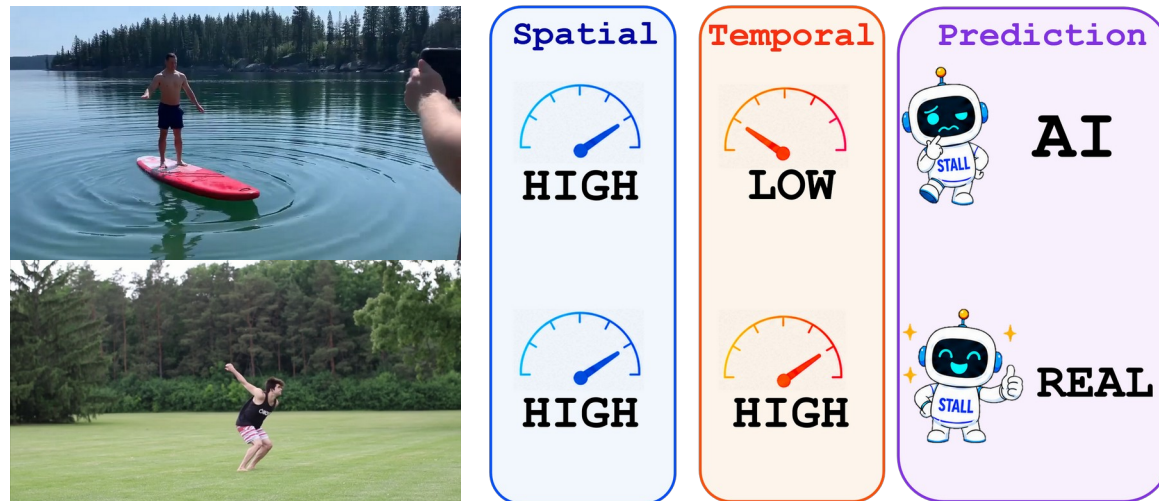


Training-free Detection of Generated Videos via Spatial-Temporal Likelihoods

CVPR 2026



Omer Ben Hayun, Roy Betser, Meir Yossef Levi, Levi Kassel, Guy Gilboa
Technion – Israel Institute of Technology

Which is generated by AI?



AI



REAL

Motivation

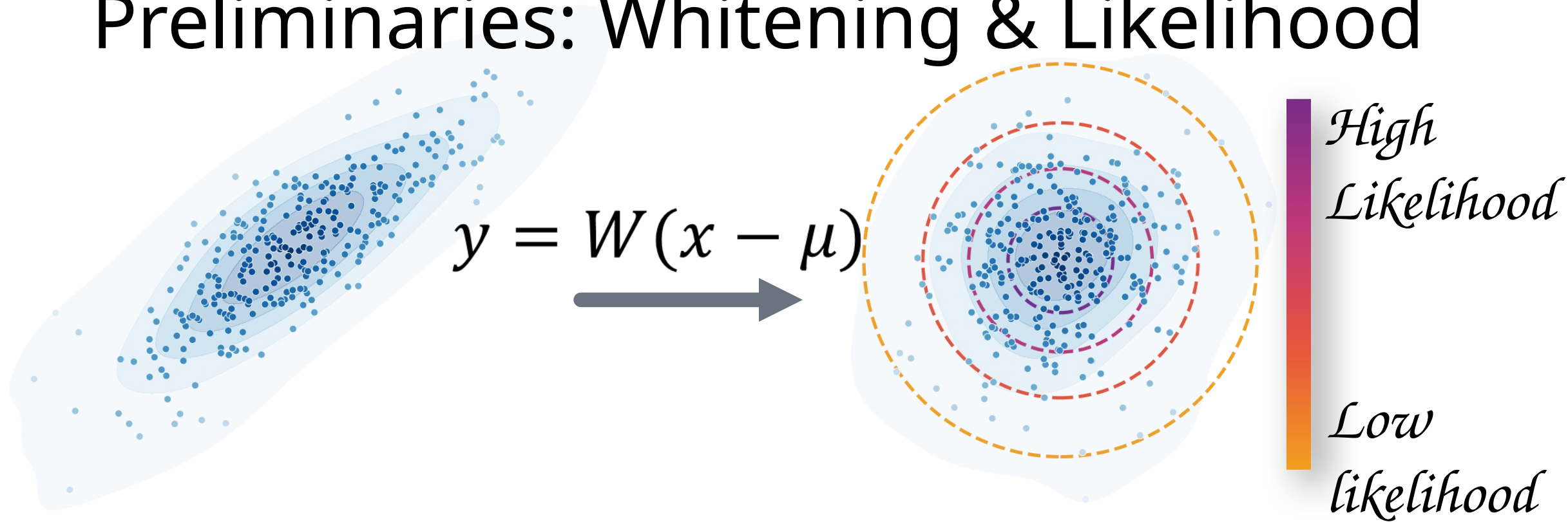
- Consent & likeness
- Misinformation risk
- Intellectual property



Research gap

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

Preliminaries: Whitening & Likelihood



\mathcal{X} Pretrain embeddings

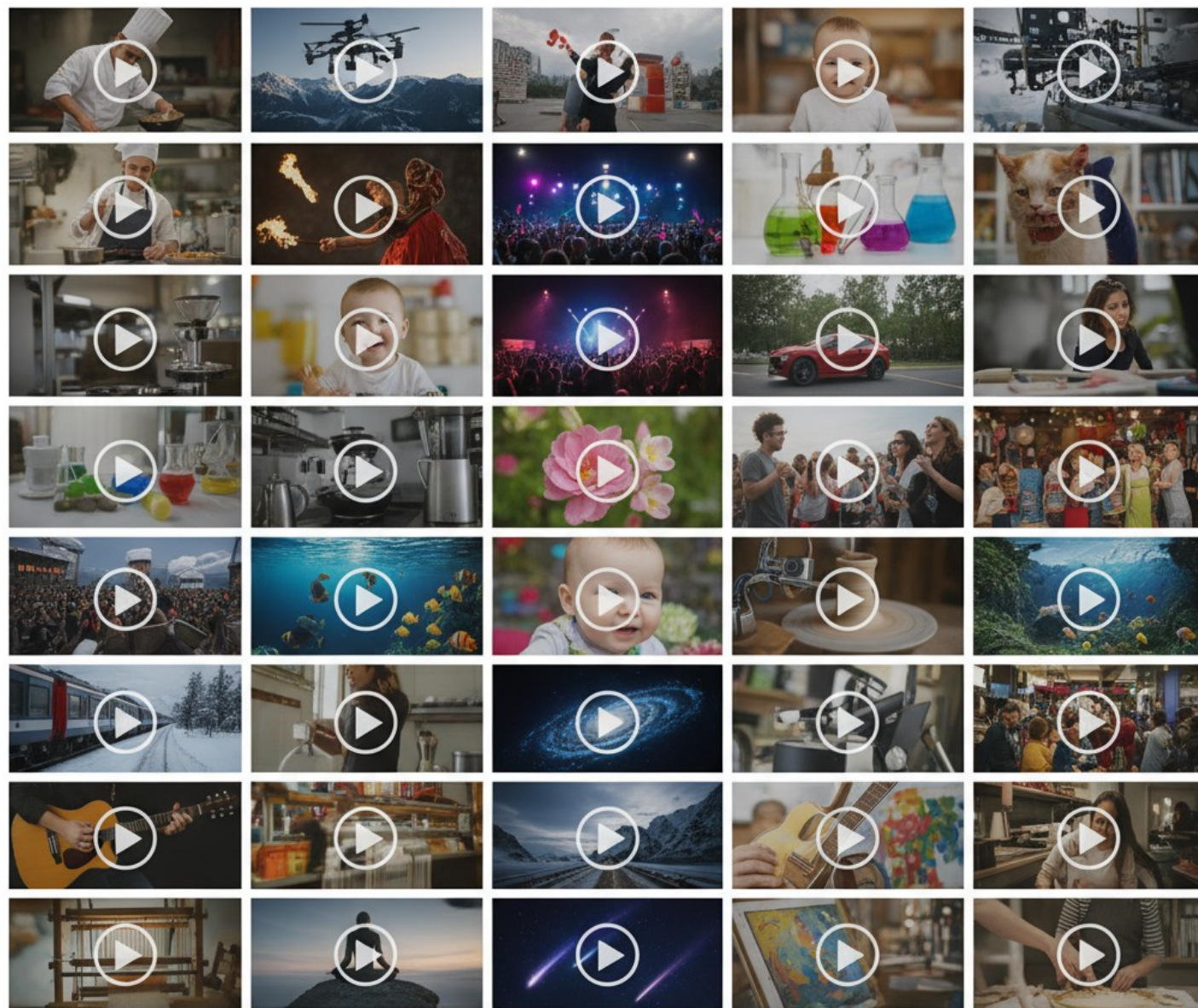
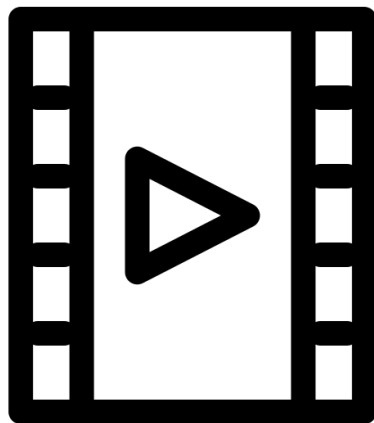
- $x \sim \mathcal{N}$ coordinate-wise
- Shifted and correlated

\mathcal{Y} Whitened embeddings

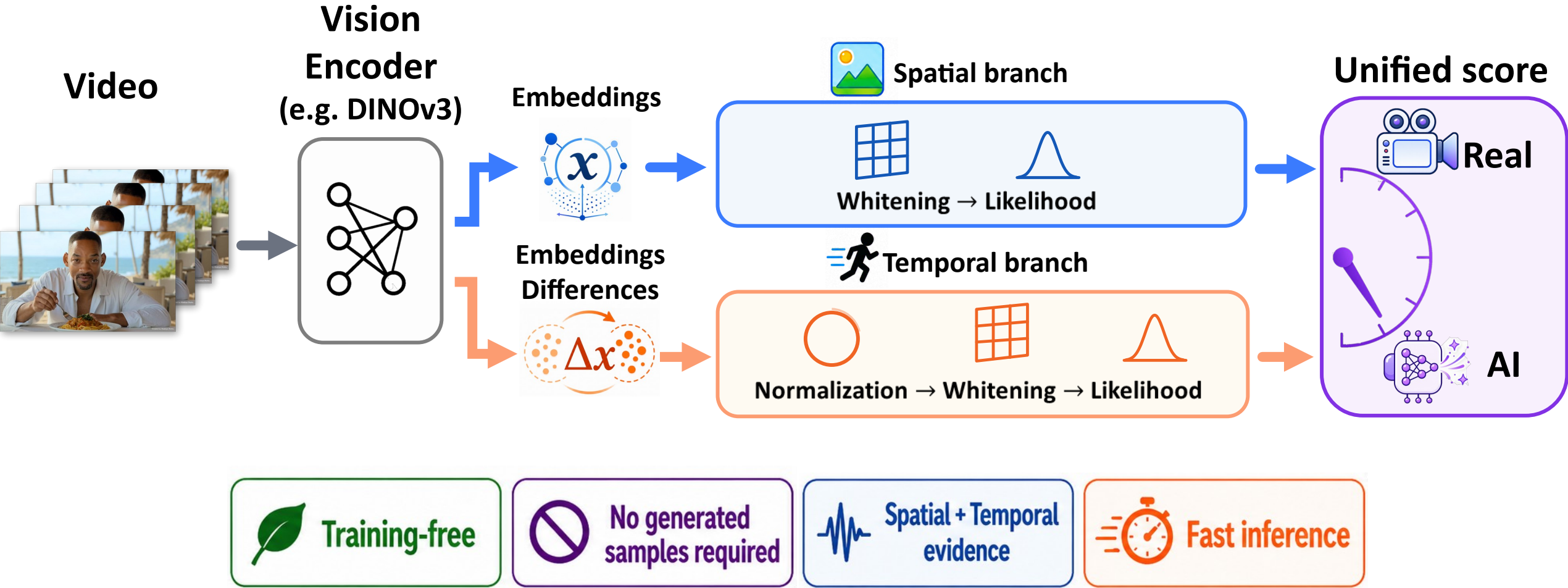
- $y \sim \mathcal{N}(0, I)$
- $\ell(y) = -0.5[d \log(2\pi) + \|y\|^2]$

Calibration set

- Dataset of real videos
- Disjoint from test set
- Large & varied
- Whiteing transform μ, W



Method

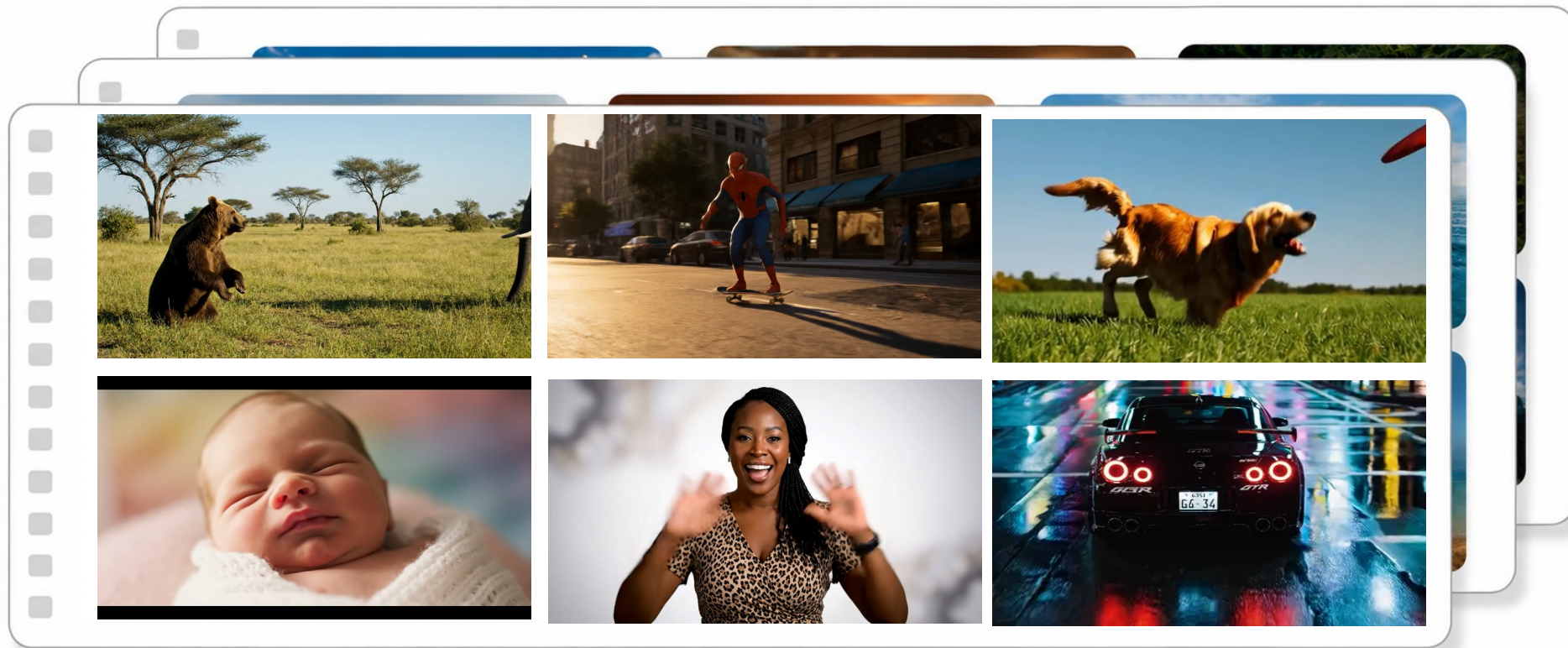


Frames from youtube video: "Will Smith Eating a spaghetti in 2025 | 4K" by a3xrfgb

New Benchmark: ComGenVid



Hugging Face



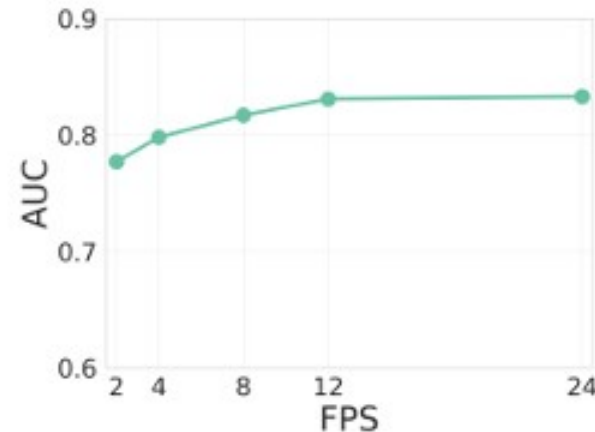
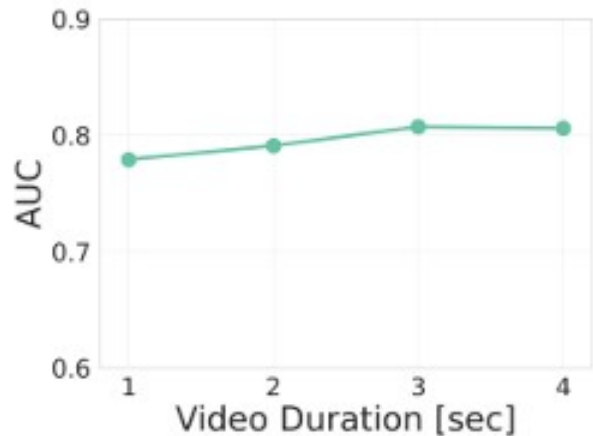
*~3,400 generated videos from **Sora** (OpenAI) and **Veo 3** (Google)*

Results - Average AUC (↑)

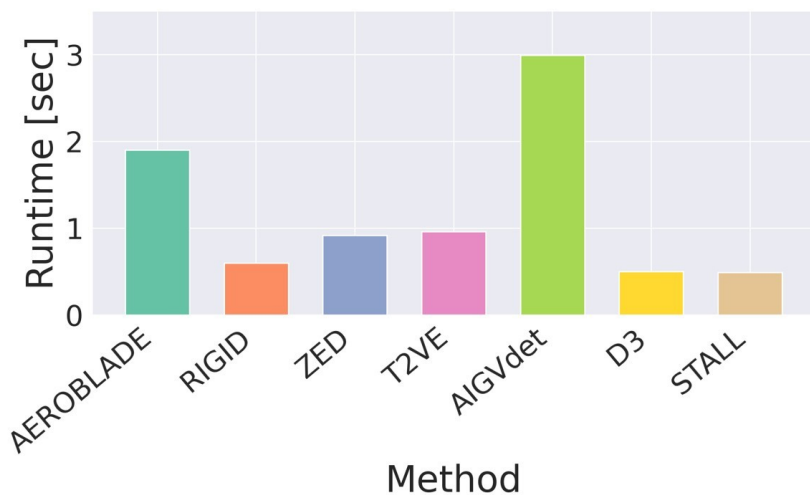
Method	Domain	Benchmark		
		VideoFeedback	GenVideo	ComGenVid (ours)
AEROBLADE	Images	0.58	0.59	0.69
RIGID	Images	<u>0.63</u>	0.65	0.57
ZED	Images	0.54	0.55	0.55
D3 (L2)	Videos	0.54	<u>0.72</u>	<u>0.73</u>
D3 (cos)	Videos	0.55	0.70	<u>0.73</u>
STALL (Ours)	Videos	0.83	0.80	0.85

Ablation studies

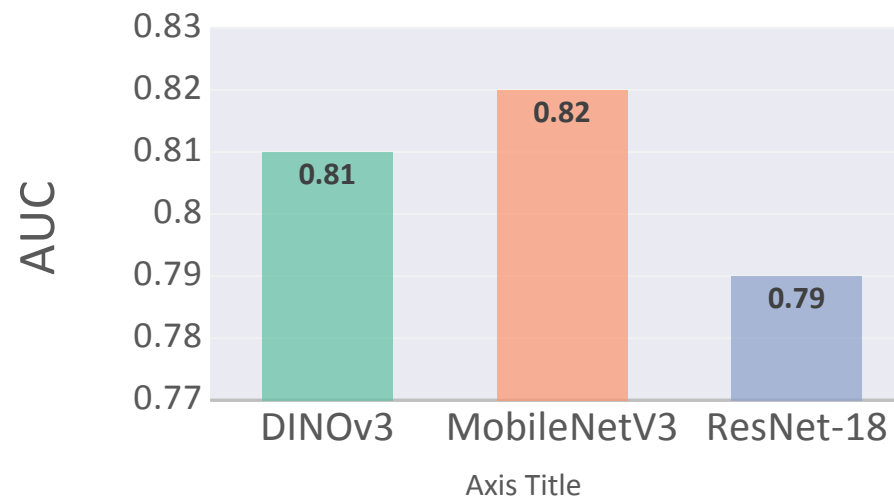
Temporal ablations



Inference runtime



Backbone encoders



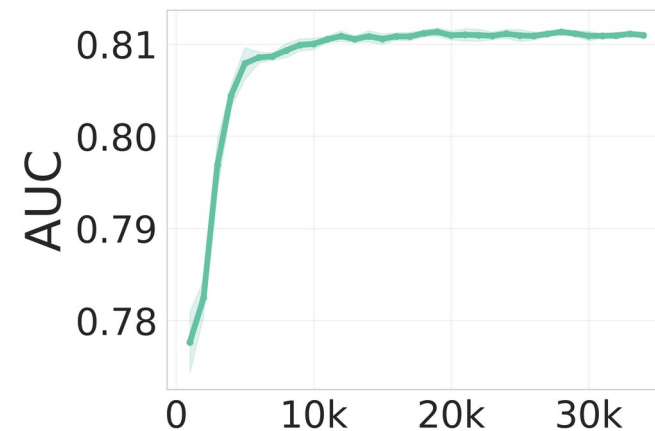
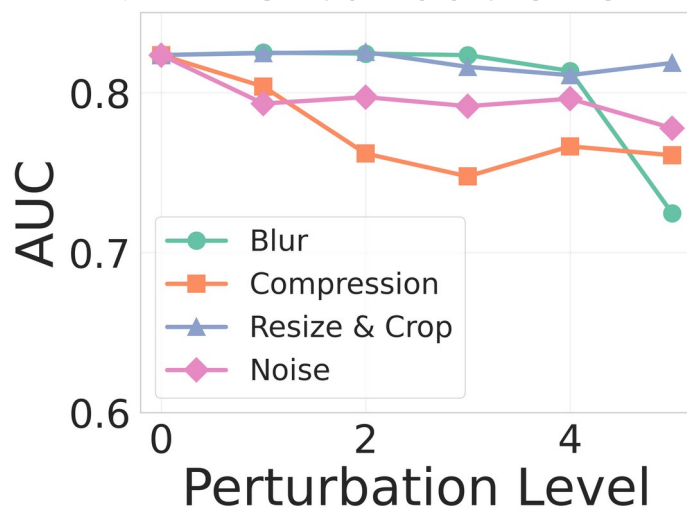
Ablation studies

Calibration set

Benchmark	Vatex	Kinetics400	DiDeMo	Panda-70M	MSR-VTT
VideoFeedback	0.82	0.82	0.86	0.76	0.73
GenVideo	0.78	0.77	0.76	0.83	0.83
ComGenVid	0.82	0.81	0.87	0.75	0.76

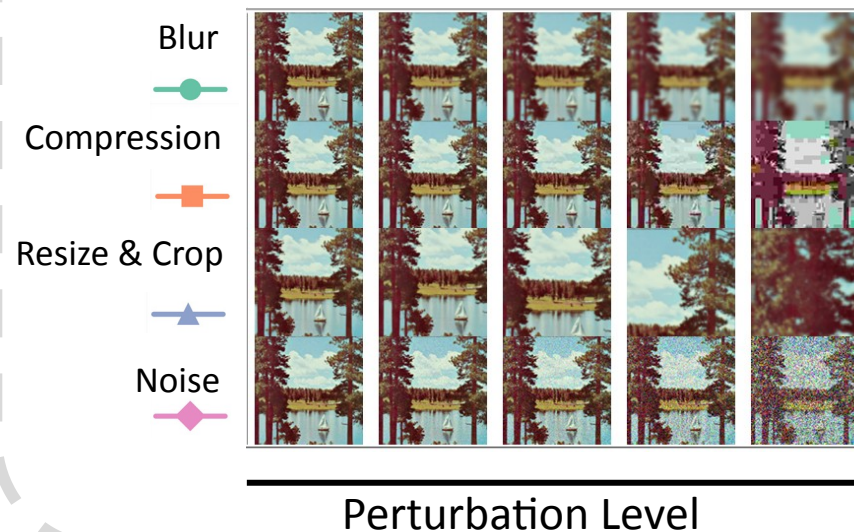
Calibration set variations AUC (↑)

Perturbations



Calibration set size

Perturbations examples



Conclusions



Training-free



**No generated
samples required**



**Spatial + Temporal
evidence**



Fast inference



Project page