www.rungta.ac.in

RUNGTA
International Journal of
Computer Science and
Information Technology

REVIEW ARTICLE

A Literature Survey on Deepfake Detection

Richa Sharma[1]*, **Tripti Sharma**[2]
[1,2] *Rungta college of Engineering and Technology, Bhilai, India-490024*
*Corresponding Author: richa.sharma@rungta.org

ABSTRACT:
The quick development of deep learning-based synthetic media production has made deepfake detection an important research topic. False information by deepfakes causes serious threats to Security and privacy. It may alter audio, video, and images to produce incredibly lifelike fakes. There are many deepfake detection methods are examined in this literature review, which are divided into three categories" deep learning, watermarking, and blockchain". The difficulties with datasets, generalization, and the adversarial robustness of detection models are also covered. Many developments have been done such as multimodal detection frameworks, interpretable explainable AI, and the effect of generative adversarial networks (GANs) on detection effectiveness are highlighted in the study. Lastly, open challenges and future research areas are described to improve deepfake detection's scalability, performance, and practicality.

KEYWORDS: Deepfake detection, systematic literature review, Deep learning, Watermarking, Blockchain, Dataset.

1. Introduction

The term "Deepfake" is a portmanteau of "deep learning" and "fake." It refers to highly realistic synthetic media, primarily video and audio, created using artificial intelligence (AI) and machine learning (ML) algorithms. Since their emergence in late 2017, deepfakes have rapidly evolved from a niche internet phenomenon into a profound societal challenge. The underlying technology largely relies on Generative Adversarial Networks (GANs) [1] and powerful autoencoders, which can seamlessly swap a target person's face into a source video or synthesize an entirely artificial voice with remarkable fidelity [2].

While synthetic media has legitimate applications in entertainment, education, and accessibility, the malicious use of deepfakes poses severe threats to global security, personal privacy, and democratic integrity. Deepfakes have been weaponized to create non-consensual pornography, commit financial fraud via impersonation, and launch political disinformation campaigns aimed at swaying elections [3]. The sheer realism of modern deepfakes has eroded the traditional paradigm of "seeing is believing," leading to an urgent need for robust, reliable detection mechanisms.

The rapid development of deepfake generation techniques has catalyzed a corresponding surge in deepfake detection research [4]. Detecting these manipulated media files is essentially a high-stakes arms race between the creators (forgers) and the detectors (forensic analysts). Early deepfakes suffered from obvious visual artifacts—such as lack of eye blinking or poor blending around the face mask—which made them relatively easy to spot. However, modern generation algorithms have resolved these superficial flaws, necessitating the deployment of highly sophisticated detection strategies.

This literature survey systematically examines the current landscape of deepfake detection. We categorize the existing solutions into three primary paradigms: (1) Reactive Deep Learning-based detection, which analyzes the media *post facto* for forensic artifacts; (2) Proactive Watermarking, which embeds verifiable tags during media creation; and (3) Blockchain-based provenance tracking, which ensures a secure, immutable ledger of media origin.

Furthermore, this paper comprehensively reviews the benchmark datasets that form the backbone of detection research, discusses the critical challenge of model generalization, and highlights the emerging role of Explainable AI (XAI) in forensic analysis. Finally, we map out the open challenges

Received on 25.02.2026 Revised on 29.04.2026
Accepted on 10.05.2026 Published on 20.05.2026
Rungta International Journal of Computer Science and Information Technology. 2026; : 52.

and promising future directions required to transition deepfake detection from isolated laboratory experiments into scalable, practical, real-world defense systems.

2. Taxonomy of Deepfake Creation

To effectively detect deepfakes, one must understand the methodologies used to create them. The generation of deepfakes can generally be classified into four distinct categories based on their primary manipulation objective [5, 4].

1. **Face Swapping (Identity Swap):** This is the most common form of deepfake. It involves replacing the face of a person in a target video with the face of a source person. Popular implementations include DeepFaceLab and Faceswap, which utilize paired encoder-decoder architectures to learn the facial expressions of both individuals and superimpose them.

2. **Face Reenactment (Expression Swap):** Unlike swapping identities, reenactment transfers the facial expressions and head movements of a source actor onto a target individual while maintaining the target's identity. Techniques like Face2Face allow real-time manipulation of facial features, often used in puppeteering attacks.

3. **Entire Face Synthesis:** This involves the creation of entirely non-existent faces using GANs, most notably StyleGAN. These images are highly photorealistic but do not correspond to any real human being. They are frequently used to create fake social media profiles or "sock puppet" accounts for astroturfing.

4. **Audio/Voice Synthesis:** Often overlooked but equally dangerous, audio deepfakes (voice cloning) synthesize realistic speech using a few seconds of a target's audio sample. Architectures like WaveNet and Tacotron have been used to spoof voice authentication systems and orchestrate CEO fraud.

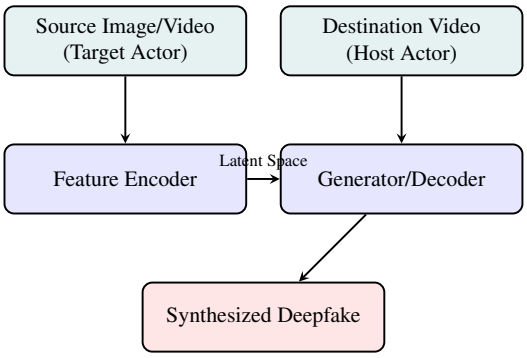

Figure 1: Generic pipeline for deepfake creation using an Encoder-Decoder architecture.

3. Deep Learning-Based Detection Methods

Deep learning currently dominates the field of deepfake detection. These methods frame detection as a binary classification problem (Real vs. Fake) and attempt to learn the subtle, latent traces left behind by the generation process. We categorize these approaches into spatial, frequency, and temporal analysis [6].

3.1. Spatial and Visual Artifact Analysis

Early detection systems focused on identifying spatial artifacts—inconsistencies within a single frame of a video. GANs and autoencoders often struggle to render high-frequency details perfectly, leaving behind convolutional traces or "fingerprints" [7].

Afchar et al. [8] introduced MesoNet, a compact Convolutional Neural Network (CNN) specifically designed to detect the microscopic properties of forged images, focusing on the mesoscopic level of detail rather than high-level semantic features.

Li et al. [9] proposed Face X-ray, which detects face swapping by searching for the blending boundary. Face X-ray operates on the assumption that manipulated faces are typically blended into the background image, creating an unnatural boundary where the internal face resolution differs slightly from the external image context. This method proved highly effective because it does not rely on artifacts from a specific GAN, granting it better generalization across unseen algorithms.

3.2. Frequency Domain Analysis

While deepfakes can look flawless in the RGB spatial domain, they often exhibit severe anomalies in the frequency domain. Generative models utilize up-sampling operations (like transposed convolutions) that inadvertently alter the spectral distribution of the image.

Qian et al. [10] demonstrated "Thinking in Frequency," introducing a Frequency in Face Forgery Network (F^3-Net). Their architecture extracts frequency-aware clues using Discrete Cosine Transform (DCT) and leverages cross-attention to fuse spatial and frequency domain features. This dual-domain approach significantly improves robustness against common video compression techniques, which often destroy spatial artifacts.

3.3. Temporal Analysis in Videos

A single frame might appear authentic, but analyzing a sequence of frames often reveals temporal inconsistencies. Deepfake generators typically synthesize videos frame-by-frame, failing to preserve strict temporal coherence—resulting in flickering or unnatural physiological signals.

Masi et al. [11] proposed a two-branch Recurrent Neural Network (RNN) that isolates deepfakes by examining both the raw visual data and the frequency components across the temporal dimension. By feeding sequences of frames into a Long Short-Term Memory (LSTM) network or Vision Transformer (ViT), the model learns the natural progression of facial movements.

Furthermore, some models analyze biological signals. Real human faces exhibit micro-color changes corresponding to heartbeat (remote photoplethysmography). Deepfakes rarely synthesize this underlying biological pulse accurately. Models tracking these temporal, biological inconsistencies have shown high accuracy, though they struggle with highly compressed or low-light videos.

4. Watermarking Techniques for Proactive Defense

While deep learning detection is reactive, attempting to catch fakes after they are released, **Watermarking** offers a proactive defense. Digital watermarking involves embedding a hidden, verifiable signal into the original media at the time of creation or capture [12].

If a video is suspected of being a deepfake, the absence or distortion of the watermark provides immediate forensic evidence of tampering.

4.1. Invisible Digital Watermarking

Traditional watermarking overlaid visible logos, but modern defense requires invisible, robust watermarks. These watermarks are mathematically embedded into the frequency coefficients (e.g., DWT or DCT domains) of the image. A successful watermark for deepfake defense must possess two specific properties:

1. **Invisibility:** It must not degrade the perceptual quality of the original media.

2. **Fragility or Semi-fragility:** Unlike copyright watermarks which must survive any modification, deepfake watermarks should ideally be destroyed or localized when passed through a GAN or autoencoder. If the watermark is broken in the facial region but intact in the background, it signals a localized face swap.

4.2. AI-Driven Watermarking

Recent research integrates deep learning into the watermarking process itself. Neural network-based encoders embed a latent payload into an image, and a corresponding neural decoder retrieves it. Qiao et al. [12] demonstrated that AI-driven watermarks can be specifically tailored to be highly sensitive to the up-sampling and blending operations inherent in deepfake generation, creating a targeted proactive defense layer.

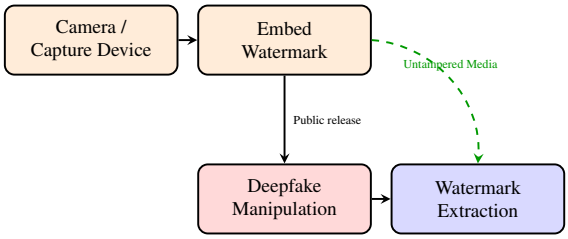

Figure 2: The proactive watermarking lifecycle. The extraction phase fails or indicates localized tampering if manipulated.

5. Blockchain-Based Provenance Tracking

Watermarking proves an image has been altered, but **Blockchain** technology proves where the image originated. By leveraging decentralized, immutable ledgers, researchers aim to create a chain of custody for digital media, combating deepfakes by ensuring traceability and data provenance [13].

5.1. Smart Contracts and Hash Verification

In a blockchain-based deepfake defense system, when an authoritative source (e.g., a news agency or government official) captures a video, a cryptographic hash of the file is generated. This hash, along with metadata (timestamp, location, device ID), is recorded as a transaction on a blockchain via a Smart Contract [14].

When an end-user views the video on a social media platform, the platform recalculates the hash of the video. If the video has been manipulated by a deepfake algorithm, even slightly, the cryptographic hash will completely change. The platform then queries the blockchain; if the hashes do not match, the video is flagged as untrusted or fabricated.

5.2. Challenges of Blockchain Integration

While conceptually powerful, blockchain faces significant practical hurdles:

- **Benign Modifications:** Social media platforms routinely compress, crop, and resize videos to save bandwidth. These benign operations change the cryptographic hash, causing false positives in the blockchain verification. To solve this, researchers are exploring ”perceptual hashing,” where minor, benign changes result in similar hashes, but drastic deepfake manipulations result in vastly different hashes.

- **Adoption Overhead:** This system requires widespread adoption. If a video is not registered on the blockchain at its inception, the system provides no defense against its manipulation.

6. Datasets for Deepfake Detection

The advancement of detection algorithms is fundamentally dependent on the quality, scale, and diversity of the datasets used for training. Early models achieved 99% accuracy on early datasets, only to fail completely on deepfakes "in the wild."

Table 1 summarizes the most critical benchmark datasets used in deepfake detection research.

The evolution of datasets reflects the arms race. FaceForensics++ [15] remains a staple for establishing baseline metrics, but models trained purely on FF++ suffer massive performance drops when evaluated on Celeb-DF [16] or the DFDC dataset [17]. This highlights that models tend to overfit to specific generation artifacts rather than learning the generalized concept of a "fake" face.

7. Challenges and Limitations of Current Methods

Despite significant algorithmic advancements, several severe limitations persist in the field of deepfake detection.

7.1. The Generalization Problem (Cross-Dataset Evaluation)

The most critical challenge is generalizability. A CNN model might achieve 98% Area Under the Curve (AUC) on the specific dataset it was trained on (e.g., FaceForensics++). However, when that exact same model is tested on deepfakes generated by an entirely different, unseen algorithm (e.g., testing on Celeb-DF), its accuracy frequently plummets to near random guessing ($\sim$50-60%) [18].

This "zero-day" deepfake problem occurs because neural networks lazily latch onto the specific statistical fingerprints of a known GAN rather than learning robust, universal forensic features. Resolving this requires self-supervised learning, few-shot learning techniques, and extracting inherently generalizable features like the blending boundaries utilized in Face X-ray [9].

7.2. Adversarial Attacks and Robustness

Detection models are neural networks, and like all neural networks, they are highly vulnerable to adversarial attacks. A forger can introduce imperceptible perturbations—adversarial noise—into the deepfake video. While the video looks identical to human eyes, the noise mathematically forces the detection CNN to misclassify the fake video as "Real" [19].

Furthermore, real-world videos on platforms like WhatsApp or Twitter undergo severe laundering: aggressive lossy compression, resizing, and frame-rate adjustments. These benign operations inadvertently act as adversarial attacks, destroying the subtle high-frequency artifacts that detectors rely on, rendering many state-of-the-art models ineffective in practical deployments.

7.3. Multimodal Deception

Current research heavily skews toward visual detection. However, modern deepfakes are multimodal threats—combining synthetic faces with perfectly cloned audio. A video where the lip-sync is generated using the target's cloned voice is significantly harder to detect if the detector only looks at spatial pixels. Analyzing the semantic alignment between the audio track and the visual lip movements (audio-visual synchrony) is a vital, yet underexplored frontier [20].

8. Recent Developments: Explainable AI and Multimodality

To address the aforementioned challenges, recent literature has shifted towards more sophisticated architectures.

8.1. Explainable AI (XAI) in Forensics

A black-box classification of "Fake: 95%" is insufficient for legal or journalistic use. Forensic analysts require interpretability—knowing *why* the model flagged the video. Explainable AI (XAI) integrates attention mechanisms into the detection pipeline. Models like Multi-Attentional Deepfake Detection [21] output heatmaps overlaid on the video, explicitly highlighting the manipulated regions (e.g., blurring around the jawline or unnatural reflections in the corneas). This interpretability not only builds human trust but allows researchers to understand exactly what artifacts the model has learned, aiding in diagnosing overfitting.

8.2. Multimodal Detection Frameworks

Recognizing that deepfakes manipulate reality across multiple senses, researchers are developing fusion frameworks. These models employ separate streams to analyze the audio waveform (e.g., using MFCC features), the visual frames, and the biological signals (pulse, eye blinking rates). The features are then concatenated into a joint representation space. If the video looks real, but the audio spectral distribution does not match the natural vocal tract of the speaker, the multimodal system will accurately flag the media.

9. Open Challenges and Future Research Directions

Based on the synthesis of the literature, several critical research directions emerge for the future of deepfake detection:

1. **Real-Time Edge Detection:** The majority of current models are computationally heavy, requiring massive

Table 1: Comparison of Benchmark Datasets for Deepfake Detection

Dataset	Size / Scale	Generation Methods Used	Key Characteristics & Limitations
UADFV (2018)	49 Real, 49 Fake videos	FakeApp	Low quality, early-generation artifacts. Highly saturated by modern detectors.
FaceForensics++ [15]	1,000 Real, 4,000 Fake videos	Deepfakes, Face2Face, FaceSwap, NeuralTextures	Industry standard benchmark. Provides multiple compression levels (RAW, HQ, LQ). Lacks diversity in ethnic backgrounds.
Celeb-DF [16]	590 Real, 5,639 Fake videos	Improved Deepfake algorithm	High visual quality with reduced blending artifacts. Significantly more challenging for cross-dataset generalization.
DFDC [17] (Deepfake Detection Challenge)	$\sim$100,000 videos	8 distinct unknown algorithms	Massive scale and highly diverse conditions (lighting, angles). Includes heavy perturbations and adversarial noise.
WildDeepfake (2021)	7,314 sequences	Collected directly from the internet	Represents real-world, "in-the-wild" deepfakes. Highly diverse but lacks controlled ground-truth environments.

GPUs. Future research must focus on model distillation and efficiency to enable real-time detection on edge devices (smartphones, browsers) to prevent the spread of deepfakes at the point of consumption.

2. **Federated Learning for Privacy:** As datasets contain massive amounts of biometric facial data, sharing them centrally poses privacy risks. Federated learning could allow organizations to collaboratively train deepfake detectors across distributed datasets without ever sharing the raw facial images.

3. **Continuous/Lifelong Learning:** The arms race never stops. Detectors must be designed with lifelong learning architectures, enabling them to continuously update their weights as new deepfake generation algorithms emerge, without catastrophically forgetting how to detect older fakes.

4. **Policy and Standardization:** Technical solutions alone are insufficient. There is a pressing need for international standards regarding digital watermarking embedded at the hardware level (cameras) and legal frameworks detailing the liability of platforms distributing synthetic media.

10. Ethical and Legal Implications of Deepfakes

The proliferation of deepfake technology has sparked intense debate regarding its ethical and legal ramifications. As synthetic media becomes democratized and easily accessible via open-source tools, the potential for mass misuse has escalated dramatically.

10.1. Consent and Privacy Rights

The most prevalent and damaging use of deepfake technology currently involves the non-consensual creation of synthetic pornography. This application disproportionately targets women and represents a severe violation of bodily autonomy and privacy. Legal frameworks globally are struggling to adapt to this threat, as traditional defamation or copyright laws are often insufficient or too slow to provide meaningful recourse for victims.

10.2. The "Liar's Dividend"

A less obvious but deeply corrosive societal impact of deepfakes is the phenomenon known as the "Liar's Dividend." As the public becomes increasingly aware that hyper-realistic fake videos exist, malicious actors can exploit this skepticism by claiming that genuine, incriminating evidence is actually a deepfake. In this scenario, the mere existence of deepfake technology provides a plausible deniability shield for perpetrators, undermining the evidentiary value of all digital media and accelerating the degradation of institutional trust.

10.3. Regulatory Responses

Governments are beginning to introduce legislation to combat deepfakes. Measures include mandating explicit labeling or watermarking of AI-generated content (e.g., the EU AI Act) and criminalizing the malicious distribution of synthetic media. However, enforcing these regulations across decentralized, global internet platforms remains an immense logistical challenge.

11. Emerging Threat Vectors: Live Deepfakes

While the majority of deepfake detection research focuses on pre-recorded, static video files (VOD), the threat landscape has recently shifted towards **Live Deepfakes**. By optimizing GANs and autoencoders for extreme low-latency inference, attackers can now perform high-fidelity face swapping and voice cloning in real-time during live video conferencing calls (e.g., Zoom, Microsoft Teams).

11.1. The Challenge of Real-Time Detection

Live deepfakes fundamentally break the current paradigm of forensic analysis.

- **Latency Constraints:** Traditional detection models that utilize complex temporal RNNs or process full video streams in the frequency domain require significant computation time (often seconds per frame). In a live call, adding even 200 milliseconds of latency to perform detection is unacceptable to the user experience.
- **Dynamic Compression:** Video conferencing software aggressively and dynamically alters resolution and bitrates based on network bandwidth. This constant fluctuation creates a noisy, unstable data stream that severely degrades the reliability of standard CNN-based detectors.

11.2. Active Interrogation Techniques

To detect live deepfakes, researchers are pivoting towards "Active Interrogation." Instead of passively analyzing the video feed, the detection system prompts the user to perform specific, unpredictable actions. For example, asking the user to rapidly turn their head in profile (a challenging angle for 2D face-swappers), wave a hand in front of their face (causing occlusion artifacts), or suddenly change lighting conditions. Generative models currently struggle to render these dynamic, complex interactions accurately in real-time, causing the deepfake mask to visually "slip" or glitch, thereby revealing the forgery to both human observers and automated monitoring systems.

12. Conclusion

Deepfakes represent a paradigm shift in the realm of digital media, transforming the theoretical risks of artificial intelligence into tangible threats to security, privacy, and truth. This literature survey systematically reviewed the state-of-the-art in deepfake detection, exploring the three main pillars of defense: reactive deep learning models, proactive invisible watermarking, and immutable blockchain provenance.

While deep learning techniques, particularly those analyzing frequency domain anomalies and temporal biological signals, have shown remarkable laboratory success, the literature reveals critical vulnerabilities. The persistent inability of models to generalize across unseen datasets and their extreme fragility to adversarial noise and standard video compression remain significant roadblocks to practical deployment.

The future of deepfake mitigation lies in adopting a "defense-in-depth" strategy. Relying solely on post-generation detection is a losing battle against the exponential improvement of GANs and diffusion models. A robust solution requires the integration of multimodal AI verification, mandated hardware-level watermarking, and decentralized blockchain registries to verify the origin of digital content.

Ultimately, combating deepfakes is not purely a technological problem; it requires a concerted, interdisciplinary effort combining advanced computer vision, cryptographic security, legislative action, and public media literacy. As synthetic media becomes increasingly indistinguishable from reality, the continued evolution of scalable, robust, and explainable detection frameworks is imperative to maintain the integrity of our shared digital reality.

12. References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[2] Yisroel Mirsky and Wenke Lee. The creation and detection of deepfakes: A survey. *ACM Computing Surveys (CSUR)*, 54(1):1–41, 2021.

[3] Shruti Agarwal, Hany Farid, Yuming Gu, Mingming He, Koki Nagano, and Hao Hao. Protecting world leaders against deep fakes. In *CVPR workshops*, volume 1, pages 38–45, 2019.

[4] Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, and Javier Ortega-Garcia. Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64:131–148, 2020.

[5] Thanh Thi Nguyen, Quoc Viet Hung Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, and Saeid Nahavandi. Deep learning for deepfakes creation and detection: A survey. *arXiv preprint arXiv:1909.11573*, 2019.

[6] Felix Juefei-Xu, Run Wang, Yihao Huang, Qing Guo, Lei Ma, and Yang Liu. Countering malicious deepfakes: Survey, battleground, and horizon. *International Journal of Computer Vision*, 130(7):1678–1734, 2022.

[7] Luca Guarnera, Oliver Giudice, and Sebastiano Battiato. Deepfake detection by analyzing convolutional traces. In *Proceedings of the IEEE/CVF conference on computer*

vision and pattern recognition workshops, pages 666–667, 2020.

[8] Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. Mesonet: a compact facial video forgery detection network. In *2018 IEEE international workshop on information forensics and security (WIFS)*, pages 1–7. IEEE, 2018.

[9] Lingzhi Li, Jianmin Bao, Zhang Ting, Hao Yang, Dongchen Chen, Fang Wen, and Baining Guo. Face x-ray for more general face forgery detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5001–5010, 2020.

[10] Yuyang Qian, Guojun Yin, Lu Sheng, Zixuan Chen, and Jing Shao. Thinking in frequency: Face forgery detection by mining frequency-aware clues. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*, pages 86–103. Springer, 2020.

[11] Iacopo Masi, Aditya Killekar, Royston Maria Mascarenhas, Shenoy Pratik Gurudatt, and Wael AbdAlmageed. Two-branch recurrent network for isolating deepfakes in videos. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*, pages 667–684. Springer, 2020.

[12] Rui Qiao et al. Watermarking for deepfake detection: A proactive defense. *IEEE Transactions on Information Forensics and Security*, 2021.

[13] Haya R Hasan and Khaled Salah. Combating deepfake videos using blockchain and smart contracts. *IEEE Access*, 7:41596–41606, 2019.

[14] Paula Fraga-Lamas et al. Blockchain and watermarking for deepfake detection. *IEEE Communications Surveys & Tutorials*, 2023.

[15] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1–11, 2019.

[16] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. Celeb-df: A large-scale challenging dataset for deepfake forensics. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3207–3216, 2020.

[17] Brian Dolhansky, Russ Howes, Ben Pflaum, Nicole Baram, and Cristian Canton Ferrer. The deepfake detection challenge (dfdc) preview dataset. *arXiv preprint arXiv:1910.08854*, 2020.

[18] Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A Efros. Cnn-generated images are surprisingly easy to spot... for now. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8695–8704, 2020.

[19] Pavel Korshunov and Sebastien Marcel. Deepfakes: a new threat to face recognition? assessment and detection. *arXiv preprint arXiv:1812.08685*, 2018.

[20] Abhijit Das et al. Towards solving the deepfake problem: An analysis on vision and audio multi-modal threat. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[21] Hanqing Zhao, Wenbo Zhou, Dongdong Chen, Tianyi Wei, Weiming Zhang, and Nenghai Yu. Multi-attentional deepfake detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2185–2194, 2021.