for future research aimed at improving the scalability, robustness, and generalization of deepfake detection systems.

2. LITERATURE SURVEY

A comprehensive survey on deepfake detection provides an in-depth review of existing methods, highlighting their strengths, limitations, and future directions. The study explores datasets such as FaceForensics++, DeepFake-TIMIT, and Celeb-DF, which contain manipulated videos created using

various deepfake generation techniques. The methodology covers deep learning-based approaches, including CNNs, RNNs, and hybrid models that leverage spatial and temporal features, as well as frequency domain analysis to detect inconsistencies in Fourier or wavelet transforms. Additionally, biological signal-based methods utilize physiological signals like heart rate and blinking patterns, while forensic analysis examines digital artifacts such as compression traces and noise patterns to identify manipulation. The survey offers a broad overview of the field, making it a valuable resource for researchers by identifying gaps and trends in deepfake detection. However, it does not propose a new methodology or detection technique and is limited to summarizing existing approaches without providing experimental results or comparisons. Furthermore, the rapid advancements in deepfake technology may render some parts of the survey outdated over time. [1].

A deepfake detection approach leverages a dual-stream deep neural network to identify subtle artifacts and inconsistencies in manipulated images and videos. The study utilizes Celeb-DF and FaceForensics++ datasets, which contain high-quality deepfake videos created using various generation techniques. The methodology integrates content-based features, extracted from pixel-level information such as unnatural textures and facial inconsistencies, with trace-based features that capture artifacts introduced during deepfake generation, such as noise patterns and compression artifacts. These two feature streams are processed independently and fused using a feature fusion module before final classification through a fully connected layer. This approach achieves high accuracy by capturing both visual and subtle artifacts, making it robust against different types of deepfakes, including high-quality manipulations. However, the dual-stream architecture is computationally expensive, requiring significant processing power and large datasets for effective training. Additionally, the model may struggle with advanced deepfakes that exhibit minimal visual artifacts. [2].

This study explores unsupervised learning techniques for deepfake video detection, eliminating the need for labeled training data. The approach leverages autoencoders and clustering algorithms to identify deepfake anomalies. The model learns patterns from real videos and detects deviations in manipulated ones. It is particularly useful for detecting unseen deepfake variations. The authors highlight the scalability and adaptability of this method. However, unsupervised techniques generally struggle with high-quality deepfakes. The accuracy of detection is lower compared to supervised models. The study suggests improving feature extraction methods to enhance detection performance. The paper also discusses the importance of leveraging multiple modalities for improved deepfake detection. Future directions include combining unsupervised learning with self-supervised techniques [3].

This review focuses on deepfake techniques related to face-swapping and expression swapping. The study provides an overview of generative adversarial networks (GANs) and autoencoders used for generating manipulated faces. It highlights the challenges in detecting deepfakes that alter facial expressions naturally. The paper categorizes different deepfake techniques based on their methodologies and applications. The study emphasizes the need for robust detection techniques that can handle subtle modifications. One key limitation is that it primarily covers face and expression swaps, excluding other deepfake manipulation techniques. The review discusses the importance of temporal inconsistencies in detection. Future work suggests the development of hybrid approaches that combine multiple detection methods for better accuracy [4].

This paper introduces a multi-modal framework for deepfake detection, integrating visual analysis, metadata verification, and blockchain-based authentication. The study emphasizes the role of blockchain in ensuring tamper-proof verification of media content. By combining multiple detection mechanisms, the framework enhances the reliability of deepfake identification. The authors discuss how metadata can help track content authenticity. However, a key limitation is the reliance on metadata, which may not always be available or reliable. The integration of blockchain increases computational complexity and costs. Despite these challenges, the approach provides a comprehensive strategy for combating deepfakes. The study suggests further research into reducing the computational overhead of blockchain-based verification. Future work may explore AI-driven metadata analysis for more efficient detection. [5].

Deepfake image detection is crucial due to the increasing risks associated with AI-generated deceptive content, which can manipulate public opinion and spread misinformation. A study explores an optimized dense CNN model for recognizing deepfake images, utilizing the Deepfake Detection Challenge (DFDC) dataset to ensure a balanced representation of real and fake content. The methodology involves training CNN architectures like DenseNet and VGG, complemented by Multi-task Cascaded Convolutional Networks (MTCNN) for face detection and feature extraction. The model demonstrates high accuracy, with MTCNN excelling in precision and F1-score, effectively identifying subtle inconsistencies in deepfake images. However, challenges remain, including the need for continuous updates as deepfake techniques evolve and the limited generalizability of the model due to dataset constraints. Additionally, training deep CNN models requires significant computational resources, making large-scale implementation costly and resource-intensive [6].

Self-supervised learning using Vision Transformers (ViTs) has emerged as an effective approach for deepfake detection by leveraging robust feature representations without relying on extensive labeled data. A study explores the use of DINO-based ViTs to extract meaningful self-supervised features for identifying deepfake faces, enhancing detection capabilities across diverse forgery techniques. The model is trained on the DFDC dataset, which includes variations in lighting, resolution, gender, age, and skin tone, ensuring improved generalization. A total of 205,589 images from 104,498 videos were used for training, with 42,108 images reserved for validation. Multiple models, including Efficient ViT and Convolutional Cross ViT, were trained and evaluated using AUC and F1-score metrics. The findings indicate that self-supervised features significantly improve performance, achieving higher AUC and F1-scores compared to baseline methods using only RGB images. However, while the method benefits from diverse training data, the complexity of ViTs with multiple attention heads does not always lead to performance gains. Additionally, the approach remains computationally intensive, highly data-dependent, and susceptible to adversarial attacks despite its improved accuracy in deepfake detection [7].

Deepfake detection has become critical due to advancements in generative models like VAEs and GANs, which create highly realistic manipulations. A major challenge is acquiring large datasets to train robust models, especially as deepfake techniques evolve. Continuously retraining models adds complexity, highlighting the need for generalizable approaches. Studies compare Vision Transformers (ViTs) and CNNs like EfficientNet for detection.

EfficientNetV2 excels at recognizing known anomalies but struggles with novel techniques. In contrast, ViTs generalize better due to self-attention mechanisms that capture broader contextual relationships. Large datasets like DFDC and ForgeryNet enhance model performance, with ForgeryNet offering diverse deepfake techniques. Research shows EfficientNet performs well on familiar manipulations but lacks adaptability. Performance evaluation focuses on accuracy and variance across different deepfake methods [8].

The paper Mastering Deepfake Detection provides an extensive review of deepfake creation and detection techniques, emphasizing GANs and Diffusion Models (DMs) as primary generative architectures. It explores deepfake creation methods and reviews detection techniques for both GAN and DM-based images, highlighting the evolution of detection methods as generative AI advances. A key contribution is a hierarchical multi-level detection approach using a dataset of 83,000 real and synthetic images. The method classifies images at three levels: distinguishing real from fake, identifying the generative model, and recognizing specific architectures. This approach achieves over 97% accuracy, surpassing state-of-the-art models. It demonstrates robustness against distortions like JPEG compression and resizing, making it practical for forensic applications. The study emphasizes the need for improved deepfake detection techniques and discusses challenges in detecting highly realistic AI-generated content. The paper highlights future research directions to enhance detection accuracy and adaptability. [9].

The paper Deepfake Generation and Detection: Case Study and Challenges provides a comprehensive survey of deepfake technologies, covering both generation and detection methods. It reviews state-of-the-art surveys that focus on deep learning (DL) and machine learning (ML) models for deepfake detection. While many surveys discuss model capabilities and dataset challenges, they often lack depth in specific areas. A major gap identified is the limited coverage of audio deepfake detection. Comparative analysis shows that some surveys focus on facial manipulations, while broader studies still overlook audio-based detection. The paper introduces a taxonomy categorizing detection techniques into image, video, and audio modalities. Existing surveys often fail to explore all three modalities in depth. To ensure a rigorous analysis, the authors conducted a systematic review of 111 articles from IEEE Xplore, ACM, Springer, and ScienceDirect. The selection process involved filtering articles using specific keywords. The study focuses on literature published between 2019 and 2023, reflecting the latest advancements in deepfake technologies. [10]

3. COMPARISION TABLE

Ref. No.	Title	Author Names	Dataset Details	Methodology	Advantages	Disadvantages
[1]	Exposing Fake Faces Through Deep Neural Networks Combining Content and Trace Feature Extractors	E. Kim and S. Cho	Celeb-DF, Face Forensics++	Combines content-based features (e.g., pixel-level artifacts) and trace-based features using a dual-stream deep neural network. The model fuses these features for classification.	High accuracy due to the combination of content and trace features. Robust against various types of deepfakes.	Computationally expensive due to dual-stream architecture. Requires large datasets for training.
[2]	DeepFake Detection for Human Face Images and Videos: A Survey	A. Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan	FaceForensics++, DeepFake-TIMIT, Celeb-DF	Provides a comprehensive review of existing deepfake detection methods, including deep learning-based approaches, frequency domain analysis, and biological signal-based methods.	Offers a broad overview of the field. Useful for researchers to identify gaps and trends.	Does not propose a new methodology. Limited to summarizing existing techniques.
[3]	Unsupervised Learning-Based Framework for Deepfake Video Detection	Li Zhang; Tong Qiao; Ming Xu; Ning Zheng; Shichuang Xie	DFDC (DeepFake Detection Challenge), FaceForensics++	Uses unsupervised learning techniques such as autoencoders and clustering algorithms to detect deepfakes.	Does not require labeled data for training. Effective for detecting unseen deepfake types.	Lower accuracy compared to supervised methods. May struggle with high-quality

[4]	DeepFake on Face and Expression Swap: A Review	Waseem, S., Bakar, S. A. R. S. A., Ahmed, B. A., Omar, Z., Eisa, T. A. E., & Dalam, M. E. E.	FaceSwap, DeepFake-TIMIT, Celeb-DF	Reviews face-swapping and expression-swappingtechniques, focusing on generative adversarial networks (GANs) and autoencoders. Discusses detection methods specific to these types of deepfakes.	Provides insights into specific deepfake types. Highlights detection challenges for face and expression swaps.	Limited to face and expression swaps. Does not propose new detection methods.
[5]	Real, Forged or Deep Fake? Enabling the Ground Truth on the Internet	Mohammad A. Hoque Md Sadek Ferdous et al	FaceForensics++, Celeb-DF, DFDC (DeepFake Detection Challenge)	The paper proposes a multi-modal framework that combines visual analysis, metadata verification, and blockchain-based authentication to verify the authenticity of media.	Provides a comprehensive approach by combining multiple verification methods. Blockchain ensures tamper-proof authentication.	Relies heavily on metadata, which may not always be available or reliable. Blockchain integration increases computational complexity.
[6]	A Survey an Optimized Dense CNN Model for Recognizing Deepfake Images	Mallikarjun Gachchannavar1, Dr. Naveenkumar J.R.2, Radha Velangi3	DFDC	The study uses deep learning models like CNNs and MTCNN for face detection and feature extraction.	The optimized dense CNN model improves deepfake detection by leveraging deep learning techniques.	Computationally Expensive – Training deep CNN models requires significant computational power and GPU resources
[7]	Realistic Facial Deep Fakes Detection Through Self-Supervised Features Generated by a Self-Distilled Vision Transformer	Jose Boaro Sergio Colcher Bruno Rocha Gomes Antonio J G Busson	FaceForensics++ Celeb-DF DF-TIMIT	The study utilizes self-supervised Vision Transformers (DINO) to extract self-attention activation maps, which are combined with CNN architectures for enhanced deep fake detection.	The approach improves deep fake detection accuracy by leveraging self-supervised learning for better feature extraction and generalization.	The method is computationally expensive, data-dependent, and vulnerable to adversarial attacks despite improved accuracy.
[8]	Cross-Forgery Analysis of Vision Transformers and CNNs for Deepfake Image Detection	Davide Alessandro Coccomini Roberto Caldelli Fabrizio Falchi Claudio Gennaro	ForgeryNet	The study evaluates the effectiveness of Vision Transformers (ViTs) and EfficientNetV2 in deepfake detection by comparing their generalization capabilities.	Vision Transformers excel in generalization, making them more effective at detecting novel deepfake techniques.	Vision Transformers are computationally expensive, while EfficientNetV2 struggles with generalization to new deepfake methods.

		Giuseppe Amato				
[9]	Mastering Deepfake Detection: A Cutting-edge Approach to Distinguish GAN and Diffusion-model Images	Luca Guarnera, Oliver Giudice, Sebastian O battato,	Dataset consisting of 83,000 images generated by nine GAN models and four Diffusion Models	The paper introduces a hierarchical multi-level approach to deepfake detection	The method achieves over 97% accuracy.	High complexity, resource-heavy, poor generalization.
[10]	Deepfake Generation and Detection: Case Study and Challenge	Yogesh Patel, Rajesh Gupta, et al	DFDC, FaceForensics++, VoxCeleb, and Celeb-DF	The paper surveys advancements in deepfake technology, analyzing generation and detection models across images, videos, and audio.	It provides a comprehensive analysis of deepfake generation and detection across multiple modalities, identifies key challenges in detection, and offers valuable future research directions.	The paper primarily identifies challenges without offering concrete solutions, relies on existing datasets that may not represent future deepfake techniques

4. CONCLUSION

Deepfake technology presents a growing challenge to media authenticity and digital security, necessitating robust and efficient detection methods. This survey reviewed recent studies focusing on lightweight convolutional neural networks such as MobileNet and ResNet, which have been widely used for detecting manipulated video content. MobileNet is favored for its low computational cost and suitability for mobile and edge devices, while ResNet achieves higher detection accuracy through its deeper architecture and residual learning capabilities. Comparative analyses on benchmark datasets like FaceForensics++ and DFDC reveal a trade-off between accuracy and efficiency, highlighting the need to balance model performance with deployment constraints. Beyond CNNs, emerging methods such as Vision Transformers, self-supervised learning models, and multi-modal frameworks offer improved generalization and detection capabilities, though often at the expense of increased computational complexity. The survey also identified ongoing challenges, including the need for larger and more diverse datasets, better generalization to unseen deepfake techniques, and resistance to adversarial attacks. Moving forward, research should prioritize the development of lightweight yet accurate models and hybrid approaches that integrate multiple data sources. Ultimately, adapting detection systems to the rapidly evolving landscape of deepfake generation is essential to ensure trust and reliability in digital media environments.

REFERENCES

[1]. Kim, E., & Cho, S. (2021). Exposing fake faces through deep neural networks combining content and trace feature extractors. IEEE Access, 9, 123493-123503

[2]. Malik, A., Kuribayashi, M., Abdullahi, S. M., & Khan, A. N. (2022). DeepFake detection for human face images and videos: A survey. IEEE Access, 10, 18757-18775.

[3]. Zhang, L., Qiao, T., Xu, M., Zheng, N., & Xie, S. (2022). Unsupervised learning-based framework for deepfake video detection. IEEE Transactions on Multimedia, 25, 4785-4799.

[4]. Waseem, S., Bakar, S. A. R. S. A., Ahmed, B. A., Omar, Z., Eisa, T. A. E., & Dalam, M. E. E. (2023). DeepFake on face and expression swap: A review. IEEE Access, 11, 117865-117906.

[5].Hoque, M. A., Ferdous, M. S., Khan, M., & Tarkoma, S. (2021). Real, forged or deep fake? Enabling the ground truth on the internet. IEEE Access, 9, 160471-160484.

[6]. Gachchannavar, M., JR, N., & Velangi, R. (2024b). A survey an optimized dense CNN model for recognizing deepfake images. International Journal for Multidisciplinary Research, 6(4). https://doi.org/10.36948/ijfmr.2024.v06i04.26429

[7] Gomes, B. R., Busson, A. J., Boaro, J., & Colcher, S. (2023, October). Realistic Facial Deep Fakes Detection through Self-Supervised Features Generated by a Self-Distilled Vision Transformer. In Proceedings of the 29th Brazilian Symposium on Multimedia and the Web (pp. 177-183).

[8]. Coccomini, D. A., Caldelli, R., Falchi, F., Gennaro, C., & Amato, G. (2022, June). Cross-forgery analysis of vision transformers and cnns for deepfake image detection. In Proceedings of the 1st International Workshop on Multimedia AI against Disinformation (pp. 52-58)

[9]. Guarnera, L., Giudice, O., & Battiato, S. (2024). Mastering deepfake detection: A cutting-edge approach to distinguish gan and diffusion-model images. ACM Transactions on Multimedia Computing, Communications and Applications.

[10]. Patel, Y., Tanwar, S., Gupta, R., Bhattacharya, P., Davidson, I. E., Nyameko, R., ... & Vimal, V. (2023). Deepfake generation and detection: Case study and challenges. IEEE Access.