Do Better-Looking Videos Stay in Memory? A VBench Study of Commercial and Brand Memorability

Hao-Tien Yu[1,*], Yi-En Dong[1]

[1]Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan

Abstract

This Quest-for-Insight paper starts from the MediaEval 2026 question: "Is there a relationship between the aesthetic quality of media and its memorability, or do these factors function independently?" We study this question in commercial videos by evaluating seven VBench dimensions on 210 VBench-covered VIDEM development videos and comparing them with a parallel movie-clip analysis. The main result is negative: the VBench *aesthetic_quality* proxy, although weakly associated with movie memorability, is essentially unrelated to commercial video memorability ($\rho = 0.0038$, $p = 0.956$, bootstrap 95% CI $[-0.1395, 0.1455]$). No VBench dimension survives false discovery rate (FDR) correction for either video or brand memorability. Brand memorability, however, is strongly coupled with video memorability ($\rho = 0.7178$, 95% CI $[0.6295, 0.7888]$). Commercial memorability appears to depend less on generic visual-quality proxies and more on brand exposure, message content, speech, text, audio, and video format.

1. Introduction

This work addresses a MediaEval 2026 Quest-for-Insight question: "Is there a relationship between the aesthetic quality of media and its memorability, or do these factors function independently?" [1] The question echoes visual memorability work showing that memorability is not simply visual pleasantness or photographic quality. Isola et al. found aesthetic quality only weakly related to photograph memorability [2]. Later large-scale image work showed that memorability prediction depends heavily on semantic and localized visual evidence, not only global appearance [3].

Video memorability work points in the same direction. Memento10k and modular video memorability models use visual, semantic, scene, event, and memory-consolidation evidence [4, 5]; gaze-attention work also treats memorability as a temporal attention problem [6]. These studies motivate a simple test: do generic visual-quality descriptors explain commercial videos, or does commercial memorability need content-specific evidence?

We extend the question from photographs and movie clips to commercial videos. Here, memorability is not just visual appeal: a commercial may be remembered because it is polished, repeats the brand, shows the product, has a distinctive message, or is easy to parse. Long-term advertisement memorability work similarly ties ad memory to multimodal content and brand exposure [7]. VIDEM lets us test whether aesthetic quality behaves like an independent visual property or contributes to video and brand recall [1, 8, 9].

MediaEval'26: Multimedia Evaluation Workshop, June 15–16, 2026, Amsterdam, Netherlands and Online
*Corresponding author.
charlessworknp@gmail.com (H. Yu); jddlake@gmail.com (Y. Dong)
0009-0008-9767-0428 (H. Yu); 0009-0006-1486-547X (Y. Dong)

Table 1
VBench visual proxies.

Dimension	Proxy
aesthetic_quality	visual appeal
imaging_quality	technical quality
temporal_flickering	frame instability
motion_smoothness	motion continuity
dynamic_degree	amount of motion
subject_consistency	foreground stability
background_consistency	background stability

We ask whether VBench-style video-quality metrics explain these targets. VBench was designed for generated video evaluation and is grounded in a large human-aligned benchmark of video quality [10]. We use seven prompt-free dimensions as visual descriptors for real videos: *aesthetic_quality*, *imaging_quality*, *temporal_flickering*, *motion_smoothness*, *dynamic_degree*, *subject_consistency*, and *background_consistency*. They can be computed directly from pixels, without new manual annotation.

A parallel analysis on the Movie Memorability Dataset [11] found a weak positive association between VBench *aesthetic_quality* and movie memorability. Commercial videos may not follow that pattern. They are communicative artifacts meant to make a brand or message memorable, not only to look pleasing. We test whether the movie-domain VBench aesthetic signal is also observed in commercial memorability and whether VBench features help explain brand recall.

This note is diagnostic rather than predictive. It tests the MediaEval aesthetic-quality question in commercial video, reports that the weak movie-domain VBench aesthetic signal is not observed, and points to brand recall and format heterogeneity for later modeling. The negative result narrows what generic visual-quality metrics can support: they describe how a video looks, but do not by themselves explain why a commercial or brand is remembered.

2. Data and Measures

We use the VIDEM development set for MediaEval 2026 Subtask 2 [8, 9]: 339 videos with *memorability_score* and *brand_memorability*. VBench evaluation was completed for 210 videos, and all commercial analyses use this subset; the remaining 129 videos were not available in the local VBench-ready subset. Matched and unmatched videos were similar in target scores (Mann-Whitney $p = 0.331$ for video memorability and $p = 0.873$ for brand memorability), duration ($p = 0.733$), and annotation count ($p = 0.837$), but matched videos had higher view counts ($p < 0.001$). We treat this as a coverage limitation, not an obvious label confound.

For comparison, we use a parallel VBench analysis of 520 MediaEval Subtask 1 Movie Memorability Dataset clips [11], using the same VBench dimensions but only movie *memorability_score*. We ask whether the movie-domain aesthetic signal also appears for commercial video and brand recall.

These are visual proxies rather than direct commercial-effectiveness judgments; prompt-dependent dimensions such as object class, scene, and color were excluded because the videos do not have generation prompts.

3. Analysis Protocol

Following prior work that examined aesthetic–memorability association rather than a strictly linear effect [2], Spearman correlation is our primary statistic. It tests whether higher VBench

Table 2
Core evidence. Effects are Spearman correlations except for the movie-commercial aesthetic difference; brackets are 95% confidence intervals.

Claim	Effect	p	FDR q
Aesthetic proxy vs. commercial memory	0.0038 [−0.1395, 0.1455]	0.9559	0.9559
Brand vs. video memory	0.7178 [0.6295, 0.7888]	< 0.001	–
Movie-commercial aesthetic difference	0.1537 [−0.0077, 0.3178]	0.0595	0.2081

scores tend to accompany higher memorability scores without assuming equal intervals or a linear relationship. We correlate each VBench dimension with *memorability_score* and *brand_memorability* on the 210 matched videos, compare with the movie-clip VBench analysis, stratify by duration and YouTube metadata category, and analyze brand residuals. Brand residuals are ordinary least-squares residuals from predicting *brand_memorability* from *memorability_score*. We also test VBench predictive value with five-fold cross-validated ridge regression using metadata features (duration, view count, engagement rate, annotation count, YouTube category) and the seven VBench dimensions.

We apply Benjamini-Hochberg false discovery rate correction within result families. We use 10,000 permutation tests for main and focused subset correlations, and 10,000 bootstrap resamples for main-correlation and movie-commercial aesthetic-comparison confidence intervals. Uncorrected effects are exploratory.

4. Results

4.1. Generic visual quality is not predictive

The VBench *aesthetic_quality* proxy is essentially uncorrelated with commercial video memorability ($\rho = 0.0038$, $p = 0.9559$, $q = 0.9559$; bootstrap 95% CI $[-0.1395, 0.1455]$). The permutation test gives the same conclusion ($p = 0.9554$, FDR $q = 0.9669$). The commercial-video estimate is closer to independence than the weak relationships reported in prior image and movie settings. No VBench dimension survives FDR correction for video memorability. For brand memorability, *imaging_quality* has a weak nominal negative association ($\rho = -0.1417$, $p = 0.0402$, bootstrap 95% CI $[-0.2733, -0.0058]$), but it disappears after correction ($q = 0.5625$; permutation FDR $q = 0.5108$) and is treated as exploratory.

4.2. The movie aesthetic signal is not observed in commercials

In movie clips, *aesthetic_quality* has a weak positive association with memorability ($\rho = 0.1575$, $p = 0.000312$; approximate 95% CI $[0.0725, 0.2402]$), but the commercial correlation is nearly zero. The movie-commercial difference is suggestive but not significant after correction (difference = 0.1537, Fisher $p = 0.0595$, FDR $q = 0.2081$; bootstrap 95% CI $[-0.0077, 0.3178]$), so we report it descriptively rather than as a confirmed interaction.

4.3. Brand recall follows video recall

Brand memorability strongly tracks general video memorability: $\rho = 0.7178$ (bootstrap 95% CI $[0.6295, 0.7888]$) in the full matched set, and high correlations in short videos ($\rho = 0.8153$), videos under five minutes ($\rho = 0.7351$), ad-like non-news videos ($\rho = 0.6925$), news/politics

videos ($\rho = 0.7296$), and education videos ($\rho = 0.7282$). Brand recall should therefore be modeled with general video recall, not as a separate visual-quality target.

4.4. Format heterogeneity limits generic predictors

A heuristic typology shows that most videos fall in a news/politics metadata group, not in short advertisement-like groups. We assign formats from simple metadata rules: *News & Politics* videos are the news/politics group, *Education* videos are education explainers, remaining videos under 60 seconds are short ad-like, remaining videos under 300 seconds are ad-like or explainers, and longer remaining videos are long presentations. These shorthand labels are not manually verified, but mark formats where visual quality may operate differently. The typology is long-tailed: 129 matched videos are in the *news/politics group*, 35 are *education explainer*, 26 are *ad-like or explainer*, 12 are *long presentation*, and only eight are *short ad-like*. In videos under five minutes, *dynamic_degree* has weak nominal positive associations with video and brand memorability, while *imaging_quality* has a weak nominal negative association with video memorability. In news/politics videos, *imaging_quality* and *background_consistency* have nominal negative associations with brand residuals after controlling for video memorability. None survive FDR correction, so they remain future-work hypotheses.

4.5. VBench adds little predictive value

Ridge regression mirrors the correlations. For video memorability, metadata-only, VBench-only, and metadata-plus-VBench models all produce negative out-of-fold Spearman correlations (-0.0815, -0.0997, and -0.0852). For brand memorability, metadata-only reaches $\rho = 0.0498$, VBench-only reaches $\rho = 0.0047$, and metadata plus VBench reaches $\rho = 0.0942$. VBench alone adds little signal, and metadata plus VBench gives only a small brand-recall gain.

5. Discussion and Outlook

In VIDEM, generic VBench visual-quality proxies do not explain commercial memorability well. The VBench dimensions used here capture only a limited view of video quality: aesthetics, technical quality, motion, flicker, and consistency. They do not measure logo visibility, spoken brand names, message text, product presence, or persuasive structure. This fits prior work tying memorability to semantic, localized, and event-level evidence beyond global aesthetics [3, 4, 5].

The weak movie signal for *aesthetic_quality* is not observed in VIDEM, where news/politics videos, explainers, presentations, and short ad-like videos serve different communication goals. A single generic visual-quality score is unlikely to behave consistently across these formats.

Three limitations qualify these results. VBench scores cover 210 of 339 development videos, and the matched subset has higher view counts than the unmatched videos; the analyzed subset may be more public-facing or popular. VBench was designed for generated-video evaluation, so its dimensions should be read as visual-quality proxies rather than validated advertising or memory features. The commercial typology is heuristic and should not replace manually verified format labels.

Future work should add commercial-communication features: logo and product visibility, OCR text, speech transcripts, brand-name mentions, shot rate, audio energy, narrative structure, and verified format labels. Brand memorability should also be predicted jointly with video memorability, since the two targets are strongly coupled.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT/Codex in order to: Grammar and spelling check, Paraphrase and reword. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] MediaEval 2026 Predicting Movie and Commercial Memorability Task Organizers, Memorability: Predicting movie and commercial memorability, https://multimediaeval.github.io/editions/2026/tasks/memorability/, 2026. Accessed: 2026-06-10.

[2] P. Isola, J. Xiao, D. Parikh, A. Torralba, A. Oliva, What makes a photograph memorable?, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (2014) 1469–1482. doi:10.1109/TPAMI.2013.200.

[3] A. Khosla, A. S. Raju, A. Torralba, A. Oliva, Understanding and predicting image memorability at a large scale, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2390–2398.

[4] A. Newman, C. Fosco, V. Casser, A. Lee, B. McNamara, A. Oliva, Multimodal memorability: Modeling effects of semantics and decay on video memorability, in: A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (Eds.), Computer Vision – ECCV 2020, Springer International Publishing, Cham, 2020, pp. 223–240. doi:10.1007/978-3-030-58517-4_14.

[5] T. Dumont, J. S. Hevia, C. L. Fosco, Modular memorability: Tiered representations for video memorability prediction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 10751–10760. doi:10.1109/CVPR52729.2023.01035.

[6] P. Kumar, E. Khandelwal, M. Tapaswi, V. Sreekumar, Seeing eye to AI: Comparing human gaze and model attention in video memorability, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2025, pp. 2082–2091.

[7] H. Si, S. Singh, Y. K. Singla, A. Bhattacharyya, V. Baths, C. Chen, R. R. Shah, B. Krishnamurthy, Long-term ad memorability: Understanding and generating memorable ads, 2023. doi:10.48550/arXiv.2309.00378. arXiv:2309.00378.

[8] I. Martín-Fernández, A. Ganesh, M. G. Constantin, C.-H. Demarty, M. Gil-Martín, S. Halder, B. Ionescu, A. Matran-Fernandez, R. Savran Kiziltepe, A. García Seco de Herrera, Overview of the MediaEval 2026 predicting movie and commercial memorability task, In *Proceedings of the MediaEval 2026 Workshop*, Amsterdam, The Netherlands and Online, 2026. To appear.

[9] R. S. Kiziltepe, S. Saha, R. Valladares Santana, F. Doctor, K. Paterson, D. Hunstone, A. García Seco de Herrera, VIDEM: VIDeo effectiveness and memorability dataset, in: Advances in Computational Intelligence, volume 16008 of *Lecture Notes in Computer Science*, Springer, 2025, pp. 41–54. doi:10.1007/978-3-032-02725-2_4.

[10] Z. Huang, Y. He, J. Yu, F. Zhang, C. Si, Y. Jiang, Y. Zhang, T. Wu, Q. Jin, N. Chanpaisit, Y. Wang, X. Chen, L. Wang, D. Lin, Y. Qiao, Z. Liu, Vbench: Comprehensive benchmark suite for video generative models, arXiv preprint arXiv:2311.17982 (2023). doi:10.48550/arXiv.2311.17982.

[11] R. Cohendet, K. Yadati, N. Q. K. Duong, C.-H. Demarty, Annotating, understanding, and predicting long-term video memorability, in: Proceedings of the 2018 ACM International Conference on Multimedia Retrieval, 2018, pp. 178–186. doi:10.1145/3206025.3206056.