Global Journal of Research in
Science and Technology
Journal homepage: https://gsjournals.com/gjrst/
ISSN: 2980-4167 (Online)

(REVIEW ARTICLE)

Check for updates

GENERATIVE ADVERSARIAL NETWORKS AND DEEPFAKE FORENSICS: DETECTING SYNTHETIC AUDIO, VIDEO, AND DOCUMENT MANIPULATION

Marcus E. Vance [1,*] and Sarah L. Jenkins [2]

[1] Department of Computer Science and Digital Forensics, College of Information Technology, Husson University, Bangor, Maine, USA.
[2] Department of Cybersecurity and Information Assurance, School of Computing and Data Sciences, Dakota State University, Madison, South Dakota, USA.
* Corresponding Author

Global Journal of Research in Science and Technology, 2025, 03(03), 001–008

Publication history: Received on 04 June 2025; revised on 19 July 2025; accepted on 24 July 2025

Article DOI: https://doi.org/10.58175/gjrst.2025.3.3.0076

ABSTRACT

The proliferation of Generative Adversarial Networks (GANs) and related deep generative models has democratized the creation of hyper-realistic synthetic media—commonly known as deepfakes—posing unprecedented challenges to digital media authenticity, privacy, and public trust. This review provides a comprehensive and critical examination of the forensic landscape for detecting GAN-generated synthetic audio, video, and document manipulation. We systematically survey passive detection techniques that exploit artifacts inherent to the generation process, including frequency-domain anomalies, physiological inconsistencies (such as eye-blinking patterns and remote photoplethysmography signals), and Photo Response Non-Uniformity (PRNU) noise analysis. The adversarial arms race between forensic detection and anti-forensic generation is critically evaluated, highlighting how increasingly sophisticated GAN architectures continuously erode the reliability of existing detectors. We examine the legal admissibility of AI-generated forensic evidence, addressing the evidentiary challenges posed by deepfake media in civil and criminal proceedings under frameworks such as FRE 901(a). Standardized benchmarks—particularly FaceForensics++ and the Deepfake Detection Challenge (DFDC) dataset—are assessed for their role in driving reproducible research and enabling cross-method comparison. Our analysis reveals that while forensic detection methods have advanced substantially, the generalization gap across datasets and generation techniques remains a critical limitation. We identify unresolved questions regarding real-time deployment, cross-modality detection, and the integration of explainable AI for courtroom admissibility. Finally, we offer perspectives on emerging trends, including proactive fingerprinting, foundation models for forensics, and regulatory frameworks for synthetic media governance.

Keywords: Generative Adversarial Networks, Deepfake Detection, Forensic Analysis, Frequency-Domain Artifacts, PRNU Analysis, Faceforensics++, DFDC

1 INTRODUCTION

The advent of Generative Adversarial Networks (GANs), introduced by Goodfellow et al. (2014), has fundamentally transformed the landscape of digital media creation. GANs and their variants—including StyleGAN, StarGAN, and

autoencoder-based architectures—have enabled the generation of photo-realistic images, videos, and audio that are increasingly indistinguishable from authentic recordings to the human eye and ear. As noted by recent surveys, a photo-realistic image can be easily generated from a random vector, and images generated by advanced GANs are so realistic that even a well-trained viewer has difficulties distinguishing artificial from real images.

The societal implications of this technological capability are profound. Deepfakes—synthetic media created using deep learning techniques—have been deployed for malicious purposes ranging from political disinformation and identity impersonation to financial fraud and revenge porn. High-profile incidents include a fraud case where a CEO was induced via a deepfake-powered telephone call to transfer a large sum of money, and a family custody dispute where deepfake audio was submitted as evidence. The quality and performance of deepfake manipulation have reached a point where deepfake livestreaming is now possible.

The forensic detection of deepfakes has thus emerged as a critical research area within multimedia forensics and computer vision. Early detection methods focused on spatial artifact analysis, relying on inconsistencies in facial textures, blending artifacts, and abnormal pixel distributions. While effective against early-generation deepfakes, such methods demonstrated limited robustness when applied to newer synthesis techniques. The rapid evolution of generation methods has created an adversarial arms race between forensic detectors and generative models, where each advance in detection is met with corresponding improvements in generation quality and anti-forensic countermeasures.

This review provides a comprehensive and critical examination of the state-of-the-art in GAN-based deepfake forensics. We organize our analysis around three thematic pillars: (1) passive detection techniques targeting generation artifacts across frequency domains, physiological signals, and sensor noise patterns; (2) the adversarial arms race between forensic detection and anti-forensic generation; and (3) the legal and evidentiary challenges posed by deepfake media, including admissibility standards and the role of standardized benchmarks. Throughout, we identify persistent limitations, unresolved questions, and promising directions for future research.

2 GENERATIVE ADVERSARIAL NETWORKS AND DEEPFAKE FORENSICS

2.1 Passive Detection Techniques: Exposing Generation Artifacts

Passive forensic detection methods operate without embedding any prior information into the media content, instead relying on artifacts introduced during the generation or manipulation process. These artifacts manifest across multiple domains—spatial, frequency, temporal, and physiological—providing complementary signals for detection.

2.1.1 Frequency-Domain Artifacts

Despite the visual realism of GAN-generated images, hidden artifacts can be exposed in the frequency domain. The upsampling operations inherent to GAN architectures—used to progressively increase image resolution—leave discernible artifacts in the frequency spectrum. These artifacts arise from the interpolation and convolutional operations that are fundamental to the generation process, creating periodic patterns that are imperceptible in the spatial domain but detectable through frequency analysis.

Recent work has leveraged learnable frequency-domain filtering kernels and frequency-domain networks to thoroughly learn and extract these discriminative features. Sun et al. (2023) introduced a two-stream neural network that separates the human face area using landmarks before mining forgery patterns in the frequency domain. Their approach combines a single-frame pathway with a multi-frame pathway to mine frequency clues, demonstrating good results with limited training data on the FaceForensics++ and Celeb-DF datasets.

The effectiveness of frequency-domain detection stems from the fundamental differences between natural image statistics and the statistics of GAN-generated images. While GANs are trained to match the distribution of real images in the spatial domain, they do not perfectly replicate the higher-order statistical properties that characterize natural images—and these discrepancies are most apparent in the frequency domain.

2.1.2 Physiological and Biological Inconsistencies

Deepfake generation processes often fail to faithfully replicate the subtle physiological signals that characterize authentic human behavior. These inconsistencies provide a rich source of forensic evidence.

Eye-blinking patterns represent one of the earliest and most widely studied physiological cues for deepfake detection. Li et al. (2018) demonstrated that deepfake videos exhibit abnormal eye-blinking patterns compared to authentic videos. The assumption underlying this approach is that there is a difference in blinking patterns between authentic

video and deepfake-generated video, as GANs and autoencoders were not explicitly trained to model the natural frequency and duration of eye blinks. However, as noted by Patel et al. (2023), with development in deepfakes, these visual artifacts are becoming increasingly difficult to find, and current advanced deepfake videos are arduous to detect with visual features alone.

Remote photoplethysmography (rPPG) offers a more sophisticated physiological signal for deepfake detection. rPPG enables the extraction of heartbeat signals from video by analyzing subtle color variations in facial skin caused by blood volume changes. Deepfake generation processes typically do not model these physiological signals, resulting in inconsistencies between the rPPG signal extracted from synthetic videos and that expected from real human physiology.

Head pose and facial landmark inconsistencies have also been exploited for detection. Yang et al. (2019) attempted to use inconsistency between head pose and other body parts using facial landmarks. However, as with other visual features, these approaches become less reliable as generation techniques advance.

Biological features including eyebrow recognition, eye movement detection, ear and mouth detection have been systematically surveyed for deepfake detection. The premise is that generative models, being primarily optimized for visual plausibility rather than biological fidelity, introduce subtle anomalies in these features that can be detected through specialized forensic analysis.

2.1.3 Photo Response Non-Uniformity (PRNU) Analysis

Photo Response Non-Uniformity (PRNU) refers to the unique pixel-level noise pattern introduced by the manufacturing imperfections of digital camera sensors. This pattern serves as a fingerprint that can authenticate the provenance of digital images and videos.

PRNU analysis has been applied to deepfake detection based on the observation that GAN-generated images lack the characteristic sensor noise pattern of authentic camera-captured images. By analyzing the PRNU pattern through cross-correlation methods, forensic analysts can distinguish between images captured by a physical camera and those generated synthetically.

However, the application of PRNU analysis to deepfake detection faces several limitations. The PRNU signal is weak and can be easily corrupted by image compression, resizing, and other post-processing operations commonly applied to deepfakes. Furthermore, as noted in the literature, early PRNU studies used limited video samples (as few as 10 videos), raising questions about the generalizability of the approach.

2.1.4 Temporal Inconsistencies in Video Deepfakes

Video deepfakes introduce temporal inconsistencies arising from frame-to-frame manipulations, typically manifesting as unnatural facial movements. These artifacts stem from the difficulty of maintaining perfect temporal coherence when manipulating individual frames or sequences of frames.

Detection approaches exploiting temporal inconsistencies include analysis of lip movement synchronization, rPPG artifacts, and head pose inconsistency. Sun et al. (2023) observed that the forced mixing of manipulated faces in the generation process of deepfakes causes spatial distortions and temporal inconsistencies in crucial facial regions.

Video-level generative artifacts reflecting temporal discontinuities between frames have been extensively studied. Approaches range from frame-level voting (where each frame is classified independently) to more sophisticated spatio-temporal detectors using 3D CNNs, recurrent models, and transformers. Hybrid multimodal approaches that integrate spatial, temporal, and physiological features have shown particular promise for robust detection.

Table 1: Passive Detection Techniques for Deepfake Forensics: Artifacts, Methods, and Limitations

Artifact Category	Detection Method	Key Studies	Principle	Limitations
Frequency-Domain Artifacts	Two-stream neural network with frequency-domain filtering	Sun et al. (2023)	Upsampling operations in GANs leave periodic artifacts in frequency spectrum	Performance degrades with high-quality GANs; requires careful preprocessing
Physiological Inconsistencies	Eye-blinking pattern analysis	Li et al. (2018)	Deepfakes exhibit abnormal blinking frequency and duration	Visual artifacts become difficult to find with advanced deepfakes
Physiological Inconsistencies	Remote photoplethysmography (rPPG)	Anil et al. (2023)	Heartbeat signals from video are inconsistent in synthetic media	Susceptible to video compression and lighting variations
Physiological Inconsistencies	Head pose and facial landmark analysis	Yang et al. (2019)	Inconsistency between head pose and body parts	Visual features unreliable with advanced deepfakes
Sensor Noise Analysis	PRNU (Photo Response Non-Uniformity)	Various (2023)	GAN-generated images lack characteristic sensor noise fingerprint	PRNU signal weak; limited sample sizes in early studies
Temporal Inconsistencies	Frame-to-frame inconsistency detection	Sun et al. (2023)	Frame-to-frame manipulations create temporal discontinuities	Sophisticated generation methods increasingly maintain temporal coherence
Temporal Inconsistencies	Spatio-temporal detectors (3D CNNs, RNNs, Transformers)	Various	Video-level artifacts reflect temporal discontinuities	Computationally intensive; limited generalization

2.2 The Adversarial Arms Race: Forensics vs. Anti-Forensics

The relationship between deepfake generation and detection is fundamentally adversarial. As forensic detection methods improve, generation techniques evolve to evade detection, creating a continuous arms race.

2.2.1 Evolving Generation Techniques

Early deepfake generation relied on relatively simple autoencoder architectures that left obvious visual artifacts. The introduction of GANs, particularly StyleGAN and its successors, dramatically improved visual quality while simultaneously reducing the detectability of generation artifacts. More recent approaches have incorporated post-processing operations to circumvent color mismatch, temporal flickering, and inaccurate face masks.

The diversity of generation methods complicates forensic detection. As noted in the literature, existing surveys have mainly focused on the detection of deepfake images and videos, but comprehensive reviews covering the full spectrum of generation techniques are needed. The FaceForensics++ dataset, for example, includes videos generated using four different manipulation methods: FaceSwap, Deepfakes, Face2Face, and NeuralTextures. The DFDC dataset employs five different methods: Deepfake Autoencoder, MM/NN, NTH, FSGAN, and StyleGAN.

2.2.2 Adversarial Attacks on Forensic Detectors

The vulnerability of forensic detectors to adversarial attacks represents a critical concern. Lou et al. (2023) proposed a black-box attack method against GAN-generated image detectors using a contrastive learning strategy. Their approach applies imperceptible perturbation to input synthetic images to remove GAN fingerprints, effectively deceiving detectors. Extensive experimental results demonstrated that the proposed attack effectively reduces the accuracy of three state-of-the-art detectors on six popular GANs while maintaining high visual quality.

The emergence of such attacks highlights the fundamental insecurity of purely passive detection approaches. As Zhang et al. (2023) demonstrated, local perturbation generation methods can effectively evade GAN-generated face anti-forensics detection. The development of robust detectors that are resilient to adversarial perturbations remains an open challenge.

2.2.3 Proactive Defense Strategies

In response to the limitations of passive detection, proactive defense strategies have been proposed. These approaches embed identifiable fingerprints within generative models themselves, enhancing forensic analysis. Proactive defenses have explicitly targeted critical initial steps in GAN-based deepfake pipelines, such as face detection and identity encoding, to prevent forgeries at their earliest stages.

The proactive approach represents a paradigm shift from reactive detection to preventive authentication. By embedding forensic information at the point of generation, these methods enable reliable provenance verification regardless of subsequent post-processing or adversarial attacks.

2.3 Legal Admissibility and Standardized Benchmarks

The translation of deepfake forensic techniques from research to practice requires addressing both evidentiary standards and evaluation methodologies.

2.3.1 Legal Admissibility of AI-Generated Evidence

The admission of AI-generated or AI-detected evidence in legal proceedings raises novel evidentiary challenges. As noted in the forensic literature, deepfake media has recently started to appear as evidence in isolated cases, including a UK family custody dispute where deepfake audio was used as evidence. The growing accessibility and realism of deepfake media pose practical questions regarding authenticity, integrity, and provenance of media evidence.

In the United States, the admissibility of digital evidence is primarily governed by Federal Rule of Evidence (FRE) 901(a), which requires authentication through evidence sufficient to support a finding that the matter is what its proponent claims. The application of this standard to AI-generated evidence is complicated by the difficulty of establishing authenticity when the content itself may be synthetically generated.

Grossman et al. (2023) explored evidentiary issues that must be addressed by the bench and bar to determine whether actual or asserted GenAI output should be admitted as evidence in civil and criminal trials. Their analysis emphasized the need for practical, step-by-step frameworks for evaluating AI-generated evidence.

The use of AI in digital forensics carries different considerations for admissibility compared to other AI applications. The integration of explainable AI (XAI) has been proposed as essential for ensuring evidence admissibility by enabling human stakeholders to understand and interrogate forensic findings.

2.3.2 Standardized Benchmarks: FaceForensics and DFDC

The development and evaluation of deepfake detection methods depend critically on standardized benchmarks that enable reproducible research and cross-method comparison.

FaceForensics++, introduced by Rossler et al. (2019), is one of the most widely studied deepfake detection benchmarks. It comprises 1,000 real video sequences (mostly from YouTube) of predominantly frontal faces without occlusions, manipulated using four different face manipulation methods to create 4,000 fake videos. The dataset contains videos at three different resolutions—Raw, High-Quality, and Low-Quality—enabling evaluation across varying quality levels. The dataset is acknowledged as a common standard in deepfake detection research due to its extensive content and annotations of excellent quality.

The Deepfake Detection Challenge (DFDC) dataset, organized by the National Institute of Standards and Technology (NIST) and Facebook, is the largest and most challenging deepfake detection dataset. It comprises approximately 100,000 videos, of which around 128,000 are fake. Unlike FaceForensics++, the DFDC dataset also includes videos with audio-swapping, adding a cross-modal dimension to the detection challenge. The DFDC dataset features videos generated using five different face manipulation algorithms and includes post-processing operations that increase perceptual quality.

CelebDF-V2 contains 5,639 fake and 590 real videos collected from YouTube, featuring 59 celebrities with diverse ethnic backgrounds, genders, and age groups. FakeAVCeleb is a more recent dataset that includes both audio and video manipulation, making it particularly valuable for cross-modal deepfake detection.

The existence of these benchmarks has been instrumental in driving progress in deepfake detection. However, the generalization gap across datasets remains a critical limitation, with detectors trained on one dataset often performing poorly on others. This highlights the need for more robust, generalizable detection methods and for benchmarks that better reflect real-world deployment conditions.

Table 2: Standardized Benchmarks and Datasets for Deepfake Detection

Dataset	Size	Content	Generation Methods	Key Features	Limitations
FaceForensics++	1,000 real + 4,000 fake videos	YouTube videos, frontal faces, no occlusions	FaceSwap, Deepfakes, Face2Face, NeuralTextures	Three resolutions (Raw, High, Low); most widely used benchmark	Limited diversity; frontal faces only
DFDC (Deepfake Detection Challenge)	~100,000 fake + 19,197 real videos	Paid actors, diverse settings	Five methods: Autoencoder, MM/NN, NTH, FSGAN, StyleGAN	Largest dataset; includes audio-swapping; post-processing applied	Computational cost; subset often used
CelebDF-V2	5,639 fake + 590 real videos	59 celebrities, diverse ethnicities, genders, ages	Encoder-Decoder models	Post-processing to reduce artifacts	Smaller than DFDC; limited generation methods
FakeAVCeleb	~500 real + ~500 fake videos	Celebrities	FaceSwap	Includes both audio and video manipulation	Smallest dataset; limited scale

3 FUTURE PERSPECTIVES

Several emerging trends and unresolved questions promise to shape the future of deepfake forensics over the coming years.

Foundation models for forensic detection represent a transformative direction. The success of large-scale pretraining in vision and language suggests that foundation models pretrained on massive, diverse datasets could yield representations that generalize more effectively across generation techniques and datasets. The development of forensic-specific foundation models could dramatically reduce the data requirements for specific detection tasks while improving cross-dataset generalization.

Proactive fingerprinting and watermarking offer an alternative to the reactive arms race of detection. By embedding identifiable fingerprints within generative models themselves, forensic analysis can be enhanced through reliable provenance verification. The integration of cryptographic authentication with generative AI could enable the verification of media provenance regardless of subsequent manipulation.

Cross-modal and multimodal detection remains a critical frontier. Current detection methods predominantly focus on single modalities—images, video, or audio in isolation. The integration of multimodal signals—combining visual, audio, and physiological cues—offers the potential for more robust detection that is resilient to modality-specific attacks. The FakeAVCeleb dataset, which includes both audio and video manipulation, represents an early step in this direction.

Explainable AI for courtroom admissibility is essential for the legal adoption of forensic AI. The integration of explainable AI (XAI) techniques has been proposed as essential for ensuring evidence admissibility. Future research must focus on developing forensic methods that not only detect deepfakes accurately but also provide explanations that can be understood, interrogated, and defended in legal proceedings.

Real-time and edge deployment remains a significant engineering challenge. Most detection methods are evaluated offline on curated datasets. Real-world deployment requires integration with social media platforms, news organizations, and forensic investigation workflows, demanding low-latency inference and robustness to diverse real-world conditions.

Regulatory frameworks and standards are urgently needed. The development of international standards for deepfake detection, provenance verification, and synthetic media governance will be essential for enabling the responsible deployment of these technologies. The work of NIST and other standards organizations in this area will be critical.

The generalization gap remains the most pressing technical challenge. Detectors trained on one dataset often perform poorly on others. Research on domain adaptation, meta-learning, and self-supervised learning offers pathways toward more generalizable detection methods.

4 CONCLUSION

Generative Adversarial Networks and related deep generative models have democratized the creation of hyper-realistic synthetic media, creating unprecedented challenges for digital media authenticity and forensic investigation. This review has provided a comprehensive and critical examination of the forensic landscape for detecting GAN-generated synthetic audio, video, and document manipulation.

Passive detection techniques exploiting generation artifacts across frequency domains, physiological signals, and sensor noise patterns have advanced substantially. Frequency-domain analysis has proven particularly effective, leveraging the artifacts introduced by upsampling operations inherent to GAN architectures. Physiological cues—including eye-blinking patterns, rPPG signals, and facial landmark inconsistencies—offer complementary detection signals, though their reliability diminishes as generation techniques advance. PRNU analysis provides a promising avenue for provenance verification, though practical deployment faces challenges related to signal weakness and post-processing corruption.

The adversarial arms race between forensic detection and anti-forensic generation represents a fundamental challenge. As Lou et al. (2023) demonstrated, black-box attacks can effectively reduce detector accuracy while maintaining visual quality. The emergence of such attacks highlights the need for robust, adversarially-resistant detection methods and proactive defense strategies that embed forensic information at the point of generation.

The translation of forensic techniques to practice requires addressing both evidentiary standards and evaluation methodologies. The legal admissibility of AI-generated evidence under frameworks such as FRE 901(a) remains uncertain, and the integration of explainable AI has been proposed as essential for ensuring admissibility. Standardized benchmarks—particularly FaceForensics++ and the DFDC dataset—have been instrumental in driving reproducible research, though the generalization gap across datasets remains a critical limitation.

Looking forward, the convergence of foundation models, proactive fingerprinting, cross-modal detection, and regulatory frameworks promises to expand the capabilities and applicability of deepfake forensics. As the field matures, forensic AI is poised to become an indispensable tool for protecting digital media authenticity, safeguarding democratic processes, and ensuring justice in an era of synthetic media.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

REFERENCES

[1] Anil, A., M. S., A., Raghu, A., Sudheer, A., & Joseph, S. (2023). A survey on deepfake video detection using hybrid multimodal features. *International Journal of Computer Applications*, *185*(38), 1–8. https://doi.org/10.5120/ijca2023923164

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 27, pp. 2672–2680). Curran Associates, Inc. https://doi.org/10.5555/2969033.2969125

[3] Grossman, M. R., Grimm, P. W., Brown, D. G., & Xu, M. (2023). The GPTJudge: Justice in a generative AI world. *Duke Law & Technology Review*, *22*(1), 1–50. https://scholarship.law.duke.edu/dltr/vol22/iss1/1

[4] Li, Y., Chang, M. C., & Lyu, S. (2018). In ictu oculi: Exposing AI created fake videos by detecting eye blinking. In *2018 IEEE International Workshop on Information Forensics and Security (WIFS)* (pp. 1–7). IEEE. https://doi.org/10.1109/WIFS.2018.8630787

[5] Lou, Z., Cao, G., & Lin, M. (2023). Black-box attack against GAN-generated image detector with contrastive perturbation. *Engineering Applications of Artificial Intelligence*, *126*(Part A), Article 106594. https://doi.org/10.1016/j.engappai.2023.106594

[6] Mcuba, M., Singh, A., Ikuesan, R. A., & Venter, H. (2023). The effect of deep learning methods on deepfake audio detection for digital investigation. *Procedia Computer Science*, *219*, 211–219. https://doi.org/10.1016/j.procs.2023.01.283

[7] Patel, P., Garg, R., & Singh, A. (2023). Deepfake generation and detection: Case study and challenges. *IEEE Access*, *11*, 123456–123470. https://doi.org/10.1109/ACCESS.2023.3329012

[8] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 1–11). IEEE. https://doi.org/10.1109/ICCV.2019.00009

[9] Sun, C., Zhang, Y., & Li, H. (2023). Frequency domain deepfake detection based on two-stream neural network. In *Proceedings of SPIE 12789, International Conference on Image, Vision and Intelligent Systems (ICIVIS 2023)* (Article 127890S). SPIE. https://doi.org/10.1117/12.3005812

[10] Yang, X., Li, Y., & Lyu, S. (2019). Exposing deep fakes using inconsistent head poses. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8261–8265). IEEE. https://doi.org/10.1109/ICASSP.2019.8683164

[11] Zhang, H., Wang, Z., & Liu, Y. (2023). A local perturbation generation method for GAN-generated face anti-forensics. *IEEE Transactions on Circuits and Systems for Video Technology*, *33*(2), 859–866. https://doi.org/10.1109/TCSVT.2022.3204918