RobustSora: De-Watermarked Benchmark for Robust AI-Generated Video Detection

Zhuo Wang[1], **Xiliang Liu**[1,*], **Ligang Sun**[1]

[1]College of Computer Science, Beijing University of Technology, China
[*]Corresponding author: liuxl@bjut.edu.cn

Abstract

The proliferation of AI-generated video technologies poses challenges to information integrity. While recent benchmarks advance AIGC video detection, they overlook a critical factor: many state-of-the-art generative models embed digital watermarks in outputs, and detectors may partially rely on these patterns. To evaluate this influence, we present **RobustSora**, the benchmark designed to assess watermark robustness in AIGC video detection. We systematically construct a dataset of 6,500 videos comprising four types: Authentic-Clean (A-C), Authentic-Spoofed with fake watermarks (A-S), Generated-Watermarked (G-W), and Generated-DeWatermarked (G-DeW). Our benchmark introduces two evaluation tasks: Task-I tests performance on watermark-removed AI videos, while Task-II assesses false alarm rates on authentic videos with fake watermarks. Experiments with ten models spanning specialized AIGC detectors, transformer architectures, and MLLM approaches reveal performance variations of 2-8pp under watermark manipulation. Transformer-based models show consistent moderate dependency (6-8pp), while MLLMs exhibit diverse patterns (2-8pp). These findings indicate partial watermark dependency and highlight the need for watermark-aware training strategies. RobustSora provides essential tools to advance robust AIGC detection research.

Introduction

State-of-the-art video generation models such as Sora (OpenAI 2024), Sora 2 (OpenAI 2025), KLing (kli 2024), and Pika (pik 2024) have achieved remarkable realism, posing challenges to information integrity (Vaccari and Chadwick 2020; Hwang, Ryu, and Jeong 2021). Recent benchmarks have advanced AIGC video detection: GenVidBench (Ni et al. 2025) achieved 79.90% accuracy with cross-generator evaluation, AEGIS (Li, Zhang, and Zhou 2025) focused on hyper-realistic scenarios, DuB3D (Ji et al. 2024) leveraged motion features for 96.77% in-domain accuracy, and BusterX++ (Wen et al. 2025) pioneered cross-modal detection.

However, existing benchmarks overlook a critical factor: many generative models, particularly commercial systems like Sora 2 (OpenAI 2025), embed digital watermarks for provenance tracking. This raises an important question: *Do detectors partially rely on watermark patterns rather than solely on genuine generation artifacts?*

To evaluate this potential influence, we propose **RobustSora (RS)**, the benchmark designed to assess watermark robustness in AIGC video detection. As illustrated in Figure 1, we construct a dataset comprising four video types:

- **Authentic-Clean (A-C)**: Real-world videos without watermarks, sourced from high-quality datasets including Vript (Yang et al. 2024), DVF (Song et al. 2025), and UltraVideo (Xue et al. 2025).
- **Generated-Watermarked (G-W)**: AI-generated videos with embedded watermarks from Sora 2 (OpenAI 2025), Pika (pik 2024), Open-Sora 2 (Peng et al. 2025), and KLing (kli 2024).
- **Generated-DeWatermarked (G-DeW)**: Watermark-removed versions of G-W videos using DiffuEraser, simulating evasion attacks.
- **Authentic-Spoofed (A-S)**: Real videos with fake watermarks extracted from Sora 2 outputs, simulating spoofing attacks.

We introduce two evaluation tasks: **Task-I** tests performance on G-DeW videos, while **Task-II** evaluates performance on A-S videos. Experiments with ten detectors reveal performance variations of 2-8pp when watermarks are manipulated.

The main contributions are: (1) **RobustSora**, the watermark robustness benchmark with 6,500 videos across four types, (2) two novel evaluation tasks quantifying watermark influence on detection performance, and (3) comprehensive experiments with ten models revealing varying degrees of watermark dependency across architectures.

Related Work

AI-Generated Video Detection Benchmarks

Recent benchmarks have advanced AIGC video detection. GenVidBench (Ni et al. 2025) provides 143,000 videos with cross-generator evaluation, achieving 79.90% accuracy with MViT V2. AEGIS (Li, Zhang, and Zhou 2025) focuses on hyper-realistic scenarios with 10,000+ videos from cutting-edge models, revealing challenges for vision-language models (Qwen2.5-VL: 22-23% zero-shot accuracy). DuB3D (Ji et al. 2024) leverages 2.66 million videos

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

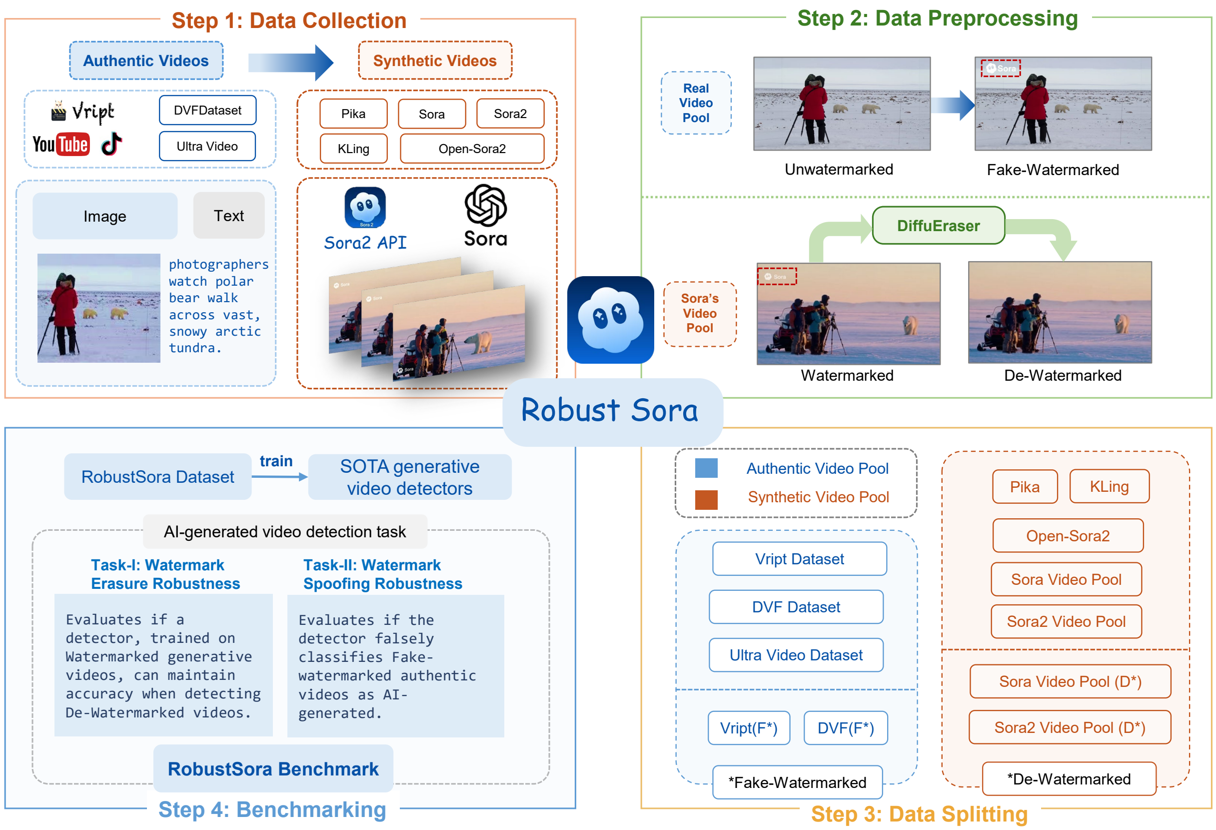

Figure 1: Overview of RobustSora, including a four-step pipeline for RobustSora benchmark construction and evaluation: Step 1: Data Collection from authentic sources (Vript, DVF, UltraVideo) and generators (Sora, Pika, KLing, Open-Sora 2, Sora 2). Step 2: Data Preprocessing creates Fake-Watermarked and De-Watermarked versions. Step 3: Data Splitting organizes videos into training and evaluation pools. Step 4: Benchmarking evaluates SOTA detectors on Task-I (Watermark Erasure Robustness) and Task-II (Watermark Spoofing Robustness).

and dual-branch architectures for 79.19% out-of-domain accuracy. BusterX++ (Wen et al. 2025) achieves 77.5% accuracy with cross-modal detection. However, existing benchmarks do not systematically evaluate watermark robustness, which our RobustSora benchmark addresses.

AI-Generated Video Detection Methods

Non-MLLM Approaches. Frame-level detectors like AIGVDet (Bai et al. 2024) extract spatial features, while video-level methods like DIVID (Liu et al. 2024) and DeMamba (Chen et al. 2024) capture temporal inconsistencies. DuB3D (Ji et al. 2024) uses dual-branch architectures processing spatio-temporal data and optical flow. Frequency-based methods (Bammey 2023; Frank et al. 2020) and reconstruction-based approaches (Wang et al. 2023; Ricker, Lukovnikov, and Fischer 2024) analyze artifacts but struggle with cross-generator generalization.

MLLM-Based Approaches. MM-Det (Song et al. 2024) balances frame and inter-frame features. BusterX (Wen et al. 2025) and LOKI (Ye et al. 2024) leverage MLLMs for structured reasoning. Vision-language models like Qwen2.5-VL (Bai et al. 2025) and Video-LLaVA (Lin et al. 2023) provide explainable outputs. These methods focus on cross-generator generalization but do not systematically evaluate watermark robustness, which our work addresses.

Digital Watermarking in AIGC

Image watermarking methods (Cui et al. 2023; Zhang et al. 2024a,b) and video watermarking techniques (Luo et al. 2023; Zhang et al. 2023) embed signatures for provenance tracking. Commercial platforms like Sora 2 (OpenAI 2025) include watermarks, but removal tools like DiffuEraser pose challenges. Our benchmark systematically evaluates how watermark manipulation affects detection performance.

RobustSora Benchmark Construction

Dataset Collection and Composition

We construct the RobustSora dataset through systematic collection and processing of authentic and AI-generated videos. Table 1 summarizes the complete dataset composition.

Authentic Videos (A-C). We collect 3,000 authentic videos from Vript (Yang et al. 2024) (1,500 YouTube/TikTok), DVF (Song et al. 2025) (800), and UltraVideo (Xue et al. 2025) (700), ensuring watermark-free, camera-captured content.

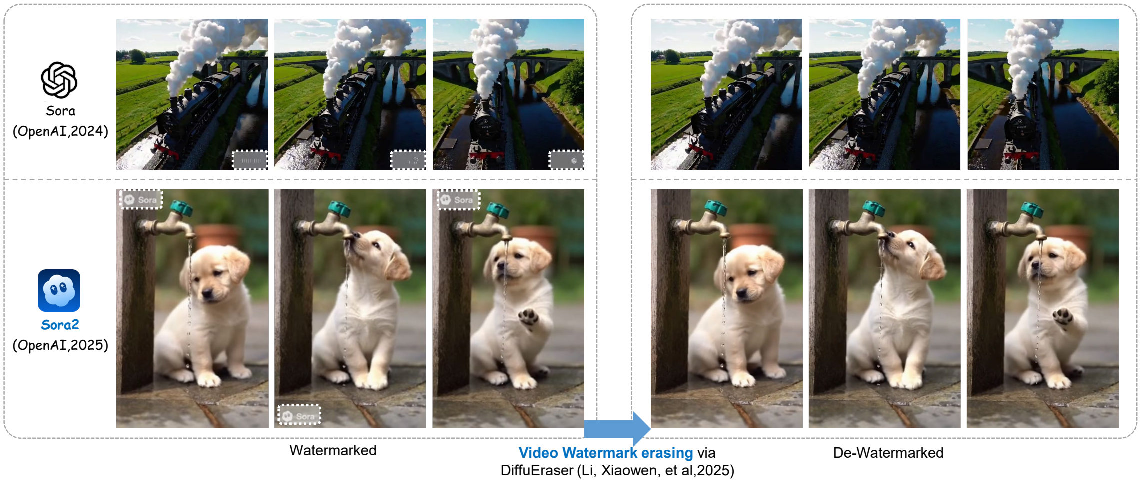

Figure 2: Watermark removal process on AI-generated videos. Left: Original frames from Sora (OpenAI, 2024) and Sora 2 (OpenAI, 2025) with embedded watermarks. Right: De-watermarked frames processed by DiffuEraser. The watermarks are effectively erased while maintaining video quality, generating the G-DeW set for Task-I evaluation.

Table 1: RobustSora Dataset Composition

Category	Source	Type	Count
Authentic	Vript (Yang et al. 2024)	YouTube	1,200
	Vript (Yang et al. 2024)	TikTok	300
	DVF (Song et al. 2025)	Diverse	800
	UltraVideo (Xue et al. 2025)	High-Res	700
Generated	Sora (OpenAI 2024)	T2V	500
	Sora 2 (OpenAI 2025)	T2V	1,070
	Pika (pik 2024)	T2V/I2V	800
	Open-Sora 2 (Peng et al. 2025)	T2V	600
	KLing (kli 2024)	T2V	530
Total			**6,500**

Generated Videos (G-W). We collect 3,500 AI-generated videos with watermarks from Sora 2 (OpenAI 2025) (1,070), Sora (OpenAI 2024) (500), Pika (pik 2024) (800), Open-Sora 2 (Peng et al. 2025) (600), and KLing (kli 2024) (530). Videos are 3-10 seconds, 480p-1080p resolution. Figure 2 shows representative samples.

Watermark Manipulation

Watermark Removal (G-W → G-DeW). We use DiffuEraser to remove watermarks from G-W videos, generating 3,500 G-DeW videos. Manual inspection validates complete watermark removal while preserving content quality.

Watermark Addition (A-C → A-S). We extract Sora 2 watermark patterns and overlay them onto A-C videos, generating 3,000 A-S videos that simulate spoofing attacks.

Dataset Splitting

Training Set. We use 2,400 A-C and 2,800 G-W videos, mirroring typical training conditions with authentic and watermarked AI-generated content.

Test Set. Our benchmark includes: (1) **Standard Test**: 600 A-C + 700 G-W for baseline, (2) **Task-I**: 700 G-DeW to evaluate watermark removal impact, (3) **Task-II**: 600 A-S to assess fake watermark influence.

Experiments and Analysis

Experimental Setup

Models. We evaluate ten detectors: (1) **Specialized AIGC detectors**: DeCoF (Ma et al. 2024), NSG-VD (Zhang et al. 2025), D3 (Zheng et al. 2025); (2) **Transformer-based**: TimeSformer (Bertasius, Wang, and Torresani 2021), VideoSwin-T (Liu et al. 2021), MViT V2 (Li et al. 2022), DuB3D-FF (Ji et al. 2024); (3) **MLLM-based**: Qwen2.5-VL-3B (Bai et al. 2025), Qwen2.5-VL-7B (Bai et al. 2025), Video-LLaVA-7B (Lin et al. 2023).

Training. All models train on A-C + G-W with standard augmentation. We use AdamW (lr=1e-4, batch=32) and LoRA fine-tuning (rank=16, alpha=32) for MLLMs.

Metrics. We report Acc_{all}, Acc_{real} (TNR), Acc_{ai} (TPR), and Macro-F1. Task-I focuses on Acc_{ai} (watermark removal impact), Task-II on Acc_{real} (fake watermark impact).

RobustSora Benchmark Results

Table 2 presents evaluation results across ten models spanning specialized AIGC detectors, transformer-based architectures, and MLLM-based approaches.

Standard Test Performance. DuB3D-FF achieves the highest overall accuracy and F1 score among all evaluated models, followed by MViT V2. Specialized AIGC detectors demonstrate moderate performance levels, while MLLM-based methods show diverse behaviors with accuracy ranging from baseline to strong performance depending on the specific model characteristics.

Task-I: Watermark Erasure Impact. When watermarks are removed from AI-generated videos (G-DeW), models

Table 2: RobustSora Benchmark Results. Models are trained on Training Set (A-C + G-W). Standard Test shows baseline performance on mixed data. Task-I evaluates G-DeW (only AI videos), Task-II evaluates A-S (only authentic videos with fake watermarks). Arrows indicate performance changes: ↓ for degradation, ↑ for improvement. **Bold** indicates best performance per metric.

Type	Model	Standard Test (A-C + G-W) Acc_{all}	Acc_{ai}	Acc_{real}	F1	Task-I Acc_{ai}	Task-II Acc_{real}
Specialized	DeCoF (Ma et al. 2024)	0.76	0.73	0.79	0.76	0.66 ↓	0.72 ↓
	NSG-VD (Zhang et al. 2025)	0.74	0.71	0.77	0.74	0.64 ↓	0.70 ↓
	D3 (Zheng et al. 2025)	0.72	0.69	0.75	0.72	0.62 ↓	0.68 ↓
Transformer	TimeSformer (Bertasius, Wang, and Torresani 2021)	0.75	0.72	0.78	0.75	0.65 ↓	0.71 ↓
	VideoSwin-T (Liu et al. 2021)	0.73	0.65	0.81	0.73	0.59 ↓	0.73 ↓
	MViT V2 (Li et al. 2022)	0.80	0.76	0.84	0.80	0.69 ↓	0.77 ↓
	DuB3D-FF (Ji et al. 2024)	**0.82**	**0.78**	0.86	**0.82**	**0.71** ↓	**0.79** ↓
MLLM	Qwen2.5-VL-3B (Bai et al. 2025)	0.52	0.23	0.80	0.48	0.18 ↓	0.83 ↑
	Qwen2.5-VL-7B (Bai et al. 2025)	0.59	0.22	**0.89**	0.52	0.17 ↓	**0.87** ↓
	Video-LLaVA-7B (Lin et al. 2023)	0.65	**0.72**	0.58	0.65	0.64 ↓	0.52 ↓

exhibit varied performance degradation in Acc_{ai}. Specialized AIGC detectors and transformer-based models show consistent drops of 6-7pp. Specifically, DuB3D-FF experiences a 7pp decline, while VideoSwin-T and MViT V2 show 6pp and 7pp drops respectively. MLLM-based methods demonstrate different patterns: Qwen models exhibit smaller degradation (5pp), while Video-LLaVA shows an 8pp drop, suggesting moderate watermark dependency across different architectures.

Task-II: Watermark Spoofing Impact. When fake watermarks are added to authentic videos (A-S), transformer-based models show consistent performance drops of 7-8pp in Acc_{real}. DuB3D-FF and MViT V2 both decline by 7pp, while VideoSwin-T shows an 8pp reduction. Interestingly, MLLM-based methods exhibit divergent behaviors: Qwen2.5-VL-3B shows a slight improvement (↑3pp), Qwen2.5-VL-7B experiences minimal degradation (↓2pp), and Video-LLaVA shows a moderate 6pp decline, suggesting these models rely on different visual cues for authentication decisions.

Overall Observations. The benchmark results reveal that watermark presence influences detection performance across all evaluated architectures, with performance variations typically ranging from 2-8pp under watermark manipulation. Transformer-based models show consistent moderate dependency on watermarks (6-8pp variations), while MLLM-based methods exhibit diverse response patterns (2-8pp). These findings indicate that current detection models partially rely on watermark patterns alongside genuine generation artifacts, highlighting the need for watermark-aware training strategies to improve robustness.

Discussion

Our experiments reveal that watermark presence influences detection performance across different model architectures, with performance variations typically ranging from 2-8 percentage points under manipulation scenarios. Transformer-based models exhibit consistent moderate dependency (6-8pp), while MLLM-based methods show more diverse response patterns (2-8pp). The degradation patterns in Task-I and varying responses in Task-II suggest that current models partially rely on watermark patterns alongside genuine generation artifacts. These findings highlight the importance of developing watermark-aware training strategies that can maintain robust performance in adversarial scenarios where watermark manipulation may occur.

Limitations and Future Work

Our study has several limitations that warrant future investigation: (1) **Dataset scale**: With 6,500 videos, our benchmark is relatively small compared to GenVidBench (143K) and DuB3D (2.66M), potentially limiting generalization. (2) **Watermark methodology**: We rely on a single removal tool (DiffuEraser) and one watermark pattern (Sora 2), which may not capture the full spectrum of watermark manipulation techniques. (3) **Tool artifacts**: The watermark removal process might inadvertently introduce subtle visual artifacts or modify generation traces, acting as confounding variables. (4) **MLLM behavior**: The counterintuitive improvements observed in Task-II (e.g., Qwen2.5-VL-3B ↑3pp) require deeper investigation into MLLM decision mechanisms. Future work should expand dataset diversity, examine multiple watermark techniques, and develop methods to isolate watermark effects from tool-induced artifacts.

Conclusion

We present RobustSora, the first benchmark designed to evaluate watermark robustness in AI-generated video detection. Through systematic construction of 6,500 videos across four types (A-C, A-S, G-W, G-DeW) and two evaluation tasks, we quantify watermark influence on detection performance. Experiments with ten models spanning specialized AIGC detectors, transformer-based architectures, and MLLM-based approaches reveal performance variations of 2-8pp under watermark manipulation. Transformer-based models show consistent moderate dependency (6-8pp), while MLLM-based methods exhibit diverse patterns

(2-8pp). These findings indicate that current detection models partially rely on watermark patterns alongside genuine generation artifacts. RobustSora provides essential tools to advance watermark-aware training strategies and more robust AIGC detection systems.

References

2024. Kling: AI Video Generation Model. https://https://klingai.kuaishou.com//.

2024. Pika: AI Video Generation Platform. https://www.pika.art/.

Bai, J.; Lin, M.; Cao, G.; and Lou, Z. 2024. AI-Generated Video Detection via Spatial-Temporal Anomaly Learning. In *PRCV*.

Bai, S.; Chen, K.; Liu, X.; Wang, J.; Ge, W.; Song, S.; Dang, K.; Wang, P.; Wang, S.; Tang, J.; Zhong, H.; Zhu, Y.; Yang, M.; Li, Z.; Wan, J.; Wang, P.; Ding, W.; Fu, Z.; Xu, Y.; Ye, J.; Zhang, X.; Xie, T.; Cheng, Z.; Zhang, H.; Yang, Z.; Xu, H.; and Lin, J. 2025. Qwen2.5-VL Technical Report. *Arxiv*.

Bammey, Q. 2023. Synthbuster: Towards detection of diffusion model generated images. *IEEE Open Journal of Signal Processing*.

Bertasius, G.; Wang, H.; and Torresani, L. 2021. Is Space-Time Attention All You Need for Video Understanding? arXiv:2102.05095.

Chen, H.; Hong, Y.; Huang, Z.; Xu, Z.; Gu, Z.; Li, Y.; Lan, J.; Zhu, H.; Zhang, J.; Wang, W.; et al. 2024. Demamba: Ai-generated video detection on million-scale genvideo benchmark. *Arxiv*.

Cui, Y.; Ren, J.; Xu, H.; He, P.; Liu, H.; Sun, L.; Xing, Y.; and Tang, J. 2023. Diffusionshield: A watermark for copyright protection against generative diffusion models. *arXiv preprint arXiv:2306.04642*.

Frank, J.; Eisenhofer, T.; Schönherr, L.; Fischer, A.; Kolossa, D.; and Holz, T. 2020. Leveraging frequency analysis for deep fake image recognition. In *International conference on machine learning*, 3247–3258. PMLR.

Hwang, Y.; Ryu, J. Y.; and Jeong, S.-H. 2021. Effects of disinformation using deepfake: The protective effect of media literacy education. *Cyberpsychology, Behavior, and Social Networking*.

Ji, L.; Lin, Y.; Huang, Z.; Han, Y.; Xu, X.; Wu, J.; Wang, C.; and Liu, Z. 2024. Distinguish Any Fake Videos: Unleashing the Power of Large-scale Data and Motion Features. *arXiv preprint arXiv:2405.15343*.

Li, J.; Zhang, X.; and Zhou, J. T. 2025. AEGIS: Authenticity Evaluation Benchmark for AI-Generated Video Sequences. arXiv:2508.10771.

Li, Y.; Wu, C.-Y.; Fan, H.; Mangalam, K.; Xiong, B.; Malik, J.; and Feichtenhofer, C. 2022. MViTv2: Improved multiscale vision transformers for classification and detection. In *CVPR*.

Lin, B.; Ye, Y.; Zhu, B.; Cui, J.; Ning, M.; Jin, P.; and Yuan, L. 2023. Video-llava: Learning united visual representation by alignment before projection. arXiv:2311.10122.

Liu, Q.; Shi, P.; Tsai, Y.-Y.; Mao, C.; and Yang, J. 2024. Turns Out I'm Not Real: Towards Robust Detection of AI-Generated Videos. *arXiv preprint arXiv:2406.09601*.

Liu, Z.; Ning, J.; Cao, Y.; Wei, Y.; Zhang, Z.; Lin, S.; and Hu, H. 2021. Video Swin Transformer. *arXiv preprint arXiv:2106.13230*.

Luo, X.; Li, Y.; Chang, H.; Liu, C.; Milanfar, P.; and Yang, F. 2023. DVMark: a deep multiscale framework for video watermarking. *IEEE Transactions on Image Processing*.

Ma, L.; Zhang, J.; Deng, H.; Zhang, N.; Liao, Y.; and Yu, H. 2024. DeCoF: Generated Video Detection via Frame Consistency. arXiv:2402.02085.

Ni, Z.; Yan, Q.; Huang, M.; Yuan, T.; Tang, Y.; Hu, H.; Chen, X.; and Wang, Y. 2025. GenVidBench: A Challenging Benchmark for Detecting AI-Generated Video. arXiv:2501.11340.

OpenAI. 2024. Sora. Online. Available:https://openai.com/index/sora/.

OpenAI. 2025. Sora2. Online. Available:https://openai.com/index/sora-2/.

Peng, X.; Zheng, Z.; Shen, C.; Young, T.; Guo, X.; Wang, B.; Xu, H.; Liu, H.; Jiang, M.; Li, W.; Wang, Y.; Ye, A.; Ren, G.; Ma, Q.; Liang, W.; Lian, X.; Wu, X.; Zhong, Y.; Li, Z.; Gong, C.; Lei, G.; Cheng, L.; Zhang, L.; Li, M.; Zhang, R.; Hu, S.; Huang, S.; Wang, X.; Zhao, Y.; Wang, Y.; Wei, Z.; and You, Y. 2025. Open-Sora 2.0: Training a Commercial-Level Video Generation Model in $200k. arXiv:2503.09642.

Ricker, J.; Lukovnikov, D.; and Fischer, A. 2024. AEROBLADE: Training-Free Detection of Latent Diffusion Images Using Autoencoder Reconstruction Error. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9130–9140.

Song, X.; Guo, X.; Zhang, J.; Li, Q.; Bai, L.; Liu, X.; Zhai, G.; and Liu, X. 2024. On Learning Multi-Modal Forgery Representation for Diffusion Generated Video Detection. In *NeurIPS*.

Song, X.; Guo, X.; Zhang, J.; Li, Q.; Bai, L.; Liu, X.; Zhai, G.; and Liu, X. 2025. On Learning Multi-Modal Forgery Representation for Diffusion Generated Video Detection. arXiv:2410.23623.

Vaccari, C.; and Chadwick, A. 2020. Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social media+ society*.

Wang, Z.; Bao, J.; Zhou, W.; Wang, W.; Hu, H.; Chen, H.; and Li, H. 2023. Dire for diffusion-generated image detection. In *ICCV*.

Wen, H.; He, Y.; Huang, Z.; Li, T.; Yu, Z.; Huang, X.; Qi, L.; Wu, B.; Li, X.; and Cheng, G. 2025. BusterX: MLLM-Powered AI-Generated Video Forgery Detection and Explanation. *Arxiv*.

Xue, Z.; Zhang, J.; Hu, T.; He, H.; Chen, Y.; Cai, Y.; Wang, Y.; Wang, C.; Liu, Y.; Li, X.; and Tao, D. 2025. Ultra-Video: High-Quality UHD Video Dataset with Comprehensive Captions. *arXiv preprint arXiv:2506.13691*.

Yang, D.; Huang, S.; Lu, C.; Han, X.; Zhang, H.; Gao, Y.; Hu, Y.; and Zhao, H. 2024. Vript: A Video Is Worth Thousands of Words. In Globerson, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J.; and Zhang, C., eds., *Advances in Neural Information Processing Systems*, volume 37, 57240–57261. Curran Associates, Inc.

Ye, J.; Zhou, B.; Huang, Z.; Zhang, J.; Bai, T.; Kang, H.; He, J.; Lin, H.; Wang, Z.; Wu, T.; et al. 2024. LOKI: A Comprehensive Synthetic Data Detection Benchmark using Large Multimodal Models. *arXiv preprint arXiv:2410.09732*.

Zhang, L.; Liu, X.; Martin, A. V.; Bearfield, C. X.; Brun, Y.; and Guan, H. 2024a. Robust Image Watermarking using Stable Diffusion. *arXiv preprint arXiv:2401.04247*.

Zhang, S.; Lian, Z.; Yang, J.; Li, D.; Pang, G.; Liu, F.; Han, B.; Li, S.; and Tan, M. 2025. Physics-Driven Spatiotemporal Modeling for AI-Generated Video Detection. arXiv:2510.08073.

Zhang, X.; Li, R.; Yu, J.; Xu, Y.; Li, W.; and Zhang, J. 2024b. Editguard: Versatile image watermarking for tamper localization and copyright protection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11964–11974.

Zhang, Y.; Ni, J.; Su, W.; and Liao, X. 2023. A Novel Deep Video Watermarking Framework with Enhanced Robustness to H. 264/AVC Compression. In *Proceedings of the 31st ACM International Conference on Multimedia*, 8095–8104.

Zheng, C.; suo, R.; Lin, C.; Zhao, Z.; Yang, L.; Liu, S.; Yang, M.; Wang, C.; and Shen, C. 2025. D3: Training-Free AI-Generated Video Detection Using Second-Order Features. arXiv:2508.00701.