AI-Generated video? Detected!

TECHNION Israel Institute of Technology

CVPR JUNE 3-7, 2026

Project page

Training-free Detection of Generated Videos via Spatial-Temporal Likelihoods

Omer Ben Hayun   Roy Betser   Meir Yossef Levi   Levi Kassel   Guy Gilboa

1. Introduction

- Training-free
- No generated samples required
- Spatial + Temporal evidence
- Fast inference

STALL

AI video generators are becoming increasingly realistic:

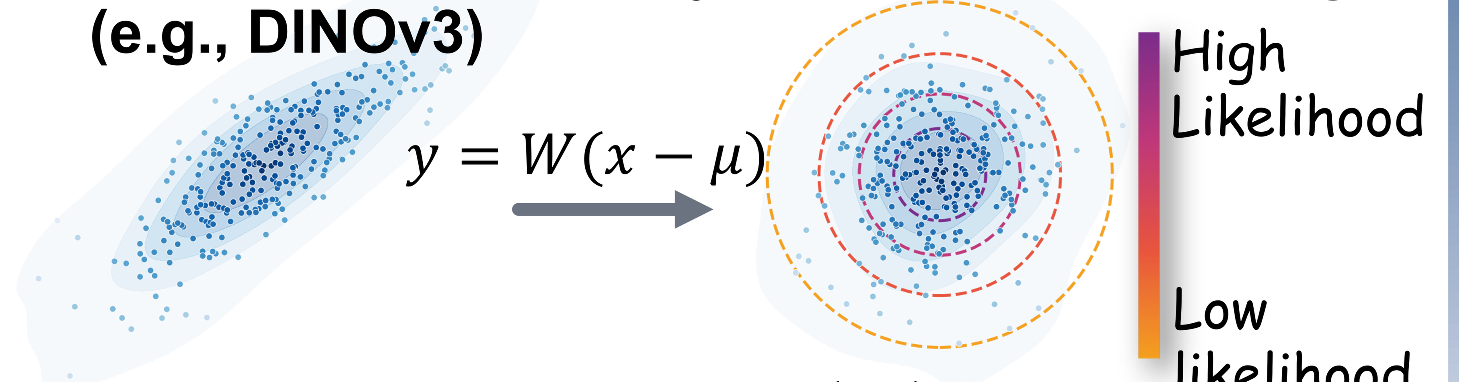
Reasearch gap: All prior detectors miss at least one key property:

Method	Train-free	Temporal cues	Spatial cues
Supervised	✗	✓	✓
Zero-shot image	✓	✗	✓
Zero-shot video	✓	✓	✗
STALL (ours)	✓	✓	✓

2. Preliminaries: Whitening & Likelihood

x Pretrain embeddings (e.g., DINOv3)

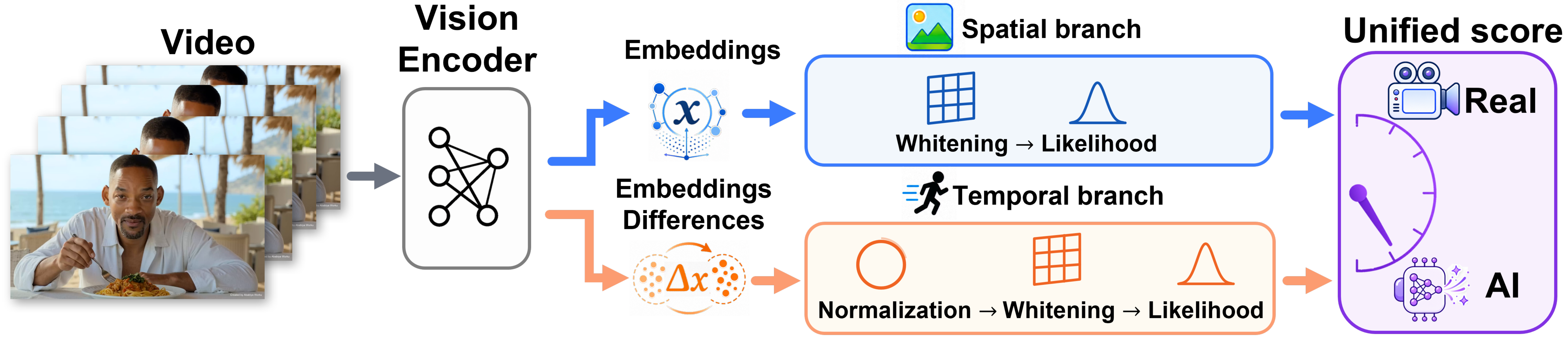
y Whitened embeddings



- $x \sim \mathcal{N}$ coordinate-wise
- Shifted and correlated

- $y \sim \mathcal{N}(0, I)$
- $\ell(y) = -0.5[d\log(2\pi) + \|y\|^2]$

3. Method: STALL (Spatial-Temporal Aggregated Log-Likelihoods)

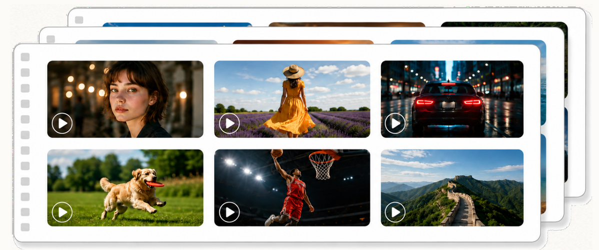


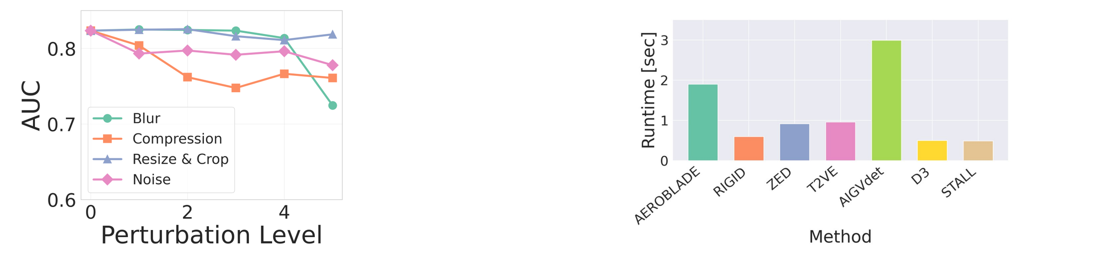
4. Experiments

Benchmark	RIGID	ZED	D3	STALL (ours)
VideoFeedback	0.63	0.54	0.54	**0.83**
GenVideo	0.65	0.55	0.72	**0.80**
ComGenVid (Ours)	0.57	0.55	0.73	**0.85**
Overall Avg	0.61	0.57	0.64	**0.82**

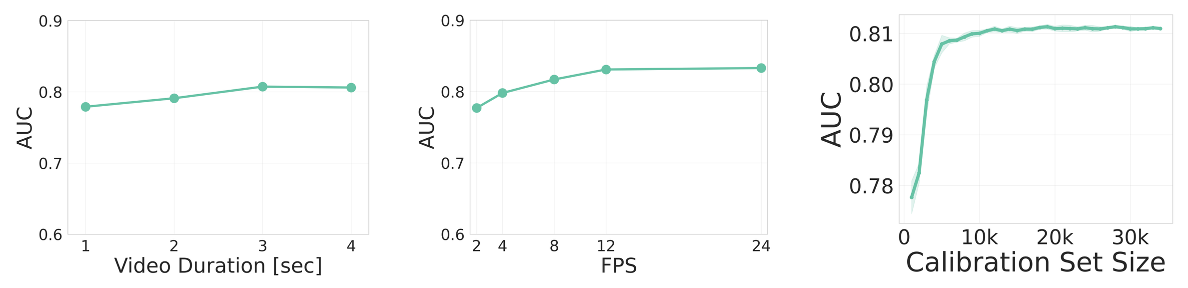
Zero-shot methods AUC across three benchmarks (higher = better).

New benchmark: ComGenVid
3,400 videos
Sora + Veo 3.

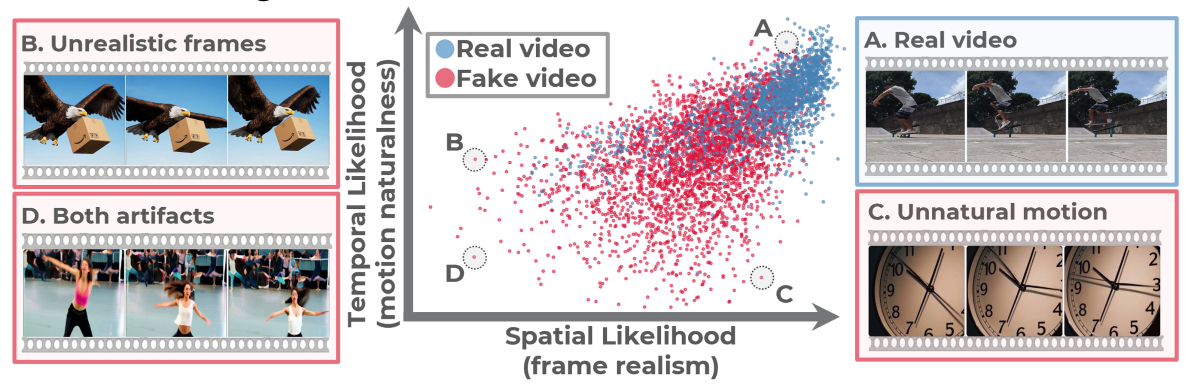




Robustness to perturbations   Latency comparison



Video length   FPS ablation   Calibration set size



Spatial-temporal scores separate real from AI-generated videos.