VIDFORENSICS-M1: AI-GENERATED VIDEO DETECTION AND FORENSICS VIA VERIFIABLE META-DETECTION REINFORCEMENT LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Recent advances in video generation models have dramatically improved the realism of synthetic videos, blurring the boundary between generated and authentic content and raising significant concerns about misinformation. Existing MLLM-based detectors predominantly rely on supervised fine-tuning or label-level reinforcement learning, where predefined or coarse-grained supervision signals limit their generalization to out-of-domain scenarios and emerging video generators. To overcome these limitations, we are the first to introduce **meta-detection** into AI-generated video detection, jointly supervising label correctness and evidence correctness during reinforcement learning to enable more reliable forgery detection. This paradigm presents two core challenges: (1) identifying evidence forms that provide reliable and scalable supervision for meta-detection, and (2) developing effective mechanisms to integrate such evidence into label-level reinforcement learning for trustworthy and generalizable synthetic video detection. Textual rationales offer semantically rich descriptions of forgery artifacts, yet their generation and verification rely heavily on external reference models, making the resulting supervision vulnerable to hallucinations and semantic biases. This motivates us to seek an evidence representation with verifiable ground truth to support rule-based meta-detection supervision. Temporal grounding meets this requirement because manipulation intervals can be precisely controlled and recorded during data construction. Building on this insight, we adopt temporal grounding as a more reliable meta-detection signal and propose an automated data construction pipeline that generates paired real and fake videos by reconstructing and replacing temporal segments using boundary-frame-conditioned video generation models. Furthermore, we propose **Evidence-Guided Reward Redistribution**, which performs evidence-aware credit assignment by redistributing rewards among label-correct responses according to their evidence quality, thereby preserving reliable label supervision while progressively encouraging the detector to acquire fine-grained and verifiable forgery localization capabilities. Extensive experiments demonstrate that **VidForensics-M1** effectively leverages verifiable temporal evidence to achieve more robust and generalizable AI-generated video detection.

1 INTRODUCTION

The rapid evolution of generative foundation models has significantly advanced the visual fidelity of AI-generated videos (Hou et al., 2026; Li et al., 2026; Xiong et al., 2026; Xu et al., 2025a;b; Zhang et al., 2026), making them increasingly difficult to distinguish from authentic content. While these models have enabled unprecedented progress in digital content creation, their widespread accessibility also introduces significant challenges to media authenticity and misinformation mitigation (Davodi et al., 2026). Consequently, developing reliable methods for AI-generated video authenticity verification has emerged as a critical task for ensuring trustworthy digital media (Vaccari & Chadwick, 2020; Chandra et al., 2024).

As shown in Fig. 1, existing MLLM-based detection methods can be broadly categorized into two paradigms. The first paradigm learns fine-grained artifact perception through supervised fine-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

tuning on visual chain-of-thought annotations provided by human annotators or teacher models (Gao et al., 2025; Fu et al., 2025b; Li et al., 2025; Tan et al., 2026; Corvi et al., 2025; Wei et al., 2022). Although these approaches enable detectors to capture fine-grained visual forgery artifacts, imitating predefined reasoning trajectories can lead to overfitting to specific annotation patterns or training distributions, thereby limiting their generalization ability in open-world scenarios. The second paradigm leverages reinforcement learning to improve generalization across diverse video generation models and data distributions. However, existing approaches mainly rely on label-level supervision signals, which may encourage detectors to exploit spurious correlations or superficial cues rather than identifying the underlying forgery artifacts (Park et al., 2025; Wen et al., 2025b; Christiano et al., 2017; Amodei et al., 2016; Guo et al., 2025).

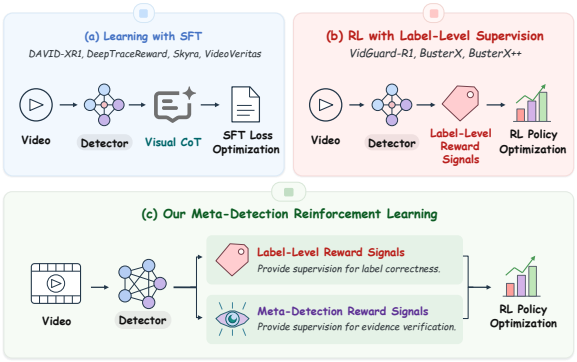

Figure 1: Comparison of MLLM-Based AI-Generated Video Detection Methods.

To address these limitations, we introduce **meta-detection**, which jointly optimizes label correctness and evidence validity within a reinforcement learning framework to enable more reliable synthetic video detection. This evidence-aware feedback provides a more informative and verifiable learning signal for policy optimization, guiding detectors toward identifying and localizing visual forgery artifacts and ultimately achieving more robust and generalizable AI-generated video detection.

A core question is what forms of evidence can serve as trustworthy and scalable supervision for meta-detection? In AI-generated video detection, meta-detection signals can naturally take two complementary forms: textual descriptions of forgery artifacts and temporal evidence that localizes when the manipulation occurs. Although textual rationales provide semantically rich descriptions of forgery artifacts, they are typically generated by powerful reference models, meaning that the resulting supervision may inherit the biases and limitations of the teacher models rather than represent objective and verifiable ground truth (Gao et al., 2025; Fu et al., 2025b; Turpin et al., 2023; Li et al., 2025; He et al., 2021). In contrast, temporal grounding provides a more objective and verifiable supervision signal, as the manipulated temporal intervals are explicitly determined through the controlled forgery construction process (Wang et al., 2025).

Based on this, we adopt temporal grounding as a more reliable meta-detection signal and propose an automated and scalable data construction pipeline that generates paired real–fake videos with verifiable temporal evidence. As shown in Fig. 2, we randomly sample a temporal interval from each real video and retain the preceding and succeeding segments as boundary contexts. The corresponding boundary frames are then fed into video generation models to reconstruct the missing content, which is inserted between the two contexts to produce the paired fake video. This controlled process naturally yields ground-truth temporal annotations for the manipulated interval. We further employ strong foundation models to describe observable forgery artifacts in each fake video, with particular emphasis on geometric deformation, temporal inconsistency, and physical violations.

Another core question is how to effectively incorporate meta-detection feedback into label-level reinforcement learning to enable trustworthy and generalizable synthetic video detection? To address this, we propose Evidence-Guided Reward Redistribution (EGRR), which reallocates rewards among label-correct responses according to their evidence quality. As shown in Fig. 3, EGRR maintains stable label-level optimization while calibrating the reward distribution based on evidence reliability, encouraging detectors to recognize fine-grained visual forgery artifacts.

Our contributions can be summarized as follows:

- We are the first to introduce the concept of **meta-detection** into AI-generated video detection and incorporate it into reinforcement learning as an evidence-aware feedback signal.
- We demonstrate that rule-based temporal grounding provides more reliable and verifiable feedback than model-based textual rationales, making it a more suitable evidence source for meta-detection.

- We propose an automated and scalable data construction pipeline through boundary-frame-conditioned temporal segment reconstruction and replacement, which constructs paired real-fake videos while providing ground-truth manipulated temporal intervals.
- We propose Evidence-Guided Reward Redistribution (EGRR), which reallocates rewards among label-correct responses according to their evidence quality, enabling effective calibration of label-level reinforcement learning with meta-detection feedback.

2 Related Work

AI-Generated Video Detection Methods. Early research on visual AIGC forensics predominantly explored image-level forgery detection (Wang et al., 2020; Ojha et al., 2023; Yan et al., 2024; Tan et al., 2024; Nguyen et al., 2024; Fu et al., 2025a; Yan et al., 2025; Yang et al., 2025). As video generation models continue to advance, increasing attention has been devoted to detecting AI-generated videos (Hou et al., 2026). DAVID-XR1 (Gao et al., 2025) distills teacher-generated visual chains of thought grounded in defect categories and spatio-temporal annotations. DeeptraceReward (Fu et al., 2025b) trains multimodal reward models from human rationales, bounding boxes, and temporal intervals. Skyra (Li et al., 2025) generates grounded artifact analyses containing temporal intervals and spatial locations. VidGuard-R1 (Park et al., 2025) combines an SFT cold start with GRPO and specialized rewards for temporal artifacts and generation complexity. BusterX (Wen et al., 2025a) bypasses SFT and trains directly with reinforcement learning using video clips and binary authenticity labels. BusterX++ (Wen et al., 2025b) extends this pure-RL formulation to unified image and video detection, preserving policy entropy to encourage cross-modal exploration. VideoVeritas (Tan et al., 2026) improves fine-grained video perception through automatically verifiable pretext tasks, including grounding and counting, rather than directly optimizing the detection objective. Unlike existing approaches, we introduce **meta-detection** into AI-generated video detection, where evidence-aware feedback is incorporated into reinforcement learning to jointly optimize prediction correctness and evidence validity.

AI-Generated Video Detection Datasets. With the rapid advancement of generative models, an increasing number of datasets have been introduced to facilitate research on AI-generated video detection (Fu et al., 2025c; Bai et al., 2024; Chen et al., 2024; Ma et al., 2025; Wen et al., 2025a; Gao et al., 2025; Fu et al., 2025b; Li et al., 2025). GVD (Bai et al., 2024) collects videos generated by diverse text-to-video and image-to-video models, while GenVideo (Chen et al., 2024) scales the benchmark to millions of real and synthetic videos. GVF (Ma et al., 2025) mitigates semantic shortcuts by constructing content-matched fake videos from prompts extracted from the corresponding real videos. GenVidBench (Ni et al., 2026) introduces cross-source and cross-generator evaluation settings together with semantic annotations. GenBuster-200K (Wen et al., 2025a) incorporates recent video generation models and large-scale real-world videos for training MLLM-based detectors. Recent evidence-oriented datasets, including DAVID-X (Gao et al., 2025), DeeptraceReward (Fu et al., 2025b), and ViF-CoT-4K (Li et al., 2025), provide artifact rationales and spatio-temporal annotations, but do not derive exact manipulated intervals from a controlled segment-replacement process. We address this gap by constructing paired real-fake videos through controlled segment manipulation with ground-truth temporal localization.

3 Verifiable Temporal Feedback for Meta-Detection

In this section, we first motivate the introduction of meta-detection into reinforcement learning. We then analyze the limitations of textual explanations as meta-detection feedback signals and introduce rule-based temporal grounding as a more reliable evidence source. Finally, we present an automated and scalable data construction pipeline for generating paired real-fake videos with verifiable temporal forgery localization annotations.

3.1 Problem Definition

Traditional RLVR-based video detectors are trained on a dataset $\mathcal{D} = \{(V_n, y_n)\}_{n=1}^{N}$, where each video $V_n = \{f_{n,t}\}_{t=1}^{T}$ consists of T frames and $y_n \in \{\text{real}, \text{fake}\}$ denotes its authenticity label. For each training video V_i, Group Relative Policy Optimization samples G responses $\{o_{ij}\}_{j=1}^{G}$ from

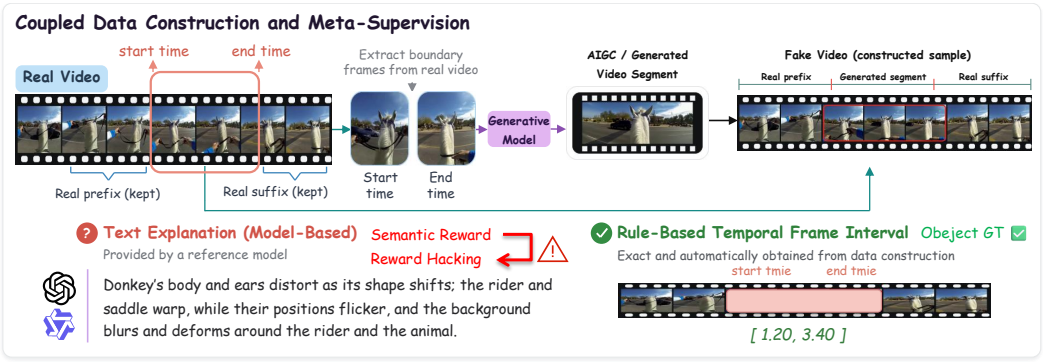

Figure 2: Overview of our automated and scalable data construction pipeline and comparison between two meta-detection feedback signals.

the current policy: $o_{ij} \sim \pi_\theta(\cdot \mid V_i)$. The predicted label $\hat{y}_{ij}$ is parsed from response o_{ij}. Its label correctness is defined as $\mathbb{I}[\hat{y}_{ij} = y_i] \in \{0, 1\}$. Traditional label-level RLVR uses the following reward:

$$R^{(i,j)} = \lambda_{\text{format}} R^{(i,j)}_{\text{format}} + (1 - \lambda_{\text{format}})\, R^{(i,j)}_{\text{label}}. \tag{1}$$

Although this reward encourages models to follow the required output format and predict correct labels, it cannot distinguish whether detectors genuinely identify visual forgery artifacts or obtain rewards by exploiting unreliable shortcuts. To address this limitation, we introduce the concept of **meta-detection**, which extracts the predicted label $\hat{y}_{ij}$ and decision evidence $\hat{e}_{ij}$ from response o_{ij}, aiming to jointly supervise binary label prediction and evidence trustworthiness for more reliable reinforcement learning.

3.2 Drawbacks of Model-Based Textual Rationales

Textual explanations offer semantically rich descriptions of visual forgery artifacts, including geometric deformation, temporal inconsistency, and physical violation. However, these explanations are typically generated by powerful reference models, making the resulting supervision reflect model-dependent interpretations of forgery traces rather than objective ground truth and potentially inheriting the biases and limitations of the reference models. Furthermore, evaluating generated textual explanations typically requires an additional auxiliary model to measure their semantic consistency with reference explanations produced by the reference model. This verification process not only introduces additional sources of model-induced bias but also incurs substantial computational overhead. Consequently, model-based textual explanations remain vulnerable to subjective biases during both generation and evaluation, potentially leading to reward hacking.

3.3 Paired Video Construction Pipeline for Verifiable Temporal Signals

In contrast, temporal grounding provides a more objective and verifiable supervision signal, as the manipulated temporal intervals can be explicitly determined through a controlled forgery construction process. Based on this observation, we propose an automated and scalable data construction pipeline that automatically obtains ground-truth forgery intervals during the construction process. As shown in Fig. 2, we collect real videos from InternVid and ActivityNet and uniformly process them into 5-second clips. For each real video, we randomly remove a temporal segment and reconstruct the missing content using boundary frames and video generation models. The generated segments are then aligned with the removed segments in terms of duration, resolution, and frame rate before being inserted back into the original temporal locations, producing the corresponding fake videos. For each fake sample, we employ GPT-5.5 (OpenAI, 2025) to generate a textual description of observable forgery artifacts, with particular emphasis on geometric deformation, temporal inconsistency, and physical violations. We then use Gemini-3.1-Pro (Google DeepMind, 2026) to filter out fake samples with unreliable explanations together with their matched real counterparts, thereby preserving a balanced 1:1 ratio between fake and real videos. All experiments in this setting are conducted on the resulting filtered subset. The quality of temporal evidence is quantified by computing the Intersection-over-Union (IoU) between the predicted temporal interval and the reference interval.

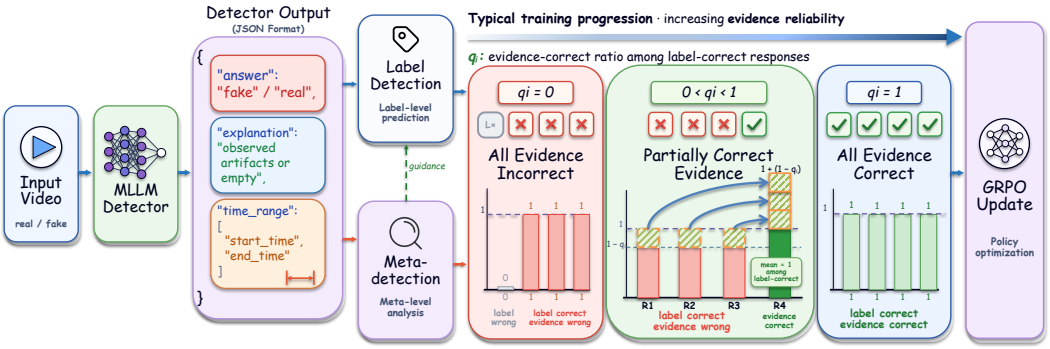

Figure 3: Overview of the Evidence-Guided Reward Redistribution (EGRR) pipeline. The detector takes videos as input and generates JSON-formatted responses containing both label-level and meta-detection signals. The meta-detection signal redistributes rewards among label-correct responses according to evidence quality, which are then used for GRPO policy optimization. The training process evolves from incorrect labels with invalid evidence, to correct labels with mixed evidence quality, and finally to jointly correct labels and reliable evidence.

Specifically, the predicted and reference intervals are represented as $\hat{t}_{ij} = (\hat{s}_{ij}, \hat{e}_{ij})$ and $t_i^* = (s_i^*, e_i^*)$, respectively, where s and e denote the start and end timestamps.

$$Q_{\text{tem}} = \text{IoU}(\hat{t}_{ij}, t_i^*) = \frac{|[\hat{s}_{ij}, \hat{e}_{ij}] \cap [s_i^*, e_i^*]|}{|[\hat{s}_{ij}, \hat{e}_{ij}] \cup [s_i^*, e_i^*]|}. \tag{2}$$

To enable a fair comparison with the binary semantic consistency scores used for textual explanations, we further adopt an IoU threshold of 0.7 to determine whether the temporal evidence is considered correct.

Our automated data construction pipeline provides the following advantages:

1. Our pipeline provides both manipulated temporal intervals and semantic descriptions as meta-detection evidence while supporting fully automated and scalable data generation.
2. Manipulated temporal intervals are derived directly from the controlled construction process, yielding reliable and auditable ground-truth signals that reduce reward-hacking risk.
3. Unlike fake videos independently generated from real videos, our paired real-fake videos share identical source content and temporal contexts, minimizing semantic shortcuts caused by content-level discrepancies.
4. Unlike approaches that generate entire fake videos, our method introduces localized and controllable manipulations within authentic videos, making the constructed forgeries better reflect complex real-world video forgery scenarios.

4 Evidence-Guided Reward Redistribution

In this section, we first motivate **Evidence-Guided Reward Reallocation (EGRR)** and present its algorithm, followed by an analysis of its theoretical properties.

4.1 Motivation and Design of EGRR

Label-level RLVR assigns feedback based solely on prediction correctness. Although such binary rewards provide reliable supervision for the final decision, they are too coarse-grained to distinguish whether a detector genuinely identifies visual forgery artifacts or reaches the correct label through superficial shortcuts. As a result, responses with different evidence quality may receive identical rewards, which can encourage reward hacking and limit generalization to out-of-distribution data.

Meta-detection provides an additional signal for resolving this ambiguity by evaluating whether the generated evidence genuinely supports the predicted label. This enables us to distinguish label-correct

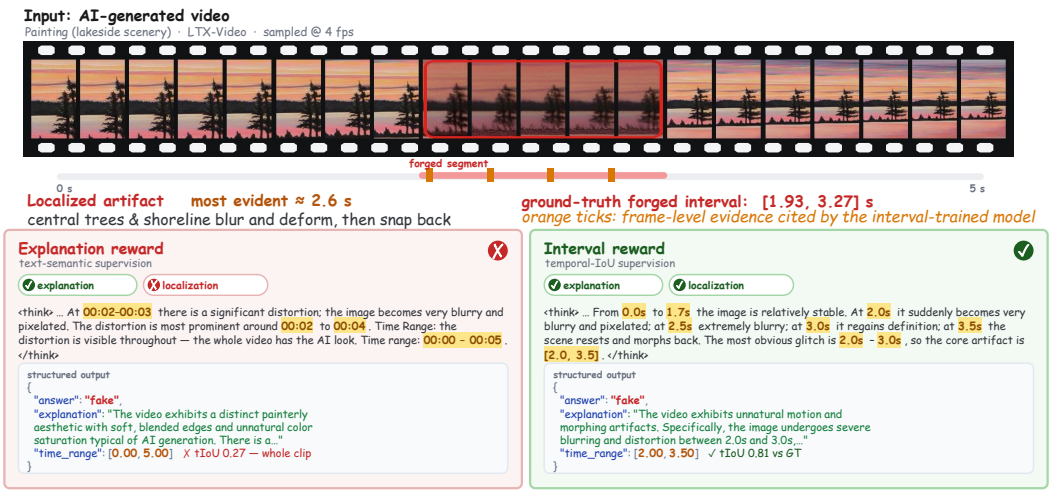

Figure 4: Comparison of detector outputs trained with two types of meta-detection feedback signals. The left and right columns show results from textual explanation-based and temporal grounding-based training, respectively. The detector trained with textual explanations fails to generalize to temporal localization, whereas the detector trained with temporal grounding produces both accurate temporal localization and reasonable textual explanations.

responses according to the quality of their supporting evidence. Rather than directly adding an independent evidence reward, we use evidence quality to redistribute the existing reward among label-correct responses. The design follows three principles: (1) label-incorrect responses remain unchanged and receive no additional evidence-based credit; (2) the total reward assigned to label-correct responses is preserved; and (3) when no label-correct response contains valid evidence, EGRR naturally degenerates to the original label-level RLVR objective. Under these constraints, evidence quality only determines the relative allocation of reward within the label-correct subset, favoring responses supported by stronger evidence without increasing the aggregate reward assigned to correct predictions.

Based on this principle, we propose **Evidence-Guided Reward Redistribution (EGRR)**. Since forgery evidence is only defined for fake videos, reward redistribution is applied only to samples with $y_i = \text{fake}$. For each sampled response o_{ij}, we define the label correctness indicator as $\ell_{ij} = \mathbb{I}[\hat{y}_{ij} = y_i] \in \{0, 1\}$ and the evidence correctness indicator as $e_{ij} \in \{0, 1\}$, where $e_{ij} = 1$ indicates that the generated evidence is sufficiently reliable to support the predicted label.

For the i-th fake video, we compute the average evidence quality among label-correct responses as

$$q_i = \begin{cases} \dfrac{\sum_{j=1}^{G} \ell_{ij} e_{ij}}{\sum_{j=1}^{G} \ell_{ij}}, & \text{if } \sum_{j=1}^{G} \ell_{ij} > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The term $\ell_{ij}(e_{ij} - q_i)$ then measures the relative evidence quality of response o_{ij} within the label-correct subset. Responses with above-average evidence receive positive adjustments, whereas those with weaker evidence are downweighted. Because the centered adjustment sums to zero over label-correct responses, EGRR preserves their aggregate reward while reallocating credit toward responses with stronger evidence. The final reward is defined as

$$R^{(i,j)} = \lambda_{\text{format}} R^{(i,j)}_{\text{format}} + (1 - \lambda_{\text{format}}) \cdot \ell_{ij} \left(1 + \mathbb{I}[y_i = \text{fake}](e_{ij} - q_i)\right). \tag{4}$$

4.2 Theoretical Advantages of EGRR

We establish two properties of EGRR: evidence-based ranking among label-correct responses, and preservation of label discrimination when evidence verification fails. We focus on fake videos, for which evidence-based redistribution is applied, and analyze the semantic reward with the unchanged

Table 1: Performance comparison on ViF-Bench in terms of Accuracy, Recall, and F1 score after removing generator-family overlap with the training data.

Method	Metric	CogV-X1.5	HunyuanVideo		Gen4-Turbo	Hailuo-02	Pika-V2	PixVerse-V4.5	Kling-V1	Sora-2	Avg.
		T2V	T2V	I2V	I2V	T2V	T2V	T2V	T2V	T2V	
Qwen3.7-Plus	Acc	67.48	57.72	52.45	54.95	54.41	67.33	77.33	55.00	55.74	60.27
	R	40.49	20.99	10.43	17.12	14.71	40.00	60.67	15.71	17.57	26.41
	F1	55.46	33.17	17.99	27.54	24.39	55.05	72.80	25.88	28.42	37.86
GPT-5.5	Acc	58.28	57.72	50.92	50.90	53.31	79.33	78.67	51.79	50.68	59.07
	R	17.79	16.67	3.07	3.60	8.09	59.33	58.67	5.00	2.70	19.44
	F1	29.90	28.27	5.88	6.84	14.77	74.17	73.33	9.40	5.19	27.53
DeepTraceReward	Acc	49.69	52.78	48.77	51.35	52.57	48.67	51.67	51.07	51.35	50.88
	R	93.25	99.38	91.41	95.50	97.79	90.67	97.33	95.00	96.62	95.22
	F1	64.96	67.79	64.09	66.25	67.34	63.85	66.82	66.00	66.51	65.96
BusterX++	Acc	61.96	59.26	50.61	55.86	56.99	62.33	65.33	56.43	52.36	57.90
	R	38.65	33.33	15.95	27.93	28.68	39.33	45.33	28.57	20.95	30.97
	F1	50.40	45.00	24.41	38.75	40.00	51.08	56.67	39.60	30.54	41.83
Qwen3.5-9B	Acc	44.24	45.12	43.33	43.30	45.62	51.99	54.93	45.04	42.67	46.25
	R	6.67	6.10	0.61	0.89	5.11	17.22	23.68	3.55	0.67	7.17
	F1	10.68	10.00	1.06	1.55	8.59	26.40	34.44	6.07	1.16	11.11
+ Label-RL	Acc	76.98	69.33	68.57	72.30	74.66	75.82	75.14	72.87	69.49	72.80
	R	67.07	61.36	60.78	63.63	61.84	67.87	66.54	65.29	60.61	63.89
	F1	74.45	66.67	65.91	69.67	70.93	73.73	72.80	70.64	66.52	70.15
+ Label-Exp-RL	Acc	78.90	75.51	76.24	73.92	75.05	80.54	79.32	72.30	73.30	76.12
	R	70.36	68.31	69.03	65.07	67.93	70.55	69.70	64.23	63.02	67.58
	F1	76.93	73.61	74.39	71.39	73.14	78.38	77.12	69.87	70.24	73.90
+ Label-Tem-RL	Acc	86.19	83.51	82.78	84.48	82.51	81.34	84.98	81.37	81.44	83.18
	R	75.98	72.94	72.00	72.96	71.19	71.10	77.15	71.20	71.25	72.86
	F1	84.62	81.56	80.70	82.46	80.28	79.21	83.70	79.26	79.33	81.24

format term omitted. Let $\mathcal{C}_i = \{j \mid \ell_{ij} = 1\}$ and $M_i = |\mathcal{C}_i|$. The EGRR semantic reward is $r_{ij}^{\mathrm{EGRR}} = \ell_{ij}(1 + e_{ij} - q_i)$, where $e_{ij} \in [0, 1]$ and q_i is the mean evidence score within $\mathcal{C}_i$ when $M_i > 0$, and zero otherwise.

Evidence-Based Ranking of Label-Correct Responses. Label-level RL assigns the same reward to all correct predictions, regardless of the quality of their supporting evidence. It therefore cannot distinguish a correct prediction supported by valid evidence from one with unreliable evidence. EGRR resolves this ambiguity by converting evidence scores into relative rewards. For any $j, k \in \mathcal{C}_i$,

$$r_{ij}^{\mathrm{EGRR}} - r_{ik}^{\mathrm{EGRR}} = e_{ij} - e_{ik}. \tag{5}$$

Thus, EGRR exactly preserves the ordering induced by evidence quality: among label-correct responses, stronger evidence receives strictly higher reward. Meanwhile, the centered adjustments satisfy $\sum_{j \in \mathcal{C}_i}(e_{ij} - q_i) = 0$. Evidence-based ranking is therefore introduced by redistributing the existing label-reward mass rather than increasing it.

Avoiding Reward Collapse under Multiplicative Coupling. A naive multiplicative joint reward for fake videos is $r_{ij}^{\mathrm{joint}} = \ell_{ij} e_{ij}$. Under this formulation, a label-correct response with invalid evidence receives the same zero reward as a label-incorrect response. Evidence failure therefore removes the reward distinction between these two cases.

EGRR preserves this distinction. Label-incorrect responses receive zero semantic reward, whereas every label-correct response satisfies

$$r_{ij}^{\mathrm{EGRR}} = 1 + e_{ij} - \frac{e_{ij} + \sum_{k \in \mathcal{C}_i \setminus \{j\}} e_{ik}}{M_i} \geq \frac{1}{M_i} > 0. \tag{6}$$

The bound follows from $e_{ij} \geq 0$ and $e_{ik} \leq 1$. Consequently, correct labels retain strictly higher semantic rewards than incorrect labels, even when their evidence is invalid.

In particular, if every label-correct response has evidence score zero, then $q_i = 0$, yielding $r_{ij}^{\mathrm{joint}} = 0$ and $r_{ij}^{\mathrm{EGRR}} = \ell_{ij}$ for all j. Multiplicative coupling collapses all semantic rewards to zero, while EGRR recovers label-level RLVR. In this case, whenever the group contains both correct and incorrect

Table 2: Performance on the fake-only subset of GenBuster-Bench after removing generator-family overlap with the training data, focusing on the capability to identify AI-generated videos.

Method	Sora	Pika	Gen3	Luma	Kling	Jimeng	Vidu	Avg.
Qwen3.7-Plus	75.5	93.0	89.0	99.0	86.0	75.0	88.7	86.6
GPT-5.5	61.0	89.0	92.0	99.0	70.0	50.0	96.7	79.7
DeepTraceReward	83.0	90.0	89.0	91.0	86.0	79.0	93.3	87.3
BusterX++	71.0	92.0	83.0	90.0	63.0	78.0	87.3	80.6
Qwen3.5-9B	36.0	60.0	66.0	77.0	53.0	46.0	58.7	56.7
+ Label-RL	76.5 +40.5	90.0 +30.0	91.0 +25.0	86.0 +9.0	78.0 +25.0	79.0 +33.0	80.7 +22.0	83.0 +26.3
+ Label-Exp-RL	78.6 +42.6	91.6 +31.6	88.6 +22.6	93.3 +16.3	82.6 +29.6	85.6 +39.6	83.9 +25.2	86.3 +29.6
+ Label-Tem-RL	**89.7** +53.7	**95.8** +35.8	**93.6** +27.6	**99.5** +22.5	**90.0** +37.0	**87.2** +41.2	**97.3** +38.6	**93.3** +36.6

labels, i.e., $0 < M_i < G$, EGRR retains nonzero semantic reward variance: $\mathrm{Var}_j(r_{ij}^{\mathrm{EGRR}}) = \frac{M_i}{G}(1 - \frac{M_i}{G}) > 0$. EGRR therefore maintains a label-discriminative reward signal when evidence verification fails, while using evidence quality to refine the ranking whenever it distinguishes label-correct responses.

5 Experiments

5.1 Experiment Setup

Implementation Details. We use Qwen3.5-9B (Qwen Team, 2026) as the base detector and uniformly sample video frames at 4 FPS. To construct diverse fake videos, we employ open-source video generation models with different scales, including LTX-Video-2B (HaCohen et al., 2024), Wan2.2-Fun-5B-InP (Wan et al., 2025), and SkyReels-V2-DF-14B-540P-Diffusers (Chen et al., 2025), to generate 45K fake videos. In addition, we incorporate 5K high-fidelity challenging fake videos generated by closed-source models, including Wan2.7-I2V (Alibaba Cloud, 2026) and Seedance1.0-Pro (BytePlus, 2025). Together with the corresponding real videos, this process results in a balanced dataset containing 100K samples. For textual evidence construction, we leverage GPT-5.5 (OpenAI, 2025) to generate descriptions of observable forgery artifacts for each fake video. We further use Gemini-3.1-Pro (Google DeepMind, 2026) to filter out samples with unreliable textual explanations and remove their corresponding real videos to maintain a 1:1 ratio between real and fake samples. During evaluation, Qwen3.5-4B (Qwen Team, 2026) is employed to measure the semantic consistency between detector-generated explanations and reference textual explanations. All experiments in this setting are conducted on the resulting filtered subset. To ensure a fair comparison between temporal evidence and textual explanations, we convert temporal grounding quality into a binary feedback signal using an IoU threshold of 0.7, consistent with the binary nature of textual evidence evaluation. We optimize the detector with the DAPO (Yu et al., 2026) algorithm using a learning rate of 1×10^{-6}. All experiments are conducted on 16 NVIDIA H200 GPUs for one training epoch.

Evaluation Metrics and Comparison Methods. We compare our method with strong proprietary MLLMs, including Qwen3.7-Plus (Qwen Team, 2026) and GPT-5.5 (OpenAI, 2025), as well as representative open-source MLLM-based detectors, including DeepTraceReward (Fu et al., 2025b) and BusterX++ (Wen et al., 2025a). To specifically evaluate out-of-distribution generalization, we exclude benchmark subsets associated with generator families that overlap with the training data. On the resulting cross-generator subset of ViF-Bench (Li et al., 2025), we report Accuracy, F1 score, and Recall. We further evaluate Recall on the corresponding fake-only subset of GenBuster-Bench (Wen et al., 2025a), which is critical for reducing false negatives in real-world scenarios. All reported results are averaged over three independent runs with different random seeds.

5.2 Main Results

Analysis of Meta-Detection Effectiveness. To validate the effectiveness of meta-detection, we compare detectors trained with and without evidence-aware feedback. As shown in Tables 1 and 2, as well as Fig. 5, detectors trained with meta-detection consistently outperform conventional label-level reinforcement learning. For example, on ViF-Bench, Label-Tem-RL improves over Label-RL by

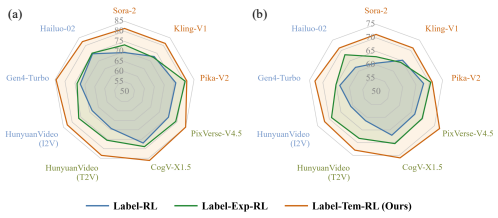

Figure 5: Performance on the cross-generator subset of ViF-Bench. (a) Accuracy . (b) Recall.

Table 3: Evaluation of Temporal Localization Capability on Five In-Domain Datasets.

Method	LTX-2B	Wan2.2	SkyReels-V2	Wan2.7	Seedance1.0
Qwen3.5-9B	11.5	8.0	9.5	6.5	5.5
+ Label-RL	21.0	17.5	19.5	14.5	13.0
+ Label-Exp-RL	26.5	24.0	23.5	18.0	18.5
+ Label-Tem-RL	79.5	73.0	77.5	66.0	61.5

10.38%, 8.97%, and 11.09% in Accuracy, Recall, and F1 score, respectively. On the fake-only subset of GenBuster-Bench, Label-Tem-RL further improves average Recall by 10.3%. The radar chart in Fig. 5 provides an intuitive visualization of these consistent gains across evaluation settings. These results demonstrate that evidence-aware reward redistribution among label-correct responses encourages detectors to learn fine-grained forgery artifacts, leading to more robust and generalizable AI-generated video detection.

Temporal Grounding Outperforms Textual Explanations as Meta-Detection Feedback. To ensure a fair comparison between temporal grounding and textual explanations, we convert temporal evidence into binary feedback using an IoU threshold of 0.7, matching the binary nature of textual evidence evaluation. As shown in Tables 1 and 2, detectors trained with temporal grounding consistently outperform those trained with textual explanations. On ViF-Bench, Label-Tem-RL improves over Label-Exp-RL by 7.06%, 5.28%, and 7.34% in Accuracy, Recall, and F1 score, respectively. On the fake-only subset of GenBuster-Bench, temporal grounding achieves improvements of 7.0% and 11.4% on OOD and Wild evaluation, respectively. These results demonstrate that rule-based temporal grounding provides more reliable and generalizable evidence supervision for meta-detection than model-based textual explanations. As shown in Fig. 4, the model trained with textual supervision can describe general forgery artifacts but fails to localize the manipulated temporal interval. In contrast, temporal grounding supervision enables accurate time-range prediction while retaining the ability to generate meaningful textual explanations, further demonstrating the superiority of temporal grounding as a reliable evidence source for meta-detection.

Evaluation of Temporal Localization Capability. To directly evaluate the temporal localization capability of the detector, we construct a held-out in-domain test set containing 1,000 fake videos from five generators, with 200 samples per generator. We compute the temporal intersection-over-union (tIoU) between the predicted and ground-truth forgery intervals and adopt the same threshold of 0.7 as in training to determine whether the manipulated segment is successfully localized. As shown in Table 3, Label-Tem-RL substantially improves temporal localization across all five generators. This result indicates that temporal-grounding supervision enables the detector to move beyond label-level discrimination and identify the specific video segments that support its predictions. Such fine-grained localization provides more interpretable forensic evidence and facilitates subsequent verification of model decisions.

6 Conclusion

In this paper, we introduce the concept of **meta-detection** into AI-generated video detection, extending reinforcement learning beyond label correctness to jointly evaluate prediction correctness and evidence validity. We develop an automated and scalable data construction pipeline that generates paired real-fake videos while deriving ground-truth forgery temporal intervals from controlled manipulation processes. We further demonstrate that rule-based temporal grounding provides more reliable and verifiable supervision than model-based textual explanations, making it a more suitable evidence source for meta-detection. Moreover, we propose **Evidence-Guided Reward Redistribution (EGRR)**, which calibrates label-level rewards by redistributing credits among label-correct responses according to evidence quality, enabling effective integration of meta-detection feedback into reinforcement learning. Extensive experiments demonstrate that **VidForensics-M1** learns fine-grained visual forgery artifacts from temporal grounding evidence, leading to more robust and generalizable AI-generated video detection.

AI USE STATEMENT

In this work, we used generative AI tools for manuscript polishing and Codex for code-assisted programming. All AI-assisted code was manually reviewed for logical correctness, and all AI-polished text was checked sentence by sentence by the authors. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

VidForensics-M1 aims to improve the reliability and interpretability of AI-generated video detection by predicting authenticity labels together with supporting forgery evidence. We expect these capabilities to support research in multimedia forensics, content provenance, and the responsible deployment of generative video technologies. However, predictions from VidForensics-M1 should not be treated as definitive evidence of authenticity or manipulation and should not serve as the sole basis for high-stakes decisions such as legal judgments, journalistic attribution, or punitive content moderation. We recommend using VidForensics-M1 primarily as an assistive forensic tool alongside provenance information, human review, and complementary verification mechanisms.

REPRODUCIBILITY STATEMENT

Our code is included in the supplementary material, and the implementation details are described in Section 5.1.

REFERENCES

Alibaba Cloud. Wan 2.7: Image-to-video api. https://www.alibabacloud.com/help/en/model-studio/image-to-video-general-api-reference, 2026.

Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in ai safety, 2016. URL https://arxiv.org/abs/1606.06565.

Jianfa Bai, Man Lin, Gang Cao, and Zijie Lou. Ai-generated video detection via spatial-temporal anomaly learning. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pp. 460–470. Springer, 2024.

BytePlus. Seedance 1.0 Pro. https://docs.byteplus.com/en/docs/ModelArk/1587798, 2025.

Bilva Chandra, Jesse Dunietz, Kathleen Roberts, Yooyoung Lee, Peter Fontana, and George Awad. Reducing risks posed by synthetic content an overview of technical approaches to digital content transparency. 2024.

Guibin Chen, Dixuan Lin, Jiangping Yang, Chunze Lin, Junchen Zhu, Mingyuan Fan, Hao Zhang, Sheng Chen, Zheng Chen, Chengcheng Ma, Weiming Xiong, Wei Wang, Nuo Pang, Kang Kang, Zhiheng Xu, Yuzhe Jin, Yupeng Liang, Yubing Song, Peng Zhao, Boyuan Xu, Di Qiu, Debang Li, Zhengcong Fei, Yang Li, and Yahui Zhou. Skyreels-v2: Infinite-length film generative model, 2025. URL https://arxiv.org/abs/2504.13074.

Haoxing Chen, Yan Hong, Zizheng Huang, Zhuoer Xu, Zhangxuan Gu, Yaohui Li, Jun Lan, Huijia Zhu, Jianfu Zhang, Weiqiang Wang, et al. Demamba: Ai-generated video detection on million-scale genvideo benchmark. *arXiv preprint arXiv:2405.19707*, 2024.

Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 4302–4310, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Riccardo Corvi, Davide Cozzolino, Ekta Prashnani, Shalini De Mello, Koki Nagano, and Luisa Verdoliva. Seeing what matters: Generalizable ai-generated video detection with forensic-oriented augmentation. *arXiv preprint arXiv:2506.16802*, 2025.

Catherine E Davodi, Sarah Barrington, Hany Farid, and Emily A Cooper. Perceptual judgments of video authenticity: An examination of viewing duration, confidence, content, and strategies. *Law Review*, 107:1753–1819, 2026.

Xinghe Fu, Zhiyuan Yan, Taiping Yao, Shen Chen, and Xi Li. Exploring unbiased deepfake detection via token-level shuffling and mixing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 3040–3048, 2025a.

Xingyu Fu, Siyi Liu, Yinuo Xu, Pan Lu, Guangqiuse Hu, Tianbo Yang, Taran Anantasagar, Christopher Shen, Yikai Mao, Yuanzhe Liu, et al. Learning human-perceived fakeness in ai-generated videos via multimodal llms. *arXiv preprint arXiv:2509.22646*, 2025b.

Xingyu Fu, Siyi Liu, Yinuo Xu, Pan Lu, Guangqiuse Hu, Tianbo Yang, Taran Anantasagar, Christopher Shen, Yikai Mao, Yuanzhe Liu, et al. Learning human-perceived fakeness in ai-generated videos via multimodal llms. *arXiv preprint arXiv:2509.22646*, 2025c.

Yifeng Gao, Yifan Ding, Hongyu Su, Juncheng Li, Yunhan Zhao, Lin Luo, Zixing Chen, Li Wang, Xin Wang, Yixu Wang, et al. David-xr1: Detecting ai-generated videos with explainable reasoning. *arXiv preprint arXiv:2506.14827*, 2025.

Google DeepMind. Gemini 3.1 Pro model card. https://deepmind.google/models/model-cards/gemini-3-1-pro/, 2026.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao Wu, Bei Feng, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chong Ruan, Damai Dai, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen, Guowei Li, H. Zhang, Hanwei Xu, Honghui Ding, Huazuo Gao, Hui Qu, Hui Li, Jianzhong Guo, Jiashi Li, Jingchang Chen, Jingyang Yuan, Jinhao Tu, Junjie Qiu, Junlong Li, J. L. Cai, Jiaqi Ni, Jian Liang, Jin Chen, Kai Dong, Kai Hu, Kaichao You, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Liang Zhao, Litong Wang, Liyue Zhang, Lei Xu, Leyi Xia, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Mingxu Zhou, Meng Li, Miaojun Wang, Mingming Li, Ning Tian, Panpan Huang, Peng Zhang, Qiancheng Wang, Qinyu Chen, Qiushi Du, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, R. J. Chen, R. L. Jin, Ruyi Chen, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shengfeng Ye, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, S. S. Li, Shuang Zhou, Shaoqing Wu, Tao Yun, Tian Pei, Tianyu Sun, T. Wang, Wangding Zeng, Wen Liu, Wenfeng Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, W. L. Xiao, Wei An, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xinyu Yang, Xinyuan Li, Xuecheng Su, Xuheng Lin, X. Q. Li, Xiangyue Jin, Xiaojin Shen, Xiaosha Chen, Xiaowen Sun, Xiaoxiang Wang, Xinnan Song, Xinyi Zhou, Xianzu Wang, Xinxia Shan, Y. K. Li, Y. Q. Wang, Y. X. Wei, Yang Zhang, Yanhong Xu, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Yu, Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Yishi Piao, Yisong Wang, Yixuan Tan, Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yuan Ou, Yuduan Wang, Yue Gong, Yuheng Zou, Yujia He, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou, Y. X. Zhu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Ying Tang, Yukun Zha, Yuting Yan, Z. Z. Ren, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zhewen Hao, Zhicheng Ma, Zhigang Yan, Zhiyu Wu, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin Li, Ziwei Xie, Ziyang Song, Zizheng Pan, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, and Zhen Zhang. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645 (8081):633–638, September 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-09422-z. URL http://dx.doi.org/10.1038/s41586-025-09422-z.

Yoav HaCohen, Nisan Chiprut, Benny Brazowski, Daniel Shalem, Dudu Moshe, Eitan Richardson, Eran Levin, Guy Shiran, Nir Zabari, Ori Gordon, Poriya Panet, Sapir Weissbuch, Victor Kulikov, Yaki Bitterman, Zeev Melumian, and Ofir Bibi. Ltx-video: Realtime video latent diffusion. *arXiv preprint arXiv:2501.00103*, 2024.

Yinan He, Bei Gan, Siyu Chen, Yichun Zhou, Guojun Yin, Luchuan Song, Lu Sheng, Jing Shao, and Ziwei Liu. Forgerynet: A versatile benchmark for comprehensive forgery analysis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4360–4369, 2021.

Dylan Xinming Hou, Juntian Zhang, Xu Gu, Yichen Wu, Nils Lukas, Gus Xia, Xiuying Chen, and Yuhan Liu. Detecting ai-generated video: A vision–language dual-view survey. In *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 32221–32255, 2026.

Lingen Li, Guangzhi Wang, Xiaoyu Li, Zhaoyang Zhang, Qi Dou, Jinwei Gu, Tianfan Xue, and Ying Shan. Cubecomposer: Spatio-temporal autoregressive 4k 360° video generation from perspective video. *arXiv e-prints*, pp. arXiv–2603, 2026.

Yifei Li, Wenzhao Zheng, Yanran Zhang, Runze Sun, Yu Zheng, Lei Chen, Jie Zhou, and Jiwen Lu. Skyra: Ai-generated video detection via grounded artifact reasoning. *arXiv preprint arXiv:2512.15693*, 2025.

Long Ma, Zhiyuan Yan, Qinglang Guo, Yong Liao, Haiyang Yu, and Pengyuan Zhou. Detecting ai-generated video via frame consistency. In *2025 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. IEEE, 2025.

Dat Nguyen, Nesryne Mejri, Inder Pal Singh, Polina Kuleshova, Marcella Astrid, Anis Kacem, Enjie Ghorbel, and Djamila Aouada. Laa-net: Localized artifact attention network for quality-agnostic and generalizable deepfake detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17395–17405, 2024.

Zhenliang Ni, Qiangyu Yan, Mouxiao Huang, Tianning Yuan, Yehui Tang, Hailin Hu, Xinghao Chen, and Yunhe Wang. Genvidbench: A 6-million benchmark for ai-generated video detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 15582–15590, 2026.

Utkarsh Ojha, Yuheng Li, and Yong Jae Lee. Towards universal fake image detectors that generalize across generative models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 24480–24489, 2023.

OpenAI. Introducing gpt-5. https://openai.com/index/introducing-gpt-5/, 2025.

Kyoungjun Park, Yifan Yang, Juheon Yi, Shicheng Zheng, Yifei Shen, Dongqi Han, Caihua Shan, Muhammad Muaz, and Lili Qiu. Vidguard-r1: Ai-generated video detection and explanation via reasoning mllms and rl. *arXiv preprint arXiv:2510.02282*, 2025.

Qwen Team. Qwen3.5: Towards native multimodal agents. https://qwen.ai/blog?id=qwen3.5, February 2026.

Chuangchuang Tan, Yao Zhao, Shikui Wei, Guanghua Gu, Ping Liu, and Yunchao Wei. Rethinking the up-sampling operations in cnn-based generative network for generalizable deepfake detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 28130–28139, 2024.

Hao Tan, Senyuan Shi, Zichang Tan, Zijian Yu, Huijia Zhu, Weiqiang Wang, Jun Wan, Zhen Lei, et al. Videoveritas: Ai-generated video detection via perception pretext reinforcement learning. In *Forty-third International Conference on Machine Learning*, 2026.

Miles Turpin, Julian Michael, Ethan Perez, and Samuel Bowman. Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Advances in Neural Information Processing Systems*, 36:74952–74965, 2023.

Cristian Vaccari and Andrew Chadwick. Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social media+ society*, 6(1): 2056305120903408, 2020.

Team Wan, Ang Wang, Baole Ai, Bin Wen, Chaojie Mao, Chen-Wei Xie, Di Chen, Feiwu Yu, Haiming Zhao, Jianxiao Yang, Jianyuan Zeng, Jiayu Wang, Jingfeng Zhang, Jingren Zhou, Jinkai Wang, Jixuan Chen, Kai Zhu, Kang Zhao, Keyu Yan, Lianghua Huang, Mengyang Feng, Ningyi Zhang, Pandeng Li, Pingyu Wu, Ruihang Chu, Ruili Feng, Shiwei Zhang, Siyang Sun, Tao Fang, Tianxing Wang, Tianyi Gui, Tingyu Weng, Tong Shen, Wei Lin, Wei Wang, Wei Wang, Wenmeng Zhou, Wente Wang, Wenting Shen, Wenyuan Yu, Xianzhong Shi, Xiaoming Huang, Xin Xu, Yan Kou, Yangyu Lv, Yifei Li, Yijing Liu, Yiming Wang, Yingya Zhang, Yitong Huang, Yong Li, You Wu, Yu Liu, Yulin Pan, Yun Zheng, Yuntao Hong, Yupeng Shi, Yutong Feng, Zeyinzi Jiang, Zhen Han, Zhi-Fan Wu, and Ziyu Liu. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*, 2025.

Haibo Wang, Zhiyang Xu, Yu Cheng, Shizhe Diao, Yufan Zhou, Yixin Cao, Qifan Wang, Weifeng Ge, and Lifu Huang. Grounded-VideoLLM: Sharpening fine-grained temporal grounding in video large language models. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (eds.), *Findings of the Association for Computational Linguistics: EMNLP 2025*, pp. 959–975, Suzhou, China, November 2025. Association for Computational Linguistics. ISBN 979-8-89176-335-7. doi: 10.18653/v1/2025.findings-emnlp.50. URL https://aclanthology.org/2025.findings-emnlp.50/.

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

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 24824–24837. Curran Associates, Inc., 2022. doi: 10.52202/068431-1800. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf.

Haiquan Wen, Yiwei He, Zhenglin Huang, Tianxiao Li, Zihan Yu, Xingru Huang, Lu Qi, Baoyuan Wu, Xiangtai Li, and Guangliang Cheng. Busterx: Mllm-powered ai-generated video forgery detection and explanation. *arXiv preprint arXiv:2505.12620*, 2025a.

Haiquan Wen, Tianxiao Li, Zhenglin Huang, Yiwei He, and Guangliang Cheng. Busterx++: Towards unified cross-modal ai-generated content detection and explanation with mllm. *arXiv preprint arXiv:2507.14632*, 2025b.

Tianwei Xiong, Jun Hao Liew, Zilong Huang, Zhijie Lin, Jiashi Feng, and Xihui Liu. Evatok: Adaptive length video tokenization for efficient visual autoregressive generation. *arXiv preprint arXiv:2603.12267*, 2026.

Xuancheng Xu, Yaning Li, Sisi You, and Bing-Kun Bao. Smrabooth: Subject and motion representation alignment for customized video generation. *arXiv preprint arXiv:2512.12193*, 2025a.

Zunnan Xu, Zhentao Yu, Zixiang Zhou, Jun Zhou, Xiaoyu Jin, Fa-Ting Hong, Xiaozhong Ji, Junwei Zhu, Chengfei Cai, Shiyu Tang, et al. Hunyuanportrait: Implicit condition control for enhanced portrait animation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15909–15919. IEEE, 2025b.

Shilin Yan, Ouxiang Li, Jiayin Cai, Yanbin Hao, Xiaolong Jiang, Yao Hu, and Weidi Xie. A sanity check for ai-generated image detection. In *International Conference on Learning Representations*, volume 2025, pp. 70702–70720, 2025.

Zhiyuan Yan, Yuhao Luo, Siwei Lyu, Qingshan Liu, and Baoyuan Wu. Transcending forgery specificity with latent space augmentation for generalizable deepfake detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8984–8994, 2024.

Zheng Yang, Ruoxin Chen, Zhiyuan Yan, Ke-Yue Zhang, Xinghe Fu, Shuang Wu, Xiujun Shu, Taiping Yao, Shouhong Ding, Zequn Qin, et al. All patches matter, more patches better: Enhance ai-generated image detection via panoptic patch learning. *arXiv preprint arXiv:2504.01396*, 2025.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, et al. Dapo: An open-source llm reinforcement learning system at scale. *Advances in Neural Information Processing Systems*, 38:113222–113244, 2026.

Ruicheng Zhang, Jun Zhou, Zunnan Xu, Zihao Liu, Jiehui Huang, Mingyang Zhang, Yu Sun, and Xiu Li. Zo3t: Zero-shot 3d-aware trajectory-guided image-to-video generation via test-time training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12708–12716, 2026.

A Theoretical Analysis of Evidence-Guided Reward Redistribution

This section establishes the two properties of Evidence-Guided Reward Redistribution (EGRR) discussed in the main text. First, EGRR distinguishes label-correct responses according to their evidence quality while conserving the original label-reward mass. Second, EGRR preserves reward separation between correct and incorrect labels when evidence verification fails, avoiding the semantic reward collapse caused by multiplicative coupling. All results are exact finite-sample statements and require no independence assumptions on the sampled responses.

A.1 Preliminaries

For a training video V_i, consider a group of $G \geq 1$ responses $\{o_{ij}\}_{j=1}^{G}$. Let $\ell_{ij} = \mathbb{I}[\hat{y}_{ij} = y_i] \in \{0, 1\}$ denote label correctness, and let $e_{ij} \in [0, 1]$ denote evidence quality. This formulation covers both continuous evidence scores and the binary evidence correctness used in our experiments.

Define $z_i = \mathbb{I}[y_i = \text{fake}]$, $\mathcal{C}_i = \{j : \ell_{ij} = 1\}$, and $M_i = |\mathcal{C}_i|$. The group-relative evidence baseline is

$$q_i = \begin{cases} \dfrac{1}{M_i} \displaystyle\sum_{j \in \mathcal{C}_i} e_{ij}, & M_i > 0, \\ 0, & M_i = 0. \end{cases} \tag{7}$$

For real videos, we set the unused evidence scores to zero.

The semantic reward of EGRR is

$$r_{ij}^{\text{EGRR}} = \ell_{ij}\big[1 + z_i(e_{ij} - q_i)\big] = \ell_{ij} + \Delta_{ij}, \qquad \Delta_{ij} = z_i \ell_{ij}(e_{ij} - q_i). \tag{8}$$

The label-only semantic reward is $r_{ij}^{\text{label}} = \ell_{ij}$. For comparison, we consider the multiplicative joint reward $r_{ij}^{\text{joint}} = \ell_{ij}[(1 - z_i) + z_i e_{ij}]$, which reduces to $\ell_{ij} e_{ij}$ for fake videos.

The complete reward is $R_{ij} = \lambda_{\text{format}} R_{ij}^{\text{format}} + (1 - \lambda_{\text{format}}) r_{ij}$, where $0 \leq \lambda_{\text{format}} < 1$. We analyze semantic rewards below; the format component is unchanged across the compared formulations.

A.2 Evidence-Based Ranking of Label-Correct Responses

We first establish reward conservation, and then show that EGRR converts differences in evidence quality into exact pairwise reward differences.

Lemma 1 (Exact finite-sample reward conservation)**.** *For every sampled group,*

$$\sum_{j=1}^{G} \Delta_{ij} = 0, \qquad \sum_{j=1}^{G} r_{ij}^{\text{EGRR}} = M_i. \tag{9}$$

If $M_i > 0$, the average semantic reward among label-correct responses is exactly one:

$$\frac{1}{M_i} \sum_{j \in \mathcal{C}_i} r_{ij}^{\text{EGRR}} = 1. \tag{10}$$

Proof. If $M_i = 0$, every correctness indicator is zero. Consequently, every adjustment and semantic reward is zero, and Eq. (9) holds immediately.

Suppose $M_i > 0$. Since $\Delta_{ij} = 0$ outside $\mathcal{C}_i$, the definition of q_i gives

$$\sum_{j=1}^{G} \Delta_{ij} = z_i \sum_{j \in \mathcal{C}_i} (e_{ij} - q_i) \tag{11}$$

$$= z_i \left(\sum_{j \in \mathcal{C}_i} e_{ij} - M_i q_i \right) = 0. \tag{12}$$

Summing Eq. (8) therefore yields $\sum_{j=1}^{G} r_{ij}^{\mathrm{EGRR}} = \sum_{j=1}^{G} \ell_{ij} = M_i$. Because label-incorrect responses receive zero semantic reward, the same total is assigned to $\mathcal{C}_i$. Dividing by M_i proves Eq. (10). □

Theorem 1 (Exact evidence-order preservation)**.** *For any two label-correct responses $j, k \in \mathcal{C}_i$ from the same fake video,*

$$r_{ij}^{\mathrm{EGRR}} - r_{ik}^{\mathrm{EGRR}} = e_{ij} - e_{ik}. \tag{13}$$

Hence, $r_{ij}^{\mathrm{EGRR}} > r_{ik}^{\mathrm{EGRR}}$ if and only if $e_{ij} > e_{ik}$. Equal evidence scores receive equal semantic rewards.

Proof. For a fake video, $z_i = 1$. Since $j, k \in \mathcal{C}_i$, both correctness indicators equal one, and the two responses share the same baseline q_i. Therefore,

$$r_{ij}^{\mathrm{EGRR}} - r_{ik}^{\mathrm{EGRR}} = (1 + e_{ij} - q_i) - (1 + e_{ik} - q_i) \tag{14}$$

$$= e_{ij} - e_{ik}. \tag{15}$$

The ordering and equality statements follow directly. □

Under label-only RLVR, the same pair receives identical rewards, $r_{ij}^{\mathrm{label}} = r_{ik}^{\mathrm{label}} = 1$, even when their evidence scores differ. Theorem 1 shows that EGRR resolves this ambiguity by introducing a reward margin equal to the difference in evidence quality. Together with Lemma 1, this establishes that evidence-based ranking is achieved by redistributing the existing label-reward mass.

Corollary 1 (Discrimination when all labels are correct)**.** *For a fake video with $M_i = G$, label-only semantic rewards have zero variance, whereas*

$$\mathrm{Var}_j\big(r_{ij}^{\mathrm{EGRR}}\big) = \frac{1}{G}\sum_{j=1}^{G}(e_{ij} - q_i)^2. \tag{16}$$

This variance is strictly positive if and only if the evidence scores are not all identical.

Proof. When $M_i = G$, every label-only reward equals one. By Lemma 1, the mean EGRR semantic reward also equals one. Since $r_{ij}^{\mathrm{EGRR}} = 1 + e_{ij} - q_i$, its deviation from the group mean is $e_{ij} - q_i$. Using the empirical variance with divisor G proves Eq. (16). A sum of squared deviations is zero exactly when every evidence score equals q_i. □

A.3 Avoiding Reward Collapse under Multiplicative Coupling

We next show that evidence-based redistribution preserves a strict reward distinction between correct and incorrect labels, including when correct predictions have invalid evidence.

Theorem 2 (Strict label separation and bounded rewards)**.** *Every label-incorrect response receives zero semantic reward. Every label-correct response from a real video receives semantic reward one. For a fake video with $M_i \geq 1$, every $j \in \mathcal{C}_i$ satisfies*

$$\frac{1}{M_i} \leq r_{ij}^{\mathrm{EGRR}} \leq 2 - \frac{1}{M_i}. \tag{17}$$

Consequently, for any label-correct response $j \in \mathcal{C}_i$ and label-incorrect response $k \notin \mathcal{C}_i$ from the same fake video,

$$r_{ij}^{\mathrm{EGRR}} - r_{ik}^{\mathrm{EGRR}} \geq \frac{1}{M_i} > 0. \tag{18}$$

Proof. If $\ell_{ij} = 0$, Eq. (8) gives $r_{ij}^{\mathrm{EGRR}} = 0$. For real videos, $z_i = 0$, so $r_{ij}^{\mathrm{EGRR}} = \ell_{ij}$.

Now consider a fake video with $M_i \geq 1$ and $j \in \mathcal{C}_i$. Expanding the baseline gives

$$r_{ij}^{\mathrm{EGRR}} = 1 + e_{ij} - \frac{e_{ij} + \sum_{k \in \mathcal{C}_i \setminus \{j\}} e_{ik}}{M_i} \tag{19}$$

$$= 1 + \frac{M_i - 1}{M_i} e_{ij} - \frac{1}{M_i} \sum_{k \in \mathcal{C}_i \setminus \{j\}} e_{ik}. \tag{20}$$

Because $e_{ij} \in [0, 1]$ and the remaining $M_i - 1$ evidence scores also lie in $[0, 1]$,

$$r_{ij}^{\mathrm{EGRR}} \geq 1 - \frac{M_i - 1}{M_i} = \frac{1}{M_i}, \tag{21}$$

$$r_{ij}^{\mathrm{EGRR}} \leq 1 + \frac{M_i - 1}{M_i} = 2 - \frac{1}{M_i}. \tag{22}$$

These inequalities also cover $M_i = 1$, for which the sum over other responses is empty and $r_{ij}^{\mathrm{EGRR}} = 1$.

The bounds are tight: the lower bound is attained when $e_{ij} = 0$ and all other label-correct evidence scores equal one; the upper bound is attained under the reverse assignment. Finally, subtracting the zero reward of a label-incorrect response proves Eq. (18). □

For binary evidence, a label-correct response with invalid evidence receives $1 - q_i$, while one with valid evidence receives $2 - q_i$. Whenever an invalid-evidence response exists within $\mathcal{C}_i$, its inclusion in the baseline implies $q_i \leq (M_i - 1)/M_i < 1$. Its semantic reward therefore remains strictly positive. Thus, EGRR distinguishes correct predictions with invalid evidence from incorrect predictions, while assigning higher rewards to correct predictions with valid evidence.

Proposition 1 (Preservation of label discrimination under evidence failure)**.** *Consider a fake video.*

First, if $j \in \mathcal{C}_i$ has $e_{ij} = 0$ and $k \notin \mathcal{C}_i$, multiplicative coupling assigns $r_{ij}^{\mathrm{joint}} = r_{ik}^{\mathrm{joint}} = 0$, whereas EGRR strictly separates their semantic rewards.

Second, if $M_i > 0$ and every label-correct response has evidence score zero, then

$$r_{ij}^{\mathrm{joint}} = 0, \qquad r_{ij}^{\mathrm{EGRR}} = \ell_{ij} \quad \textit{for all } j. \tag{23}$$

If, additionally, $0 < M_i < G$, the empirical semantic reward variances satisfy

$$\mathrm{Var}_j\big(r_{ij}^{\mathrm{joint}}\big) = 0, \qquad \mathrm{Var}_j\big(r_{ij}^{\mathrm{EGRR}}\big) = \frac{M_i}{G}\left(1 - \frac{M_i}{G}\right) > 0. \tag{24}$$

Proof. For fake videos, $r_{ij}^{\mathrm{joint}} = \ell_{ij} e_{ij}$. A label-correct response with $e_{ij} = 0$ therefore receives zero joint reward. A label-incorrect response also receives zero joint reward, regardless of its evidence score. Their joint rewards coincide, while Theorem 2 guarantees an EGRR semantic reward gap of at least $1/M_i$.

If every label-correct response has evidence score zero, Eq. (7) gives $q_i = 0$. For label-correct responses, EGRR then assigns reward one. For label-incorrect responses, it assigns reward zero. Multiplicative coupling assigns zero to both subsets, proving Eq. (23).

To compute the variance, write $p_i = M_i/G$. The EGRR semantic rewards consist of M_i ones and $G - M_i$ zeros, with mean p_i. Hence,

$$\mathrm{Var}_j\big(r_{ij}^{\mathrm{EGRR}}\big) = \frac{1}{G}\left[M_i(1 - p_i)^2 + (G - M_i)p_i^2\right] \tag{25}$$

$$= p_i(1 - p_i)^2 + (1 - p_i)p_i^2 \tag{26}$$

$$= p_i(1 - p_i). \tag{27}$$

This quantity is strictly positive when $0 < M_i < G$. All multiplicative semantic rewards are zero, so their variance is zero. □

Under the conditions of Eq. (24), the group-centered EGRR semantic rewards are $1 - p_i > 0$ for correct labels and $-p_i < 0$ for incorrect labels. Multiplicative coupling instead produces zero centered semantic rewards for every response. EGRR therefore preserves a label-discriminative comparison when evidence verification fails.

The condition $0 < M_i < G$ is necessary for this particular nonzero-variance guarantee. If all labels are incorrect, or all labels are correct with zero evidence scores, EGRR semantic rewards are constant within the group. There is then no variation in label correctness or evidence quality to distinguish the responses.

Corollary 2 (Exact fallback under non-discriminative evidence)**.** *For a fake video with $M_i > 0$, EGRR coincides with label-only semantic rewards for every response if and only if all evidence scores within $\mathcal{C}_i$ are identical. In particular, if $e_{ij} = c$ for every $j \in \mathcal{C}_i$, where $c \in [0, 1]$, then*

$$r_{ij}^{\mathrm{EGRR}} = \ell_{ij}, \qquad r_{ij}^{\mathrm{joint}} = c\ell_{ij} \quad \textit{for all } j. \tag{28}$$

When $M_i = 0$, both semantic reward formulations are zero for every response.

Proof. If all label-correct evidence scores equal c, their mean is $q_i = c$. The centered evidence adjustment therefore vanishes on $\mathcal{C}_i$ and is zero outside $\mathcal{C}_i$ by the correctness factor. Thus, $r_{ij}^{\mathrm{EGRR}} = \ell_{ij}$ for every response. The multiplicative identity follows directly from $r_{ij}^{\mathrm{joint}} = \ell_{ij} e_{ij}$.

Conversely, suppose EGRR equals the label-only reward for every response. For each $j \in \mathcal{C}_i$, this implies $1 + e_{ij} - q_i = 1$, and hence $e_{ij} = q_i$. All label-correct evidence scores are therefore identical. The case $M_i = 0$ follows from $\ell_{ij} = 0$ for all j. □

Scope of the Guarantees. The results establish exact properties of semantic rewards on each sampled group. Evidence ordering is defined with respect to the supplied evidence scores. Because the format component is unchanged, exact fallback also holds for the complete rewards. However, varying format rewards can affect the ordering of complete rewards, and nonzero semantic reward variance does not by itself guarantee a nonzero parameter update. These results do not assert preservation of the policy gradient or monotonic improvement in detection accuracy.

B Real-Fake Video Pair Examples

The figure presents examples of the constructed real-fake video pairs. For each video generation model, two examples are provided. Each example consists of three rows: (1) the original authentic video; (2) the boundary frames provided to the video generation model and the generated synthetic segment; and (3) the final forged video obtained by aligning the generated segment with the removed content in terms of frame rate and resolution before inserting it back into the original position. The AI-generated segments are highlighted with red bounding boxes. The generated segments are aligned with the removed segments in terms of duration, resolution, and frame rate, while the surrounding prefix and suffix frames are directly copied from the original videos. Since the manipulated intervals are explicitly specified during the construction process, the corresponding forgery boundaries can be automatically recorded, providing objective and auditable evidence.

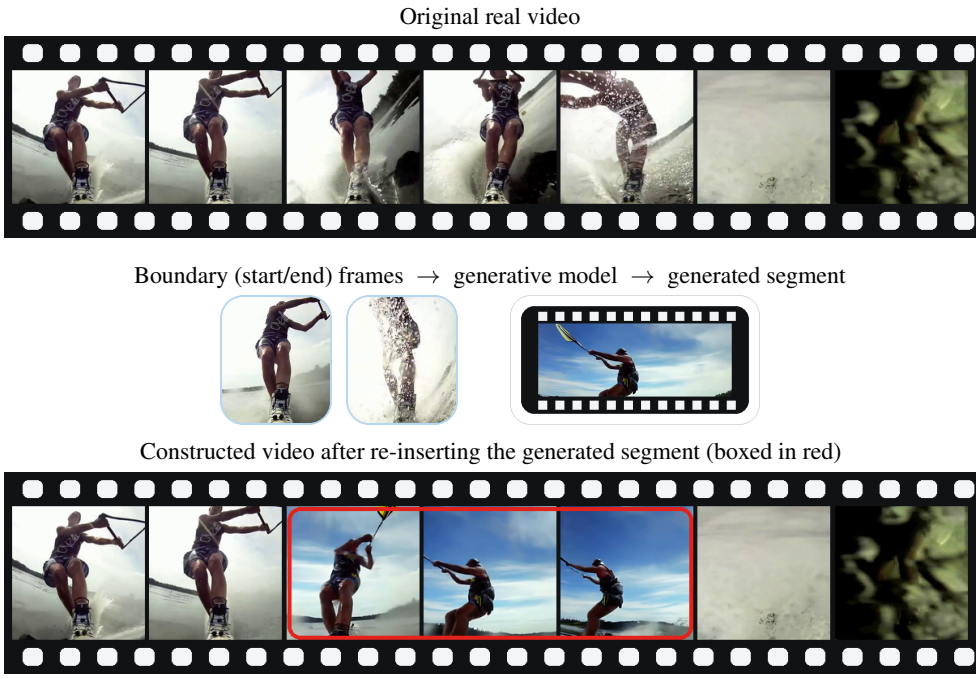

Figure 6: Data construction with **Wan2.2-Fun-In**. Source video: ActivityNet (*Waterskiing*); generation interval $[1.13, 3.40]$ s of a 5 s clip.

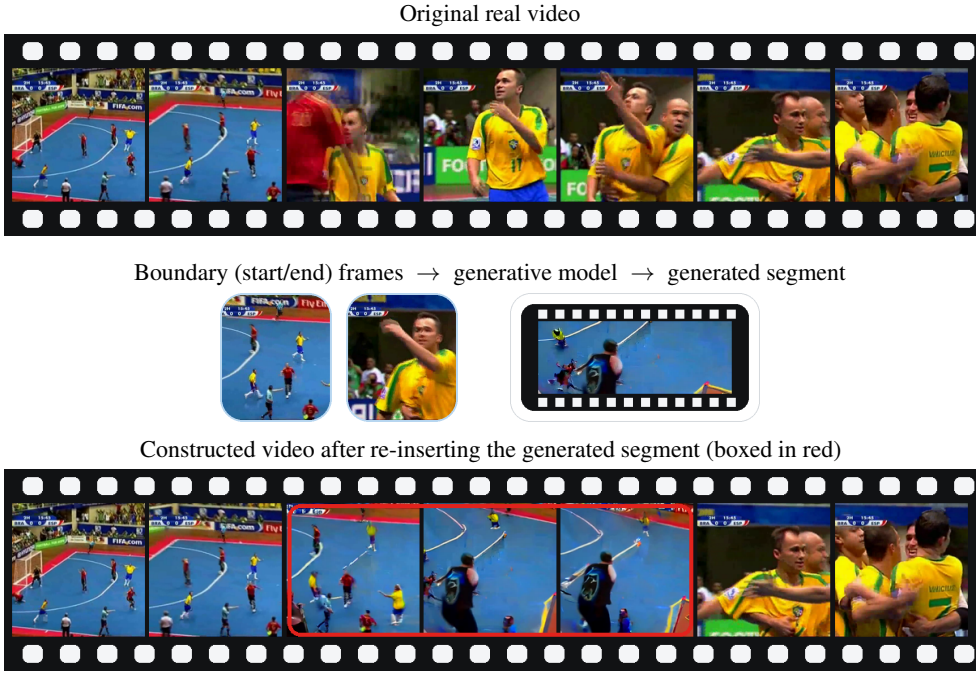

Figure 7: Data construction with **Wan2.2-Fun-In**. Source video: ActivityNet (*Futsal*); generation interval $[1.27, 3.53]$ s of a 5 s clip.

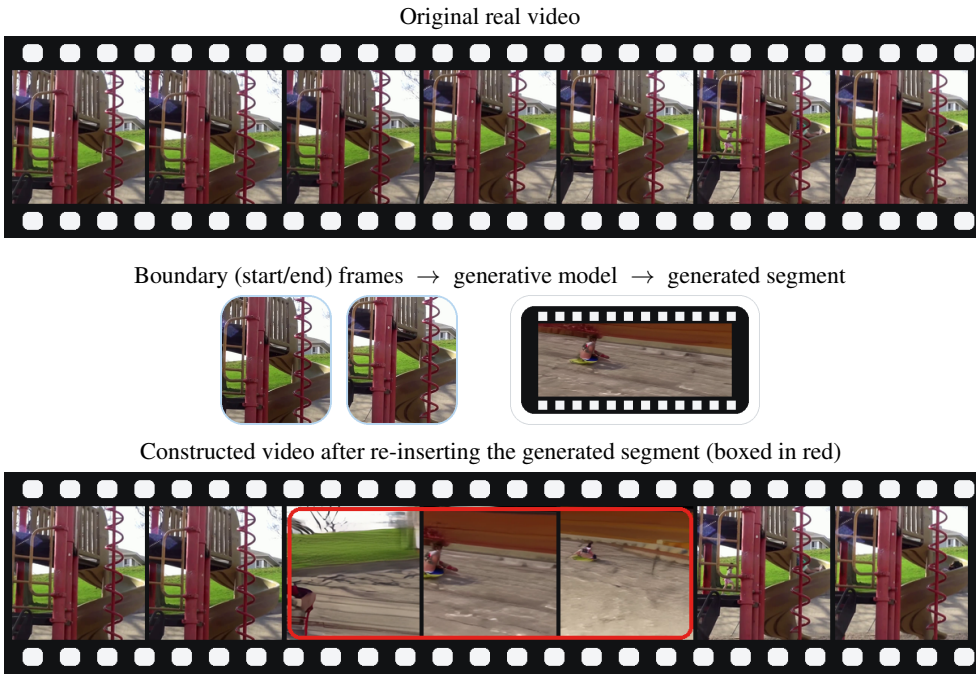

Figure 8: Data construction with **LTX-Video**. Source video: ActivityNet (*Fun sliding down*); generation interval $[0.73, 3.13]$ s of a 5 s clip.



Figure 9: Data construction with **LTX-Video**. Source video: ActivityNet (*Sumo*); generation interval $[1.47, 3.87]$ s of a 5 s clip.

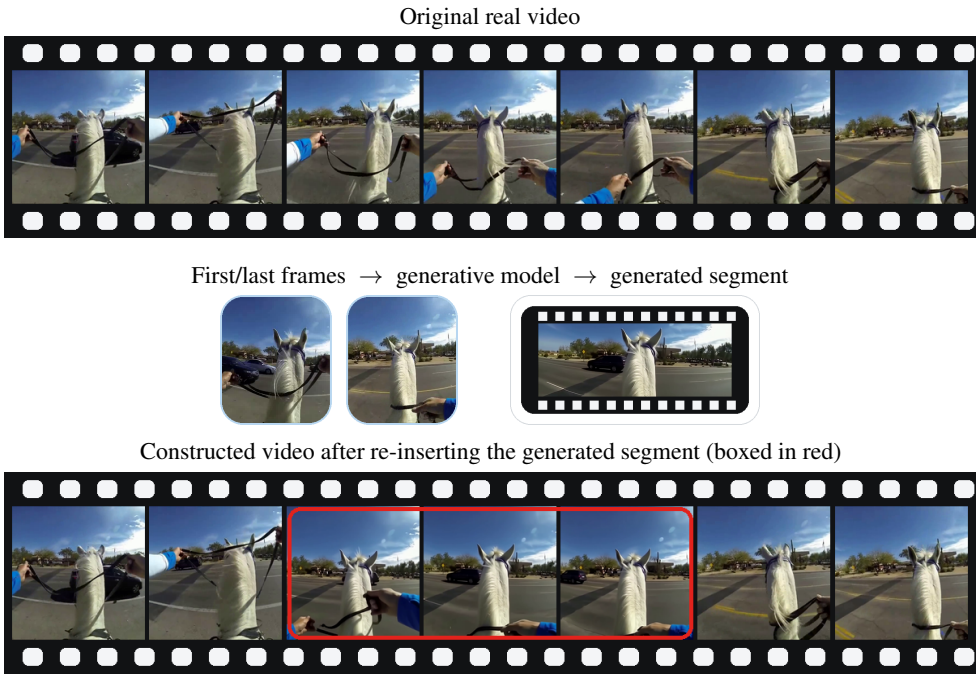

Figure 10: Data construction with **Wan2.7-i2v**. Source video: ActivityNet (*Horseback riding*); generation interval $[1.20, 3.40]$ s of a 5 s clip.

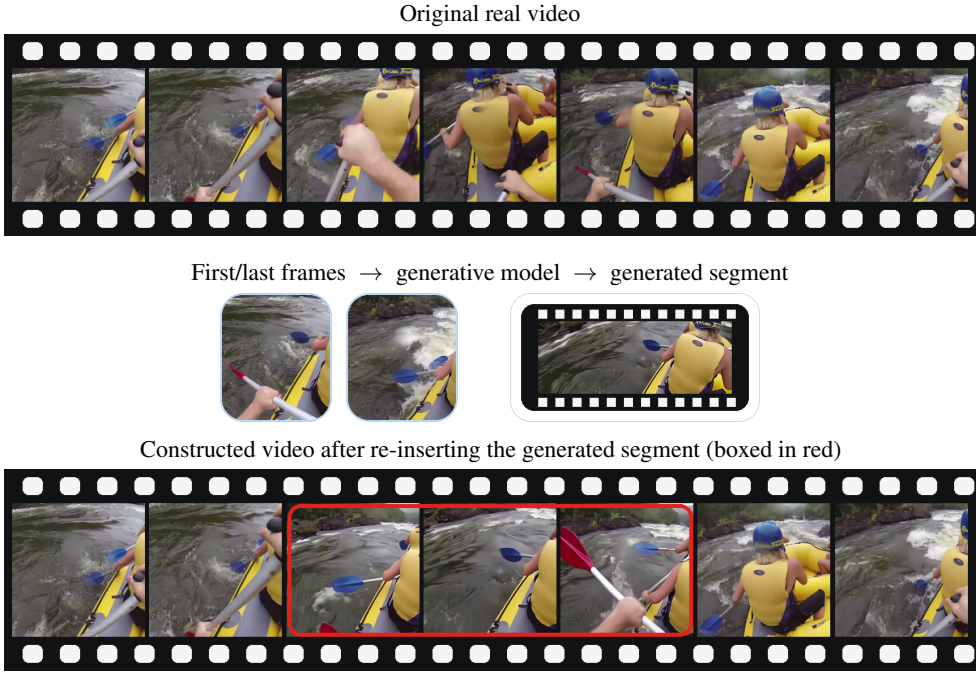

Figure 11: Data construction with **Wan2.7-i2v**. Source video: ActivityNet (*Rafting*); generation interval $[1.20, 3.40]$ s of a 5 s clip.



Figure 12: Data construction with **Seedance 1.0 pro**. Source video: ActivityNet (*Bullfighting*); generation interval $[1.20, 3.40]$ s of a 5 s clip.

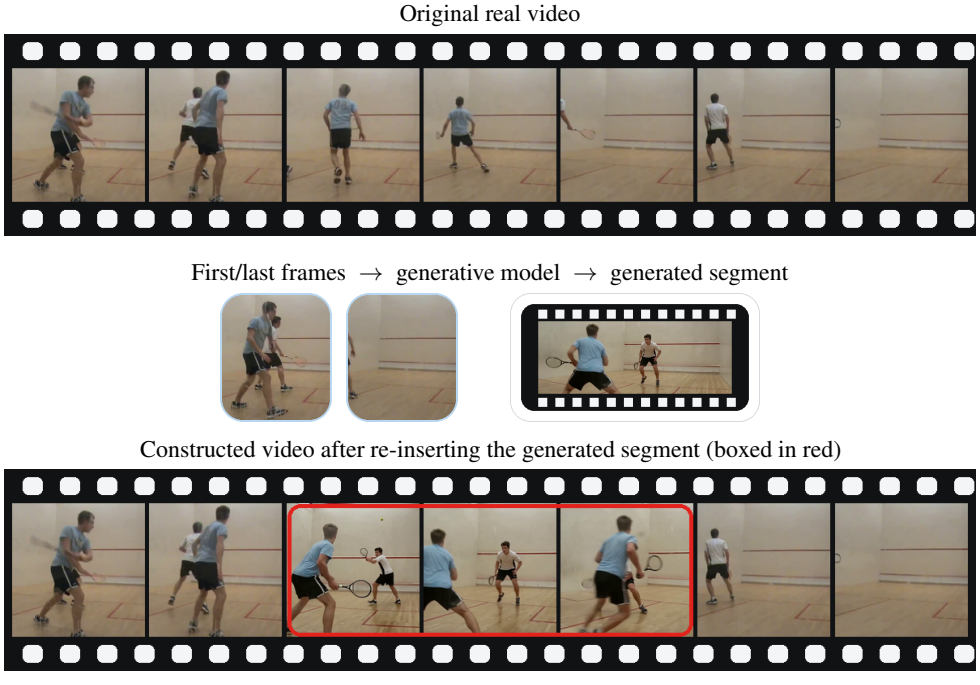

Figure 13: Data construction with **Seedance 1.0 pro**. Source video: ActivityNet (*Playing squash*); generation interval $[1.20, 3.40]$ s of a 5 s clip.

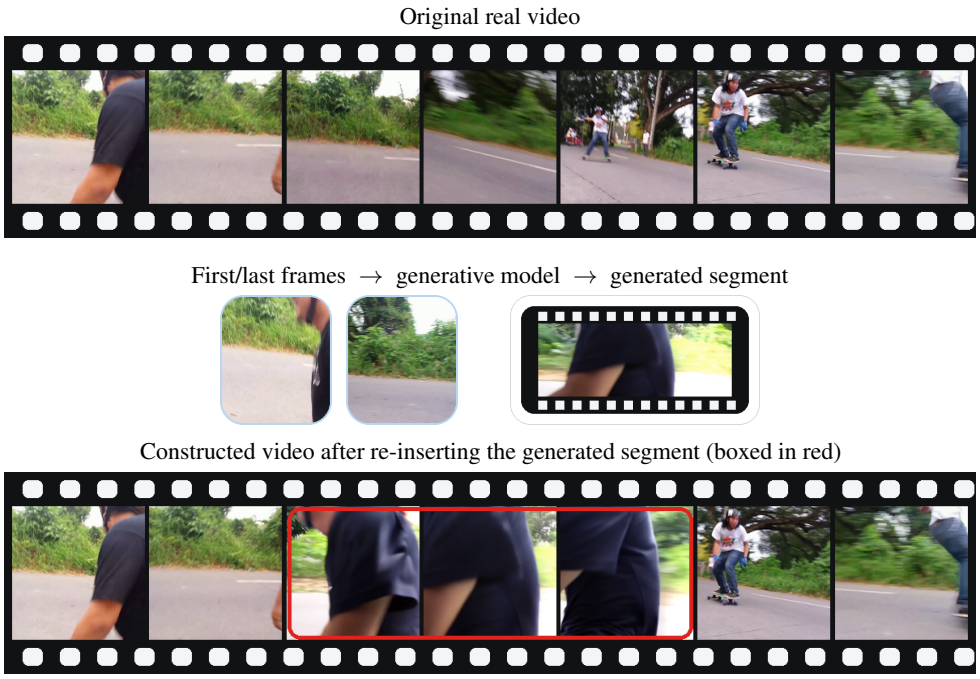

Figure 14: Data construction with **SkyReels-V2-DF-14B**. Source video: ActivityNet (*Longboarding*); generation interval $[1.20, 3.40]$ s of a 5 s clip.

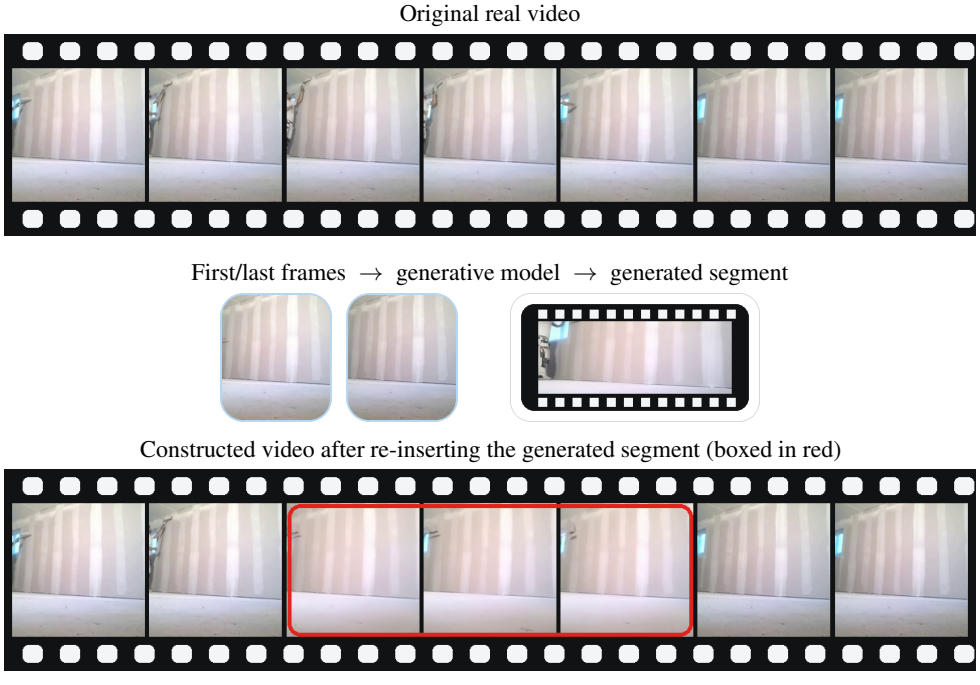

Figure 15: Data construction with **SkyReels-V2-DF-14B**. Source video: ActivityNet (*Plastering*); generation interval $[1.20, 3.40]$ s of a 5 s clip.