

Article

Not peer-reviewed version

Temporal Stability Profiling of AI-Generated Short Videos: A Training-Free Metric Suite with Controlled-Perturbation Validation and a Case Study on Public Showcase Clips

[Grace Watt](#) *

Posted Date: 12 October 2026

Keywords: AI-generated video; temporal consistency; flicker; frame freezing; shot boundary detection; no-reference video quality; reproducibility; near-duplicate detection



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Temporal Stability Profiling of AI-Generated Short Videos: A Training-Free Metric Suite with Controlled-Perturbation Validation and a Case Study on Public Showcase Clips

Grace Watt

Independent Researcher; addem4670@gmail.com

Abstract

AI-generated video often fails in time rather than in any single frame. Typical failures are global brightness flicker, frozen segments, padded duplicate frames, high-frequency texture “boil”, and abrupt discontinuities. Learned quality models can score such clips, but they need GPU inference, pinned checkpoints, and feature extractors that change between releases, which makes results hard to reproduce outside the original lab. This paper proposes TSM, a training-free Temporal Stability Metric suite that needs only FFmpeg and array arithmetic. TSM computes six interpretable measures on downsampled luma: activity, freeze ratio, flicker index, temporal jitter ratio, a flash-robust shot-boundary count, and a native-rate duplicate-frame ratio. It also defines a frame-difference time series that serves as a temporal signature for detecting re-encoded or re-framed variants of the same clip. We validate TSM in three ways. First, controlled perturbations of 15 clean clips show that each target metric responded in the expected direction in 15 of 15 clips for freezes, flicker, frame duplication, and temporal noise, and in 14 of 15 for an inserted hard cut. Second, on 57 candidate shot boundaries annotated by visual inspection, the TSM detector reached precision 0.756 and recall 0.944, against 0.875 and 0.972 for FFmpeg’s scdet; requiring both detectors gave 0.917 for both. Third, re-encoding at CRF 35 changed every continuous metric by a median of at most 3.4%. We then apply TSM to 35 public showcase files from three AI video websites (Section 6). Checksums show that 10 files are byte-identical across two sites, which leaves 25 unique clips, and the temporal signature identifies one portrait/widescreen variant pair. The case study shows how TSM separates multi-shot edits from single-shot clips before any quality comparison. It does not rank the sites or the generators behind them.

Keywords: AI-generated video; temporal consistency; flicker; frame freezing; shot boundary detection; no-reference video quality; reproducibility; near-duplicate detection

1. Introduction

Modern text-to-video and image-to-video generators produce frames that are individually convincing [1–5]. Their characteristic failures are temporal. Identity drifts across frames, textures crawl, brightness pulses, motion stalls, and some clips are padded with repeated frames to reach a target duration. Benchmarks such as VBench include dimensions for temporal flicker and motion smoothness [6], and EvalCrafter combines several motion and consistency measures [7]. These suites, however, depend on learned feature extractors, optical-flow networks, and specific model checkpoints. They are well suited to model developers. They are less suited to a practitioner who wants a fast, auditable answer to simpler questions: does this file contain frozen segments, flicker, duplicated frames, or several shots edited together?

Those simpler questions matter because they determine whether a clip can be compared at all. A 30-second montage with 29 cuts and a single 5-second shot should not be scored on the same

temporal-consistency scale: cuts look like extreme inconsistency to any frame-pair measure. Likewise, a clip padded by frame duplication will look “smooth” to measures that reward small inter-frame change. A lightweight pre-screen that detects these structural properties is a prerequisite for fair evaluation.

This paper contributes:

1. TSM, a six-metric temporal stability suite that is training-free, deterministic, and fully specified by a few fixed parameters;
2. a flash-robust shot-boundary rule that combines mean-removed frame differences with a spatial-spread test;
3. a frame-difference temporal signature for detecting re-encoded and re-framed variants that byte-level checksums miss;
4. a three-part validation, consisting of controlled perturbations with known ground truth, annotated shot boundaries compared with an established detector, and compression robustness; and
5. a case study on 35 public showcase files from three AI video websites, which demonstrates corpus auditing, deduplication, and structural profiling, with per-clip metric values and a reporting checklist.

2. Scope and Claims

2.1. In Scope

TSM targets short clips of up to about 30 s, decoded from common containers. It measures signal-level temporal properties: change magnitude, stillness, brightness oscillation, temporal roughness, shot boundaries, and duplicated frames.

2.2. Out of Scope

TSM does not measure prompt fidelity, semantic correctness, physical plausibility, identity preservation, or aesthetics. A low jitter ratio does not mean the motion is correct, and a high activity value does not mean the motion is good. TSM is not a deepfake detector, and it does not infer which model produced a clip.

2.3. Permitted Claims

A report using TSM may state that a clip or a corpus has measured temporal properties under the stated parameters. It may not rank generators or websites from TSM values alone, because those values depend on editing, encoding, content, and curation as much as on generation.

3. Related Work

Distributional video metrics such as FVD [8] compare feature statistics of sets of clips and cannot locate a failure inside one clip. VBench [6] measures temporal flicker as mean absolute difference between frames of static-scene clips and motion smoothness with a frame-interpolation prior. EvalCrafter [7] uses optical-flow-based measures and aligns metric combinations to human opinion. Warping-error measures built on RAFT flow [9] and perceptual distances such as LPIPS [10] estimate temporal consistency more precisely, at higher computational cost. No-reference video quality models such as DOVER [11] predict overall technical and aesthetic quality.

Shot-boundary detection has a long history. Histogram and pixel-difference methods were surveyed early on [12]. TRECVID formalised evaluation with tolerance windows [13], and learned detectors such as TransNet V2 [14] are now the state of the art. FFmpeg’s scdet filter [15] is a widely available, training-free baseline. Video fingerprinting and near-duplicate retrieval [16] motivate our temporal signature, which is far simpler and is intended only for auditing small corpora.

TSM sits between these lines of work. It sacrifices the precision of learned models for determinism, speed, and zero dependencies beyond FFmpeg, and every metric is defined so that its expected response to a known distortion can be tested directly.

4. The TSM Metric Suite

4.1. Preprocessing

Each clip is decoded by FFmpeg, resampled to 12 frames per second with the fps filter, scaled to 160×90 pixels with area interpolation (aspect ratio not preserved), and converted to 8-bit luma. Frames are mapped to $[0, 1]$, giving a sequence $F_1, \dots, F_T$. A second decode at the native frame rate is used only for the duplicate-frame ratio (M6). Resampling to a fixed rate makes values comparable across 24, 30, and 60 fps sources.

4.2. Definitions

The basic quantity is the mean absolute frame difference:

$$D_t = \text{mean} |F_{t+1} - F_t|, \quad t = 1, \dots, T-1 \quad (1)$$

M1, Activity. $A = \text{median}_t D_t$. The median is robust to a few cut spikes.

M2, Freeze ratio. The share of sampled transitions that are nearly static:

$$FR = (1 / (T-1)) \cdot \sum_t I[D_t < \tau_f], \quad \tau_f = 0.002 \quad (2)$$

M3, Flicker index. Global brightness oscillation, measured as the second difference of mean luma μ_t and scaled by 10^3 :

$$FI = 10^3 \cdot \text{mean}_t |\mu_{t+1} - 2\mu_t + \mu_{t-1}| \quad (3)$$

A second difference is insensitive to slow, monotonic exposure changes such as a fade, but responds strongly to oscillation. Transitions adjacent to a detected cut are excluded.

M4, Temporal jitter ratio. The share of temporal change that is non-smooth:

$$J = \sum_t \|F_{t+1} - 2F_t + F_{t-1}\|_1 / \sum_t (\|F_{t+1} - F_t\|_1 + \|F_t - F_{t-1}\|_1) \quad (4)$$

By the triangle inequality, $0 \leq J \leq 1$. A purely linear intensity ramp gives $J = 0$, independent temporal noise gives about 0.87, and strict frame-to-frame alternation gives 1. Natural motion lies in between, because moving edges produce non-zero second differences. J is therefore a relative measure, interpreted against a baseline or under perturbation, not an absolute threshold. Cut-adjacent transitions are excluded.

M5, Shot boundaries. To resist camera flashes and exposure jumps, each frame first has its mean removed: $Z_t = F_t - \mu_t$. Let $G_t = \text{mean} |Z_{t+1} - Z_t|$, and let s_t be the share of the 16 cells in a 4×4 grid whose mean absolute difference is at least 0.05. A boundary is declared at t when

$$G_t \geq 0.08 \wedge G_t \geq 3 \cdot \text{median}(G_{t-6}, \dots, G_{t+6}) \wedge s_t \geq 0.75 \quad (5)$$

The local-median condition separates isolated spikes from sustained fast motion. The spatial-spread condition rejects changes confined to part of the frame.

M6, Duplicate-frame ratio. At the native frame rate, DR is the share of consecutive frame pairs with mean absolute difference below $\tau_d = 0.0005$. Frame-rate up-conversion by repetition yields $DR \approx 0.5$, and frozen segments also raise DR.

4.3. Temporal Signature

The series $D1 \dots T-1$ is a compact fingerprint of how a clip changes over time. It is expected to be largely invariant to re-encoding, resolution changes, and moderate re-framing, because cuts, motion bursts, and pauses occur at the same moments. For two clips a and b , the signature similarity is the Pearson correlation of their series truncated to the common length:

$$\rho(a, b) = \text{corr}(D_{1 \dots n}^a, D_{1 \dots n}^b), \quad n = \min(T_a, T_b) - 1 \quad (6)$$

Pairs with $\rho \geq 0.9$ are flagged for visual review. The threshold is a screening convention, not a validated decision boundary.

Table 1. Fixed TSM v1.0 parameters. None was tuned on the case-study corpus after annotation.

Parameter	Value	Used by
Analysis rate; frame size	12 fps; 160 × 90 luma, area scaling	M1–M5, signature
Freeze threshold τ_f	0.002 (0.2% of luma range)	M2
Cut: absolute / relative / window	0.08 / $3 \times \text{local median}$ / ± 6 samples	M5
Cut: grid / cell threshold / spread	4×4 / 0.05 / 0.75	M5
Duplicate threshold τ_d	0.0005 at native rate	M6
Signature review threshold	$\rho \geq 0.9$	Section 4.3

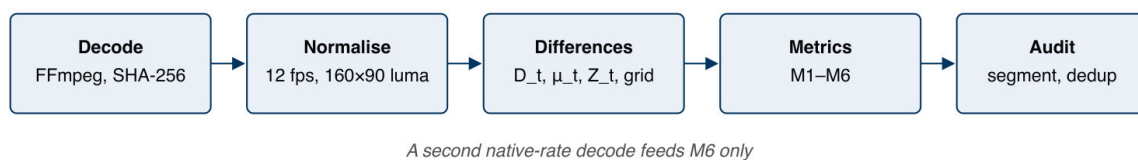


Figure 1. TSM processing pipeline. All steps are deterministic for a given FFmpeg build.

5. Validation Design

5.1. Controlled Perturbations

Ground truth for temporal failures is rarely available for real AI video, so we created it. From the case-study corpus (Section 6), every unique clip in which neither TSM nor scdet detected a boundary was selected as a base clip, giving 15 clips. Each was transformed with FFmpeg 7.1.1 into seven versions, each re-encoded with libx264 at CRF 18 unless stated otherwise:

identity: re-encode only, which serves as the reference for every comparison;
freeze_1s: the middle frame held for 24 extra frames (loop filter);
flicker_3hz: brightness offset of $0.06 \cdot \sin(2\pi \cdot 3t)$ (eq filter, per-frame evaluation); **dup_12to24:** down-sampled to 12 fps and repeated back to 24 fps;
temporal_noise: temporally varying noise of strength 14 (noise=all=14:allf=t); **hard_cut:** the first half of the clip followed by the second half of the next base clip; and **reencode_crf35:** re-encoding at CRF 35, as a robustness check.

For each perturbation, a *hit* is a change of the target metric in the expected direction relative to **identity**. Changes in non-target metrics are reported to expose coupling.

5.2. Shot-Boundary Annotation

For every unique clip in which either TSM or scdet (threshold 10) reported a boundary, all detections were pooled. Detections within 0.2 s were merged, giving 57 candidate events in 10 clips. For each event, frames were extracted at ± 0.25 s and at ± 0.07 s and inspected side by side. Each event was labelled *cut* (change of shot or viewpoint), *non-cut* (same shot, for example a lighting change, occlusion, or fast motion), or *uncertain* (for example a possible jump cut within similar framing, or a transition to black). Uncertain events are excluded from precision and recall. Because candidates

come only from the two detectors, recall is relative to the pooled set and is optimistic for both. Boundaries missed by both detectors are not counted.

5.3. Compression Robustness

Robustness is the median relative change of each continuous metric between **identity** and **reencode_crf35**, together with the maximum absolute change.

6. Test Corpus

The case study uses publicly displayed showcase videos from three AI video websites, chosen as a convenience sample of real AI video content that anyone can retrieve. On 10 October 2026, each homepage was loaded in a headless Chromium browser and scrolled, and every requested media URL ending in .mp4 was recorded. The pages were <https://dreaminai.app/> (10 URLs), <https://freepik.studio/> (10 URLs), and <https://higsfield.io> (15 URLs). All 35 files returned HTTP 200 and decoded as 24 fps H.264. Each file was hashed with SHA-256 and probed with FFprobe before analysis.

Table 2. Corpus composition after exact deduplication.

Source	Files	Unique after SHA-256	Resolutions	Duration (s)	Audio track	Samples at 12 fps
dreaminai.app	10	10 (shared set)	1280×720; 1280×548; 600×1066	10.0–30.1 (median 20.9)	9 / 10	2525
freepik.studio	10					
higsfield.io	15	15	600 wide; 338–1066 high	5.1–12.1 (median 8.0)	13 / 15	1462
Total	35	25				3987

All 10 files from freepik.studio were byte-identical to the 10 files from dreaminai.app; both sites serve them under the same relative path. The effective sample is therefore 25 clips, not 35, and the two sites are analysed as one *shared set*. Treating the duplicates as independent observations would have inflated the corpus by 40% and duplicated every statistic of the shared set. This is exactly the error that checksum auditing is meant to prevent.

The origin of showcase clips is not documented on the pages: which generator, which prompt, and whether editing or post-production was applied are all unknown. TSM values on this corpus therefore describe the files as published, not the capability of any generator or website.

7. Results

7.1. Controlled Perturbations

Table 3. Response to controlled perturbations ($n = 15$ base clips). Δ is the median change relative to **identity**.

Perturbation	Target metric	Hits	Median Δ target	Notable non-target effects (median Δ)
freeze_1s	FR	15/15	+0.102	DR +0.111 (expected: frozen frames are duplicates); A -0.002; cuts 0
flicker_3hz	FI	15/15	+64.4	A +0.046; J -0.081; cuts 0
dup_12to24	DR	15/15	+0.502	all others $ \Delta \leq 0.02$
temporal_noise	J	15/15	+0.012	FR fell to 0 in all 6 clips that had FR > 0

Perturbation	Target metric	Hits	Median Δ target	Notable non-target effects (median Δ)
hard_cut	cuts	14/15	+1	15 detections in total: one clip missed, one clip with 2 detections

Every target metric responded as designed. Three couplings are worth noting. First, flicker raises activity, because a brightness offset changes every pixel. Analysts should read A together with FI, not alone. Second, flicker lowers J, because a smooth sinusoid adds a large low-curvature component to the denominator of Equation (4). J is therefore a measure of roughness, not of oscillation. Third, temporal noise eliminated every non-zero freeze ratio: noise masks stillness. A clip with heavy grain can hide frozen content from FR, a limitation shared by all difference-based freeze detectors. The **identity** versions produced no boundaries in any of the 15 clips.

7.2. Shot-Boundary Detection

Of the 57 candidate events, 36 were labelled *cut*, 13 *non-cut*, and 8 *uncertain*.

Table 4. Boundary detection against visual labels (49 certain events; recall relative to the pooled candidate set).

Detector	TP	FP	FN	Precision	Recall	F1
TSM M5	34	11	2	0.756	0.944	0.840
FFmpeg scdet (threshold 10)	35	5	1	0.875	0.972	0.921
TSM $\wedge$ scdet	33	3	3	0.917	0.917	0.917

TSM’s 11 false positives fall into three groups: local lighting changes such as light panels switching on or dimming within one shot (6), large foreground objects or hands crossing close to the lens (4), and fast hand motion (1). Mean removal rejects global flashes but not local illumination changes, and the 0.75 spread threshold is not enough when an object covers most of a portrait frame. scdet had fewer false positives on this corpus but also responded to lighting changes and fast motion. In an earlier development run on clips from outside this study, scdet fired repeatedly on photographic flashes that TSM’s mean removal ignored, so the two detectors fail in complementary ways. For auditing, we recommend the conjunction of both detectors. It reduced false positives to 3, at the cost of two more misses than scdet alone, and gave the highest precision.

7.3. Compression Robustness

Re-encoding at CRF 35 changed the metrics by a median relative amount of 2.1% (A), 1.1% (FR), 1.3% (FI), 0.18% (J), and 3.4% (DR). The largest absolute changes were 0.042 in FR, 0.30 in FI, 0.008 in J, and 0.022 in DR. No boundary count changed. TSM values are therefore stable under the level of compression typical of web delivery, which matters because published AI video is almost always re-encoded.

7.4. Structural Profiles of the Case-Study Corpus

Table 5. TSM medians and structural flags per source (unique clips). Per-clip values are in Appendix A.

Set	<i>n</i>	A	FR	FI	J	Cuts (median)	DR	Clips with ≥ 1 cut	FR > 0.1	DR > 0.1
Shared set										
(dreaminai.app = freepik.studio)	10	0.020	0.146	3.92	0.792	1.5	0.047	6/10	6/10	4/10
higsfield.io	15	0.027	0.000	4.18	0.772	0	0.000	4/15	3/15	2/15

The two sets differ in structure more than in stability. The shared set consists of longer clips (median 20.9 s), and 6 of its 10 clips contain at least one boundary. One clip, dreamina-illuminated-portal-dance, contains 29 detected boundaries in 30 s, which makes it a music-video-style montage. Six shared-set clips also have freeze ratios above 0.1 and four have duplicate ratios above 0.1. This pattern is consistent with edited compilations that include held frames, rather than with single uncut generations. The higgsfield.io clips are shorter (median 8.0 s), mostly single-shot, and usually free of frozen or duplicated segments.

Flicker indices were similar in the two sets (medians 3.92 and 4.18). The highest values were 18.3 (higgsfield-visual-effects-wild-ride), 15.6 (dreamina-alley-chase), and 14.5 (the montage above). For the first of these, which was also a perturbation base clip, the value was unchanged after CRF 35 re-encoding (18.28 vs 18.27),

so it is not a compression artefact. Whether these values reflect intended lighting effects in the content or generative flicker cannot be determined from TSM alone.

The practical consequence is methodological. A per-clip temporal-consistency score computed on this corpus without segmentation would mostly measure editing density. Using any of the flags in Table 5 (a boundary, $FR > 0.1$, or $DR > 0.1$), TSM marks 8 of 10 shared-set clips and 6 of 15 higgsfield.io clips for segmentation or trimming before a model-level stability measure is applied. One of the latter, higgsfield-visual-effects-floating-fall, was flagged only by boundaries later labelled non-cuts (foreground objects crossing the lens), which shows why flagged clips should be reviewed rather than discarded.

7.5. Duplicates and Variants

Exact deduplication by SHA-256 removed 10 of 35 files (Section 6). The temporal signature (Equation 6) was then computed for all 300 pairs of the remaining 25 clips. One pair exceeded the 0.9 review threshold: dreamina-mountain-portrait and dreamina-mountain-widescreen ($\rho = 0.933$), whose file names describe portrait and widescreen versions of the same mountain scene. They differ in resolution (600×1066 and 1280×548) and in bytes, so checksums cannot link them. The next-highest correlation was 0.808. Pairs below the threshold were not reviewed, and the absence of further variants is not claimed.

8. Discussion

What TSM is for. TSM is a pre-screen and an auditing tool. Without a GPU or learned model, it answers whether a file contains multiple shots, frozen or padded segments, brightness oscillation, or unusual temporal roughness. It also identifies duplicate and variant files within a corpus. These answers determine whether a downstream comparison, human or learned, is fair.

What TSM is not. TSM cannot tell good motion from bad motion. A walking figure whose legs swap places, a classic generative failure, may have ordinary A and J values. Semantic and physical failures require human rating or learned models [6,7].

Recommended use. (1) Hash every file and drop exact duplicates. (2) Compute signatures and review pairs with $\rho \geq 0.9$. (3) Segment clips at boundaries detected by both TSM and scdet. (4) Report FR and DR to flag frozen and padded content. (5) Read FI with A, and J only relative to a baseline. (6) Proceed to perceptual evaluation only on segments that pass these checks.

9. Limitations

The perturbation set is synthetic, and real generative flicker or boil may differ in spectrum and spatial extent. Fifteen base clips limit the precision of hit rates. The boundary annotation was a single pass, and recall is relative to a pooled candidate set. The case-study corpus is a convenience sample of curated showcase files whose provenance is undocumented, and it does not represent the output distribution of any generator. The fixed parameters were set from first principles and a small development run. The boundary rule was revised once during development, after false positives

from a flash-lit scene and from foreground objects were inspected on clips outside the study corpus. The parameters have not been cross-validated on an independent annotated set. Area scaling to 160 × 90 discards fine texture, so very small-scale boil may be under-measured by J. Finally, TSM values depend on the decoder and scaler, so reports should state the FFmpeg version.

10. Reproducible Execution Workflow

1. Record each source page, access time, and media URL list. Download files without transcoding.
2. Compute SHA-256 and FFprobe metadata. Deduplicate exact copies before counting.
3. Run TSM v1.0 with the Table 1 parameters, and run scdet with a stated threshold.
4. Compute signatures and review pairs with $\rho \geq 0.9$.
5. Segment at conjunctive boundaries, then report M1–M6 per segment and per clip.
6. Publish the manifest, hashes, per-clip values, annotations, and scripts. Redistribute media only where permitted.

11. Conclusion

Temporal failures are the defining weakness of AI-generated video, yet the tools for measuring them are often heavy and hard to reproduce. TSM offers six interpretable, training-free measures and a temporal signature, each validated against a distortion with known ground truth. Every target metric responded to its perturbation in at least 14 of 15 clips, the values were stable under strong compression, and the boundary detector became competitive with an established baseline when the two were combined. Applied to public showcase files from <https://dreaminai.app/>, <https://freepik.studio/>, and <https://higsfield.io>, TSM reduced 35 files to 25 unique clips, identified a re-framed variant pair, and showed that much of the structural variation in one set reflects editing (multi-shot montage and held frames), which must be controlled before generation quality is compared. The central lesson is procedural: structure, duplication, and padding should be measured and controlled before any claim about temporal quality is made.

Appendix A. Per-Clip TSM Values

Table A1. TSM v1.0 values for the 25 unique clips. Shared-set clips are listed under their dreaminai.app identifiers; freepik-01 to freepik-10 are byte-identical to dreaminai-01 to dreaminai-10. SHA-256 values are truncated to 12 hexadecimal characters.

ID	File stem	W×H	Dur (s)	A	FR	FI	J	Cuts (TSM)	Cuts (scdet)	DR	SHA-256
dreaminai-01	dreamina-alley-chase	1280×720	30.0	0.0769	0.000	15.56	0.818	1	0	0.000	caddf108b920
dreaminai-02	dreamina-background-edit	1280×720	10.0	0.0037	0.150	0.73	0.705	0	0	0.021	b7a90f9b5fb2a
dreaminai-03	dreamina-city-crossing	1280×720	30.0	0.0412	0.058	2.40	0.820	0	0	0.007	d1443063a033
dreaminai-04	dreamina-eyewear-product-ad	1280×720	15.0	0.0096	0.256	2.28	0.802	0	0	0.228	4090889b1cf7
dreaminai-05	dreamina-illuminated-portal-dance	1280×720	30.1	0.0438	0.142	14.53	0.786	29	25	0.117	61638b4be4a2
dreaminai-06	dreamina-mountain-portrait	600×1066	15.0	0.0209	0.178	3.35	0.799	2	2	0.139	a4c7d74f9130

ID	File stem	W×H	Dur (s)	A	FR	FI	J	Cuts (TSM)	Cuts (scdet)	DR	SHA-256
dreamin-ai-07	dreamina-mountain-widescreen	1280×548	15.0	0.0256	0.339	1.49	0.751	3	3	0.189	8c823d389140
dreamin-ai-08	dreamina-orbital-jet	1280×720	10.4	0.0170	0.008	4.66	0.765	0	0	0.000	9006161c54e6
dreamin-ai-09	dreamina-shanghai-street-portrait	1280×720	26.8	0.0184	0.003	4.48	0.776	8	7	0.002	e65c0a470523
dreamin-ai-10	dreamina-stairwell-action	1280×720	27.8	0.0113	0.222	6.72	0.797	10	8	0.074	be9d50553a92
higgsfield-01	higgsfield-visual-effects-act-natural	600×1066	6.1	0.0388	0.000	3.58	0.774	0	0	0.000	e27bef0f2d8e
higgsfield-02	higgsfield-visual-effects-burning-man	600×800	7.1	0.0167	0.000	1.73	0.713	0	0	0.000	82d639de0b9f
higgsfield-03	higgsfield-visual-effects-cutout	600×450	8.1	0.0059	0.354	6.39	0.835	1	0	0.203	7589dc136e14
higgsfield-04	higgsfield-visual-effects-eyes-in	600×1066	7.1	0.0640	0.000	4.18	0.667	0	0	0.006	562d12452fb1
higgsfield-05	higgsfield-visual-effects-floating-fall	600×1066	12.1	0.0097	0.007	2.71	0.742	3	0	0.000	0b4f12b74f99
higgsfield-06	higgsfield-visual-effects-high-flip	600×1066	9.0	0.0236	0.000	8.99	0.728	0	0	0.018	14dcd7cc823c
higgsfield-07	higgsfield-visual-effects-incline	600×800	7.1	0.0147	0.143	2.83	0.772	0	0	0.083	57dc508c9db4
higgsfield-08	higgsfield-visual-effects-lacewalker	600×338	8.1	0.0154	0.000	2.97	0.633	0	0	0.000	58c69807fb8f
higgsfield-09	higgsfield-visual-effects-melting	600×600	5.7	0.0047	0.313	0.27	0.655	0	0	0.216	d9b57900e73b
higgsfield-10	higgsfield-visual-effects-selfception	600×800	5.1	0.0683	0.000	4.30	0.805	0	0	0.000	62027a4978df
higgsfield-11	higgsfield-visual-effects-smash-and-grab	600×1066	10.1	0.0292	0.050	4.39	0.843	1	1	0.013	e8ae653b38f0
higgsfield-12	higgsfield-visual-effects-street-colossus	600×1066	10.1	0.0270	0.000	4.68	0.753	2	2	0.017	973c4a51528f
higgsfield-13	higgsfield-visual-effects-studio-slide	600×450	8.0	0.0449	0.000	3.31	0.809	0	0	0.000	7e932c3f42e4
higgsfield-14	higgsfield-visual-effects-wild-ride	600×1066	10.1	0.0728	0.000	18.29	0.822	0	0	0.000	4364bfdcef0d
higgsfield-15	higgsfield-visual-effects-world-morphing	600×1066	8.0	0.0655	0.000	4.44	0.793	0	0	0.000	ec58fc958b61

Appendix B. Reporting Checklist

Corpus: source pages and access time; URL manifest; file count before and after SHA-256 deduplication; signature-flagged pairs and review outcome.

Processing: TSM version and parameters; FFmpeg version; analysis rate and frame size; scdet threshold.

Outputs: per-clip M1–M6; segmentation rule; per-segment values if used; annotation protocol and label counts if boundaries were verified.

Boundaries: no inference of generator identity or quality from TSM alone; no attribution of results to organizations whose names resemble a source domain.

Data Availability: No supplementary datasets or analysis code are included with this submission. Per-clip TSM values are reported in Appendix A. Source media were accessed from <https://dreaminai.app/>, <https://freepik.studio/>, and <https://higsfield.io/>; their availability is subject to change by the site operators.

Acknowledgments and AI-Assistance Disclosure: The author selected the research question, study scope, sources, analysis decisions, and publication venue, and remains responsible for the manuscript. A generative text system assisted with literature discovery, editing, document preparation, and implementation of the analysis script. The study used publicly displayed source media and did not generate new endpoint outputs.

Conflicts of Interest: The author declares no conflicts of interest and no affiliation with dreaminai.app, freepik.studio, or higsfield.io.

References

1. Ho, J., Salimans, T., Gritsenko, A., Chan, W., Norouzi, M., & Fleet, D. J. (2022). Video Diffusion Models. *Advances in Neural Information Processing Systems*, 35. <https://arxiv.org/abs/2204.03458>
2. Blattmann, A., Dockhorn, T., Kulal, S., et al. (2023). Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets. arXiv:2311.15127. <https://arxiv.org/abs/2311.15127>
3. Yang, Z., Teng, J., Zheng, W., et al. (2024). CogVideoX: Text-to-Video Diffusion Models with an Expert Transformer. arXiv:2408.06072. <https://arxiv.org/abs/2408.06072>
4. Kong, W., Tian, Q., Zhang, Z., et al. (2024). HunyuanVideo: A Systematic Framework for Large Video Generative Models. arXiv:2412.03603. <https://arxiv.org/abs/2412.03603>
5. Wan Team. (2025). Wan: Open and Advanced Large-Scale Video Generative Models. arXiv:2503.20314. <https://arxiv.org/abs/2503.20314>
6. Huang, Z., He, Y., Yu, J., et al. (2024). VBench: Comprehensive Benchmark Suite for Video Generative Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21807–21818. <https://arxiv.org/abs/2311.17982>
7. Liu, Y., Cun, X., Liu, X., et al. (2024). EvalCrafter: Benchmarking and Evaluating Large Video Generation Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. <https://arxiv.org/abs/2310.11440>
8. Unterthiner, T., van Steenkiste, S., Kurach, K., Marinier, R., Michalski, M., & Gelly, S. (2018). Towards Accurate Generative Models of Video: A New Metric & Challenges. arXiv:1812.01717. <https://arxiv.org/abs/1812.01717>
9. Teed, Z., & Deng, J. (2020). RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. *Proceedings of the European Conference on Computer Vision*, 402–419. <https://arxiv.org/abs/2003.12039>
10. Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 586–595. <https://arxiv.org/abs/1801.03924>
11. Wu, H., Zhang, E., Liao, L., et al. (2023). Exploring Video Quality Assessment on User Generated Contents from Aesthetic and Technical Perspectives. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. <https://arxiv.org/abs/2211.04894>
12. Boreczky, J. S., & Rowe, L. A. (1996). Comparison of Video Shot Boundary Detection Techniques. *Journal of Electronic Imaging*, 5(2), 122–128.
13. Smeaton, A. F., Over, P., & Doherty, A. R. (2010). Video Shot Boundary Detection: Seven Years of TRECVID Activity. *Computer Vision and Image Understanding*, 114(4), 411–418.
14. Souček, T., & Lokoč, J. (2020). TransNet V2: An Effective Deep Network Architecture for Fast Shot Transition Detection. arXiv:2008.04838. <https://arxiv.org/abs/2008.04838>

15. FFmpeg Developers. (2025). FFmpeg Filters Documentation: scdet. Retrieved 10 October 2026, from <https://ffmpeg.org/ffmpeg-filters.html#scdet-1>
16. Wu, X., Hauptmann, A. G., & Ngo, C.-W. (2007). Practical Elimination of Near-Duplicates from Web Video Search. *Proceedings of the 15th ACM International Conference on Multimedia*, 218–227.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.