CVPR JUNE 3-7, 2026 DENVER COLORADO

中国科学技术大学 University of Science and Technology of China

SWIFT: Sliding Window Reconstruction for Few-Shot Training-Free Generated Video Attribution

Presenter: Chao Wang **Date: 2026-05-14**

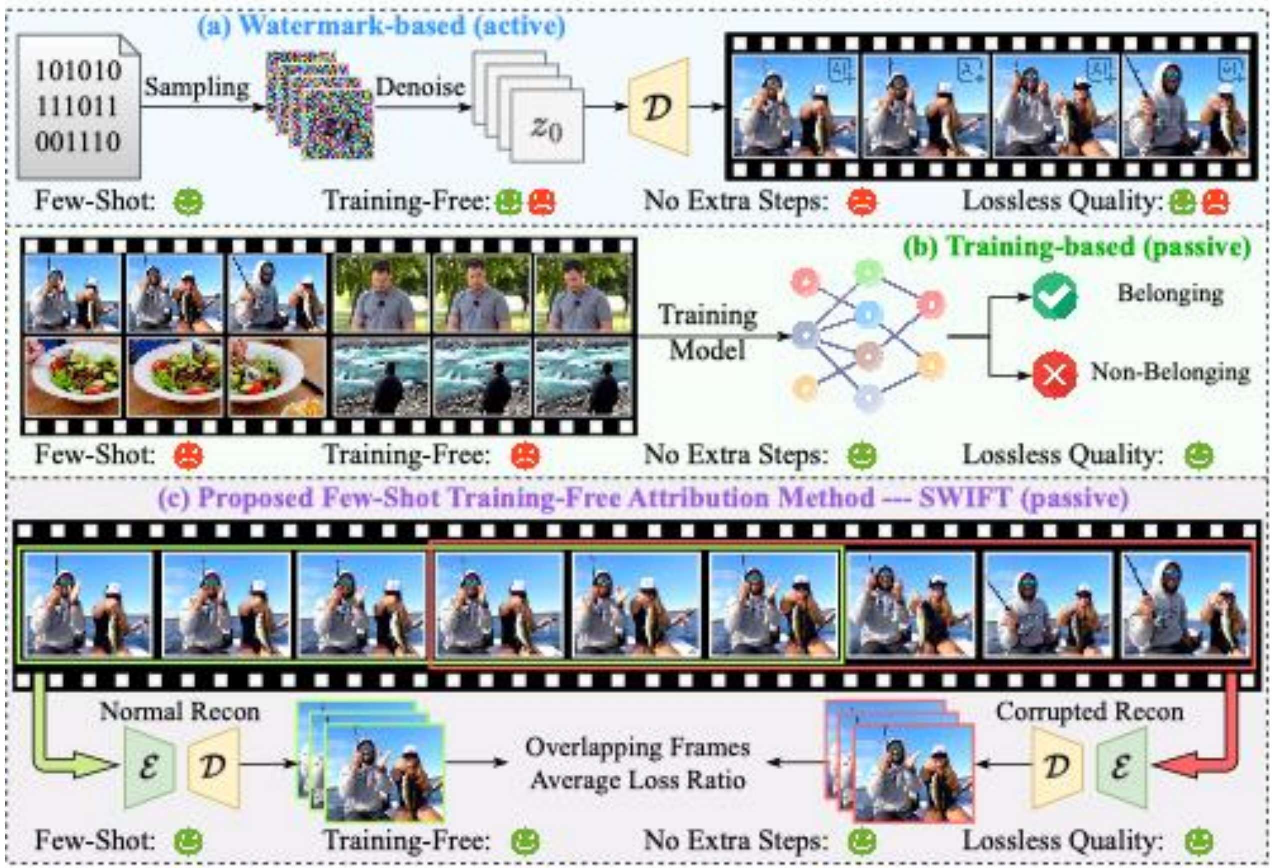
Code: https://github.com/wangchao0708/SWIFT

Email: {chaowang0708@mail, chenkj@}.ustc.edu.cn

Chao Wang[1], Zijin Yang[1], Yaofei Wang[2], Yuang Qi[1], Weiming Zhang[1], Nenghai Yu[1], Kejiang Chen[1,*]
[1]University of Science and Technology of China [2]Hefei University of Technology
chaowang0708@mail.ustc.edu.cn chenkj@ustc.edu.cn

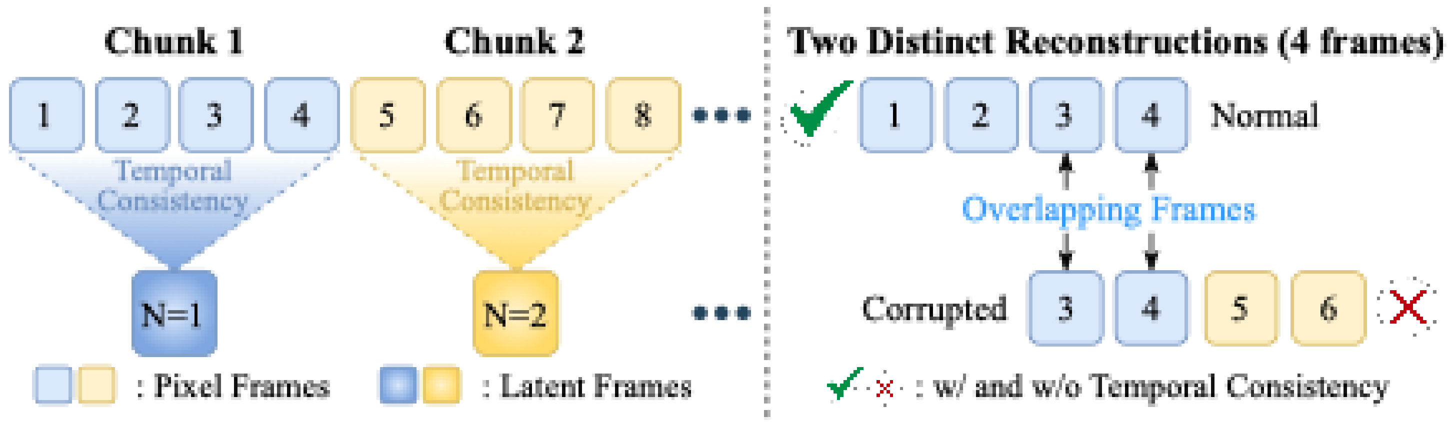
1 Background

Research Background



- ***Watermark-based (Active)***: Requires extra embedding operations; degrades visual quality and introduces deployment overhead.
- ***Training-based (Passive)***: Requires massive training samples and high cost; the model must be re-trained whenever a new generator emerges.
- ***Recon-based (Image modality)***: Accuracy drops on videos because it only captures spatial consistency and ignores **temporal coherence**.

Motivation



Mapping Relationship: Pixel Frames(many) ↔ Latent Frame(one)

Normal-Recon: preserves the relationship

Corrupted-Recon: disrupts it by shifting the window

Belonging: Corrupted Loss ↑↑ | Non-belonging: Loss ≈ unchanged

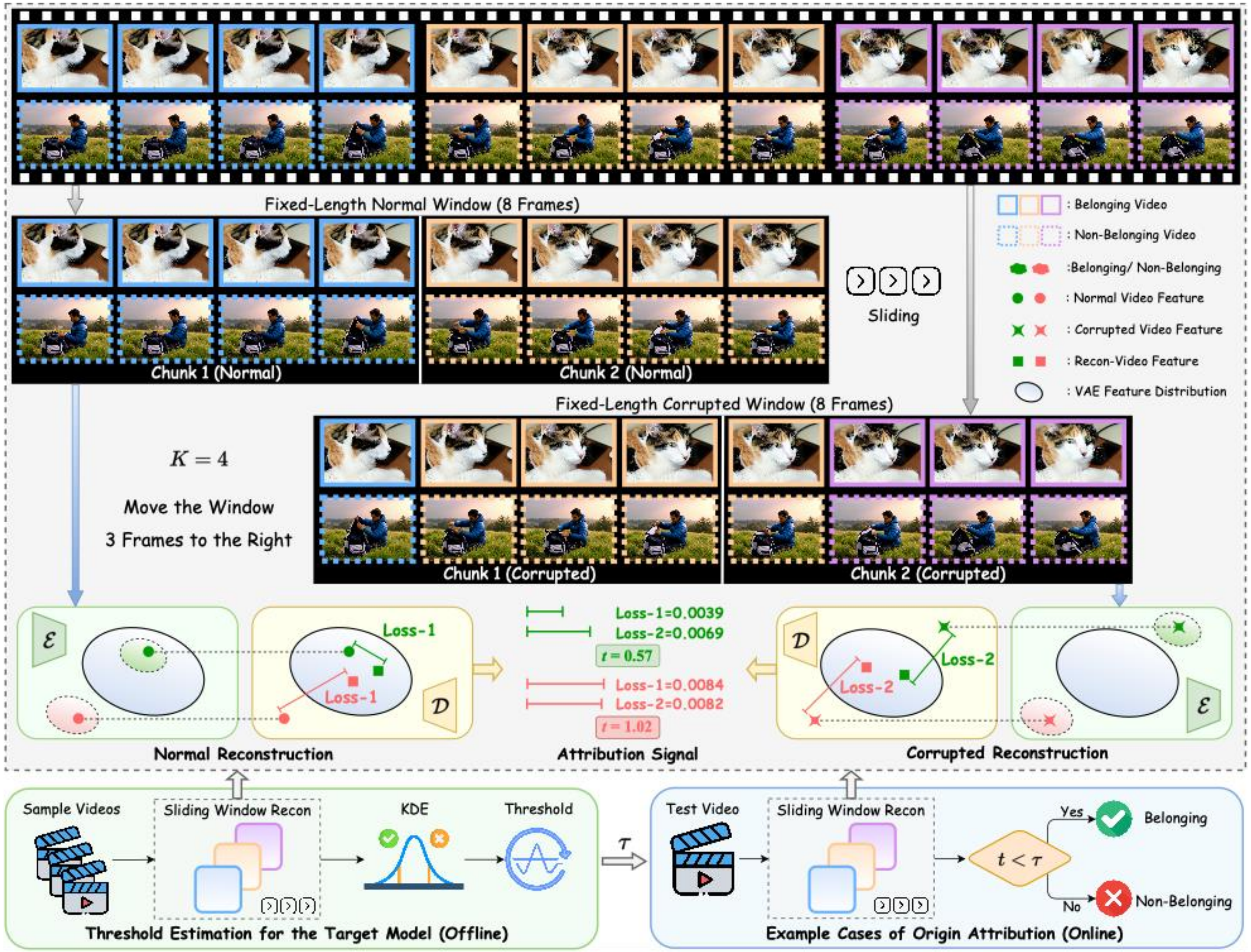
Belonging → Loss ratio ≪ 1; Non-belonging → Loss ratio ≈ 1

2 Methodology

Methodology

中国科学技术大学
University of Science and Technology of China

Three key modules: Fixed-Length Sliding Window, Dual Reconstruction, Adaptive Threshold



1. Build two windows of length K(N−1) from the test video — a **Normal one W_0** (aligned with chunks) and a **Corrupted one W_{K-1}** (shifted by K−1 frames to break chunk alignment).

$$W_{\mathrm{Nor}} : j \bmod K = 0, \quad W_{\mathrm{Cor}} : j \bmod K \neq 0.$$

2. Feed both windows through the target model's 3D VAE $\mathcal{R}$ to obtain a **Normal and a Corrupted reconstruction**, then compute per-frame MSE loss against the originals. Take the average per-frame loss ratio over the overlapping frames as signal t (**Belonging t ≪ 1; Non-Belonging t ≈ 1**).

$$t = \frac{1}{K(N-1) - K + 1} \sum_{i=K}^{K(N-1)} \frac{\mathcal{L}(F_i^{*}, F_i)}{\mathcal{L}(F_i^{**}, F_i)}.$$

3. Use **Kernel Density Estimation** on only S belonging samples (default **$S = 200$, $\alpha = 0.05$**) to non-parametrically derive threshold τ; classify as belonging if t < τ, otherwise non-belonging.

3 Experimental Results

Experimental Setup and Baseline

中国科学技术大学 University of Science and Technology of China

Real Videos: Randomly select 500 videos from **OpenVid-1M**

Minimize the impact of image diversity

Generated Videos: **5 Models** (Each model generates 700 videos using prompts from **OpenVid-1M**)

Baseline: **AEDR** (AAAI 2026 Oral, ***SOTA Image Attribution Method***)

Table 1: Detailed composition information of the dataset (S-Video).

Dataset / Model	Specific Type	Release Date	Resolution (W×H)	Frames	FPS	Quantity
OpenVid-1M [18]	Real	2024-07-01	512×360∼1024×1024	30∼196	20∼60	500
HunyuanVideo [13]	T2V	2024-12-03	960×544	129	24	700
Wan2.1 [26]	T2V-1.3B	2025-02-25	832×480	81	16	700
EasyAnimate [34]	T2V-7B-V5.1	2025-03-06	1008×576	49	8	700
LTX-Video [8]	TI2V-13B-0.9.7	2025-05-05	1216×704	121	30	700
Wan2.2 [26]	TI2V-5B	2025-07-28	1280×704	121	24	700

Figure 2: Sample video in S-Video.

Effectiveness and Efficiency

Table 2: Attribution Effectiveness

$\mathcal{M}_1$	$\mathcal{M}_2$	AEDR [27]					SWIFT (ours)				
		TP	FP	FN	TN	Acc	TP	FP	FN	TN	Acc
HYV	Wan2.1	481	465	19	35	51.6%	473	6	27	494	96.7% (+45.1)
	EA	481	375	19	125	60.6%	473	7	27	493	96.6% (+36.0)
	LTX	481	467	19	33	51.4%	473	171	27	329	80.2% (+28.8)
	Wan2.2	481	476	19	24	50.5%	473	143	27	357	83.0% (+32.5)
	Real	481	433	19	47	54.8%	473	4	27	496	96.9% (+42.1)
Wan2.1	HYV	477	80	23	420	89.7%	484	0	16	500	98.4% (+8.7)
	EA	477	45	23	455	93.2%	484	0	16	500	98.4% (+5.2)
	LTX	477	41	23	459	93.6%	484	0	16	500	98.4% (+4.8)
	Wan2.2	477	155	23	345	82.2%	484	0	16	500	98.4% (+16.2)
	Real	477	102	23	398	87.5%	484	0	16	500	98.4% (+10.9)
EA	HYV	478	323	22	171	65.5%	478	0	22	500	97.8% (+32.3)
	Wan2.1	478	263	22	237	71.5%	478	0	22	500	97.8% (+26.3)
	LTX	478	478	22	22	50.0%	478	0	22	500	97.8% (+47.8)
	Wan2.2	478	374	22	126	60.4%	478	0	22	500	97.8% (+37.4)
	Real	478	296	22	204	68.2%	478	0	22	500	97.8% (+29.6)
LTX	HYV	482	318	18	182	66.4%	474	195	26	305	77.9% (+11.5)
	Wan2.1	482	153	18	347	82.9%	474	124	26	376	85.0% (+2.1)
	EA	482	11	18	489	97.1%	474	29	26	471	94.5% (-2.6)
	Wan2.2	482	224	18	276	75.9%	474	204	26	296	77.0% (+1.1)
	Real	482	26	18	474	95.6%	474	52	26	448	92.2% (-3.4)
Wan2.2	HYV	474	179	26	321	79.5%	483	14	17	486	96.9% (+17.4)
	Wan2.1	474	371	26	129	60.3%	483	0	17	500	98.3% (+38.0)
	EA	474	205	26	295	76.9%	483	0	17	500	98.3% (+21.4)
	LTX	474	147	26	353	82.7%	483	3	17	497	98.0% (+15.3)
	Real	474	44	26	456	93.0%	483	3	17	497	98.0% (+5.0)
Avg Acc						**73.6%**					**94.0%** (+20.4)

Table 3: Few-Shot Effectiveness

S	Attribution Accuracy				
	HYV	Wan2.1	EA	LTX	Wan2.2
0	**90.1%**	81.7%	**89.3%**	66.1%	**98.4%**
5	78.6%	88.7%	80.3%	79.0%	91.8%
10	76.1%	88.3%	80.8%	78.3%	90.4%
20	85.2%	94.2%	95.0%	83.9%	92.7%
35	90.3%	96.9%	97.4%	84.8%	93.2%
50	90.7%	97.1%	96.6%	84.7%	93.3%
100	90.2%	98.4%	97.7%	85.0%	97.4%
150	90.2%	98.4%	97.8%	85.1%	97.6%
200	90.7%	98.4%	97.8%	85.3%	97.9%

Table 4: Impact of Video Length

Length	Frames	Runtime	Accuracy
100%	121	170s	**98.0%**
75%	93	134s	98.0%
50%	61	84s	97.1%
25%	33	**40s**	94.1%

T2. SWIFT achieves **94.0%** average accuracy, outperforming AEDR **(73.6%)** by **+20.4%**, with consistent gains and remarkably ***low false positive rates***.

T3. Only **20 samples** suffice for **90%** accuracy; SWIFT even achieves **~90% Zero-Shot** attribution on ***HYV, Wan2.2, and EA***.

T4. As video length shrinks from **100%** to **25%**, accuracy drops only slightly **(98.0%→94.1%)** while runtime falls drastically ***(170s→40s)***.

SWIFT is currently the best-performing training-free passive attribution method for generated videos

CVPR JUNE 3-7, 2026 DENVER COLORADO

中国科学技术大学
University of Science and Technology of China

汪超
安徽 合肥

扫二维码，添加我为朋友。

Thank you!

Presenter: Chao Wang **Date: 2026-05-14**

Code: https://github.com/wangchao0708/SWIFT

Email: {chaowang0708@mail, chenkj@}.ustc.edu.cn