Supplementary Material of "Generating Highlight Video of a User-Specified Length using Most Replayed Data"

A ADDITIONAL EXAMPLES FOR IMPLICATIONS OF MOST REPLAYED DATA AS AN INDICATOR FOR HIGHLIGHT CREATION

Extending our investigation of the implications of Most Replayed Data (MRD) as an indicator for highlight creation described in Section 2 of the main paper, we explore its effectiveness using eleven additional videos across various genres, including sports, documentaries, news, and education. For videos with explicit event timestamps such as officially marked scores in sports (e.g., goals in a soccer match or points in a basketball game), we analyzed the MRD's reflection of key events and calculated its correlation with other known engaging factors (i.e., in-video sound loudness and live chat frequency) similarly to the main paper. For videos lacking such explicit events or engaging materials like live chat data or significant variations in in-video sound (common in documentaries and educational videos), we examined whether the MRD peaks effectively capture important moments from the narrative of the original video. To achieve it, we first extracted the subtitle files (SRT files) from *YouTube*. Using ChatGPT-4, we generated a one-sentence summary of each video's main narrative with the prompt: *"Provide a one-sentence summary that captures the main idea concisely."*. Next, we identified the top five peaks with the highest MRD values using the `find_peaks` function from python `scipy` library with the prominence parameter set to 0.1. We then extracted the SRT content surrounding each peak, specifically the sentences containing the peak along with their adjacent sentences. Using ChatGPT-4 again, we summarized this contextual SRT content with the prompt: *"Given the following transcript segments, provide a concise summary reflecting context in 15 words or less."*. By executing this process, we obtained the key narrative summaries, MRD graphs, peak scenes, and contextual content of each peak. The following paragraphs provide descriptions of the analyses conducted for each video.

We first analyzed soccer [16], basketball [22], and ice hockey [25] videos (Figures A1, A2 and A3). Figure A1 shows a strong alignment between key soccer match moments and MRD peaks, with notable spikes in MRD values during goals (e.g., at 17, 28, 38, 62, 68, and 72 minutes), a close call (34 minutes), and the final decision of the World Cup gold medal winner (99 minutes). Interestingly, there is a MRD spike at 42 minutes despite no goal being scored at that time. Further investigation revealed that this spike corresponds to a free kick declared by the referee, a significant event of the game. Additionally, a smaller peak observed at 36 minutes was linked to a penalty kick awarded by the referee, leading to the goal scored around 38 minutes. The peak at 56 minutes was determined to be caused by a humorous moment in which a spectator, dressed in an unusual outfit, ran onto the field, causing the referee to briefly halt the match. Similarly, Figure A2 shows MRD spikes coinciding with key basketball moments, such as 2- or 3-point shots (at 11, 15, 18, 41, 56, and 62 minutes). These moments include fan-favorite plays such as a crucial 3-pointer made near the end of the first half (41 minutes) and an alley-oop dunk (62 minutes). Other key moments include fouls (35 and 81 minutes) and the dramatic game finish (88 minutes). In Figure A3, MRD peaks correspond to ice hockey goals (17, 51, 94, 98 minutes) and a power play (25 minutes), which occurred when one team gained a player advantage due to a penalty caused by the opposing team. Another notable spike (47 minutes) aligns with a close call, where the goalkeeper narrowly saved a goal. The final outcome of the Olympic match is reflected by a peak at 101 minutes, with the winner's medal ceremony following at 134 minutes. Interestingly, there is also a spike at 66 minutes, despite no key ice hockey event occurring; further analysis revealed that this spike corresponds to a replay of critical first-half goals.

We also examined the correlation between MRD and other known engaging factors, in-video sound and live *YouTube* chat frequency. The MRD graph shows similarities to both the in-video sound (blue curve) and chat frequency (gray dotted curve) graphs. Cosine similarities between MRD and these factors demonstrate high correlations: for soccer (in-video sound: 0.80, live chat frequency: 0.60), basketball (in-video sound: 0.82, live chat frequency: 0.80), and ice hockey (in-video sound: 0.91, live chat frequency: 0.73). Among the three graphs, the MRD graph aligns much better with the events at peak points than the other two factors. For example, in soccer (Figure A1), between 80 and 100 minutes, there are no significant match events, yet the live chat frequency and in-video sound show fluctuations. In basketball (Figure A2), the live chat frequency shows minimal changes in response to match events. In ice hockey (Figure A3), aside from a few key moments, the in-video sound and live chat frequency remain largely unaffected by other key moments, indicating that MRD provides a more reliable representation of viewer engagement around critical moments in the video than the alternatives.

We also analyzed documentary [18, 20], educational [36, 38], news [21, 23, 24], and speech [37] videos (Figures A4 to A11). We found that MRD effectively captures engaging points corresponding to the core narrative across these diverse genres. For instance, as shown in Figure A4, consider a documentary with the following one-sentence summary: *"This documentary follows the life cycle and survival challenges of harp seal pups in the melting Arctic, highlighting the urgent impact of climate change on their habitat and calling for global action to protect the fragile ecosystem."*. The indicated MRD peaks correspond to the segments discussing the harp seal pups facing survival challenges in the melting Arctic. The contextual summary for this peak accurately reflects the critical moment described in the narrative summary. In addition, MRD peaks shown in Figure A5 highlight the documentary's central themes and pivotal moments. Similarly, in educational videos, they correspond to innovative performances that captivate the audience (Figure A6) or key demonstrations of complex concepts (Figure A7). In news videos, they align with critical coverage of historic events like groundbreaking spaceflights (Figure A8) or significant policy changes (Figure A9). In the speech video, they highlight key moments where speakers share profound insights or deliver inspirational messages that deeply resonate with viewers (Figure A10). In the space phenomenon live broadcast, they coincide with awe-inspiring visuals of celestial events or expert explanations that engage and educate the audience (Figure A11).

Building on our qualitative assessment, we conducted a quantitative analysis to investigate whether moments with high MRD values capture key narrative elements more effectively than other video moments. We extracted three distinct types of points from each of the eight analyzed videos (the same videos used in Figures A4 to A11):

(1) **High MRD points:** The five highest points that were used in our analysis of eight videos, which were MRD peaks.
(2) **Random MRD points:** Five points selected randomly regardless of MRD values.
(3) **Low MRD points:** Five points selected randomly from MRD values below 0.1.

When selecting random points, we ensured that each point was at least 10 seconds apart from the others. For each type of point, we extracted the SRT content surrounding each point–specifically, the sentences containing the point along with their adjacent sentences. Each of the five SRT contexts was summarized by ChatGPT-4 (again using the prompt *"Given the following transcript segments, provide a concise summary reflecting context in 15 words or less."*) and then concatenated in chronological order to form a single paragraph. We then measured the semantic similarity between this paragraph and the one-sentence key narrative summary (previously generated by ChatGPT-4 with the prompt *"Provide a one-sentence summary that captures the main idea concisely."*). To calculate the semantic relationship between two texts, we used the sentence-transformer model `all-mpnet-base-v2` [29], a deep learning model that can perform sentence-pair regression tasks. We calculated the average and standard deviation of these similarities from eight videos. The results showed that the high MRD points achieved the highest average similarity ($M = 0.744, SD = 0.070$), followed by the random MRD points which can serve as a baseline ($M = 0.698, SD = 0.091$) and then the low MRD points ($M = 0.642, SD = 0.093$).

To supplement our analysis of MRD as an indicator for highlight creation described in Section 2 of the main paper, we have so far investigated eleven additional videos across various genres (Figures A1 to A11). In videos featuring explicit events, such as goals, we found that the spikes in MRD align with these events, outperforming other engaging factors like in-video sound and live chat frequency in highlighting key moments (Figures A1 to A3). For videos lacking such clear events, MRD peaks resonated with viewers, as their content aligned closely with each video's core narratives (Figures A4 to A11). Further quantitative comparisons showed that these MRD peaks tend to match a video's key narrative elements better than moments with low MRD values. Collectively, these findings suggest that MRD is a reliable indicator of engaging moments across various video genres.

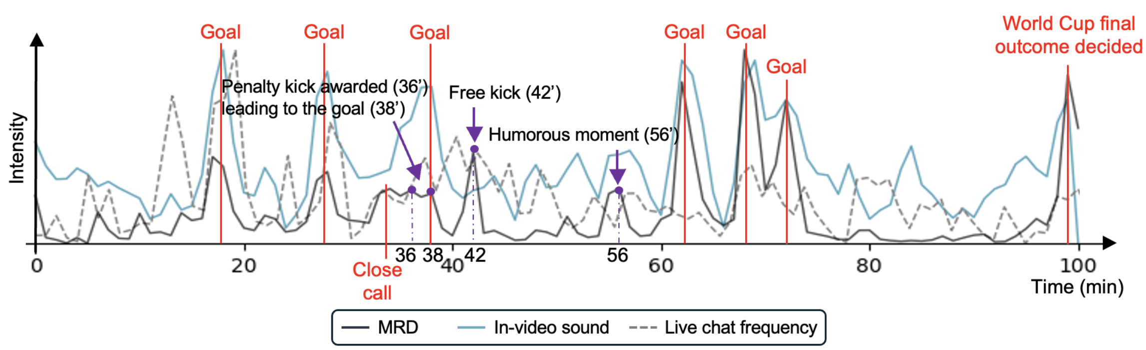

Fig. A1. The trend of MRD similar to other known engaging factors and its successful reflection of important events in a soccer match [16].

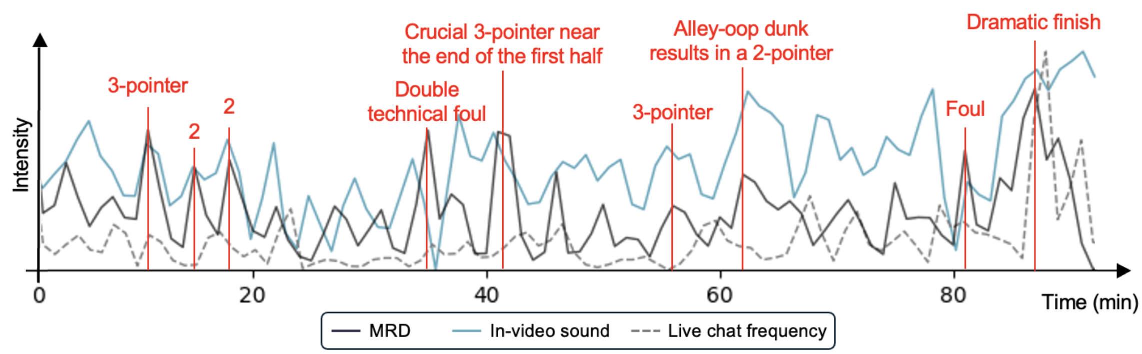

Fig. A2. The trend of MRD similar to other known engaging factors and its successful reflection of important events in a basketball match [22].

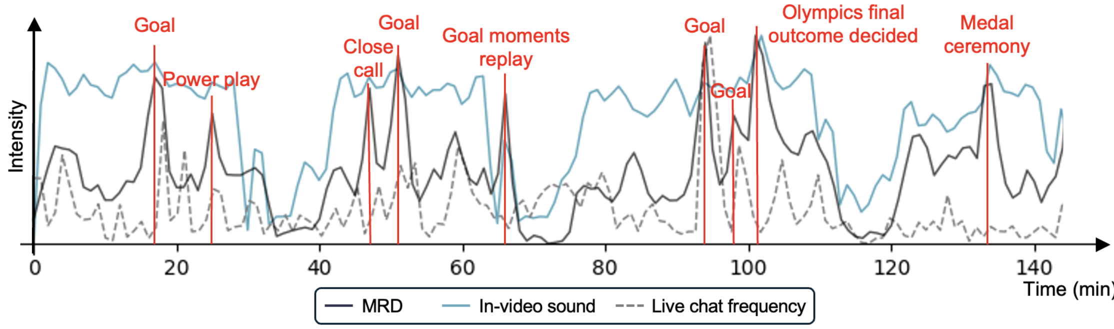

Fig. A3. The trend of MRD similar to other known engaging factors and its successful reflection of important events in an ice hockey match [25].

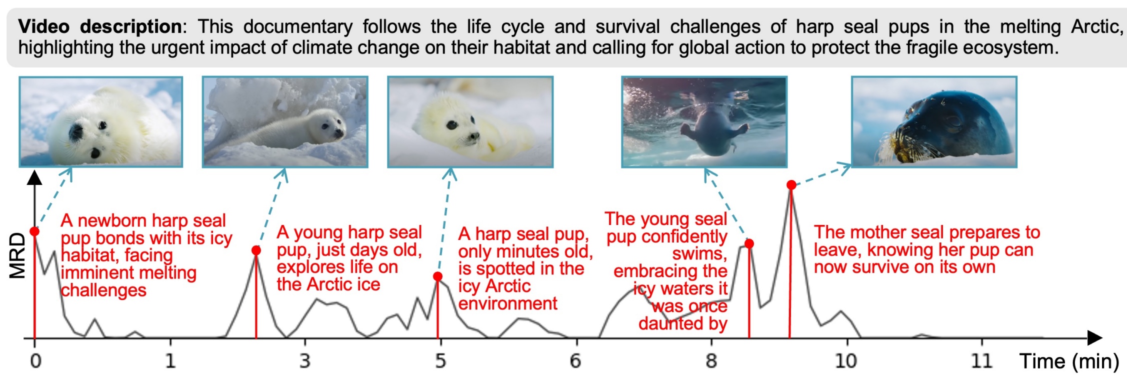

Fig. A4. MRD peaks align with critical moments in the documentary [18] about harp seal pups facing survival challenges in the melting Arctic. Image source: [18].

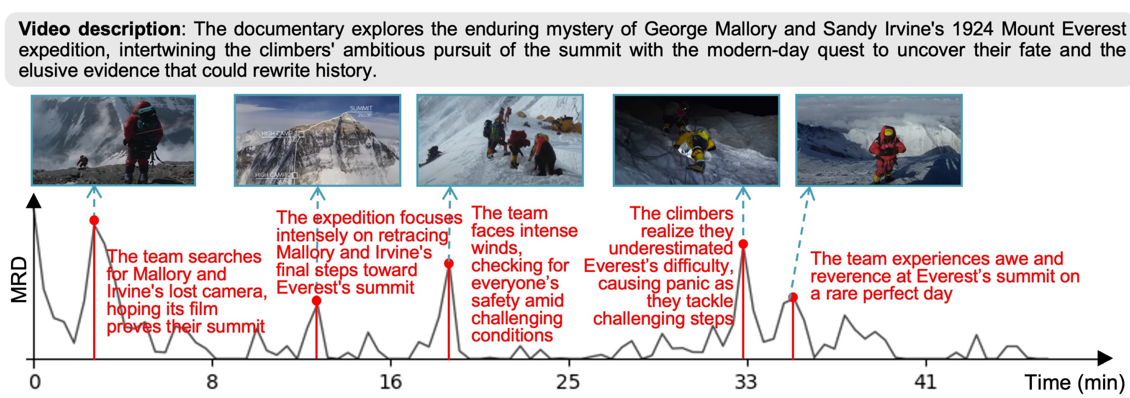

Fig. A5. MRD peaks align with critical moments in the documentary [20] such as the team's search for the lost camera and their final steps toward the summit to explore the mystery of Mallory and Irvine's 1924 Mount Everest expedition. Image source: [20].

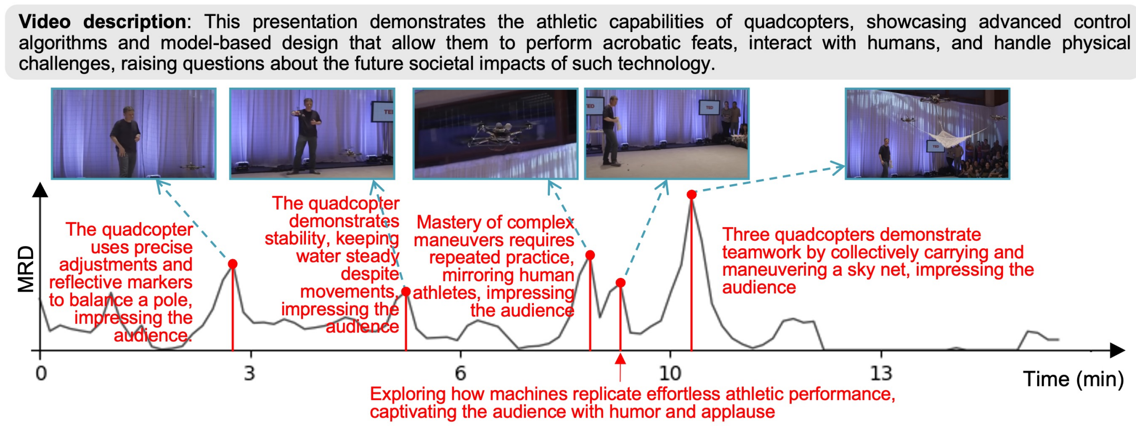

Fig. A6. MRD peaks align with critical moments in the educational video [36] showcasing quadcopters performing advanced acrobatic feats that impress the audience with applause. Image source: [36].

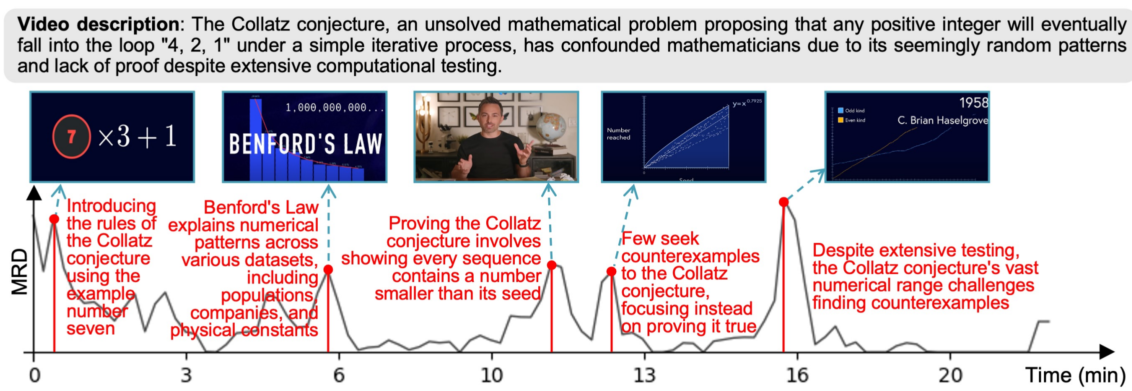

Fig. A7. MRD peaks align with critical moments in the educational video [38] explaining the definition of Collatz conjecture and its complexities. Image source: [38].

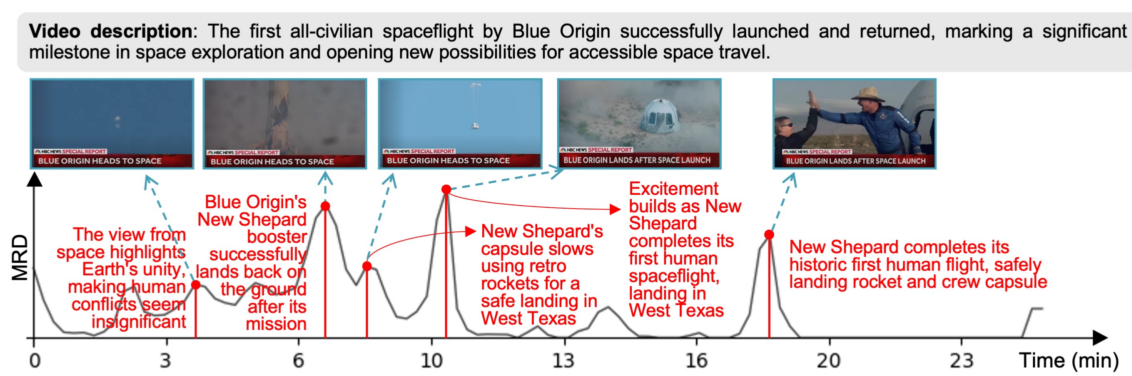

Fig. A8. MRD peaks align with critical moments in the news [24] of Blue Origin's first all-civilian spaceflight. Image source: [24].

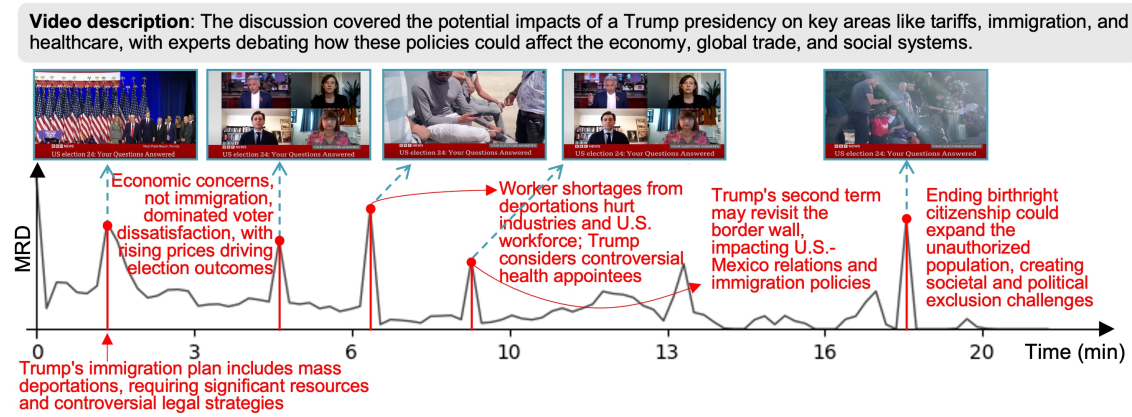

Fig. A9. MRD peaks align with critical moments in the news [23] discussing the potential impacts of a presidential policy shift. Image source: [23].

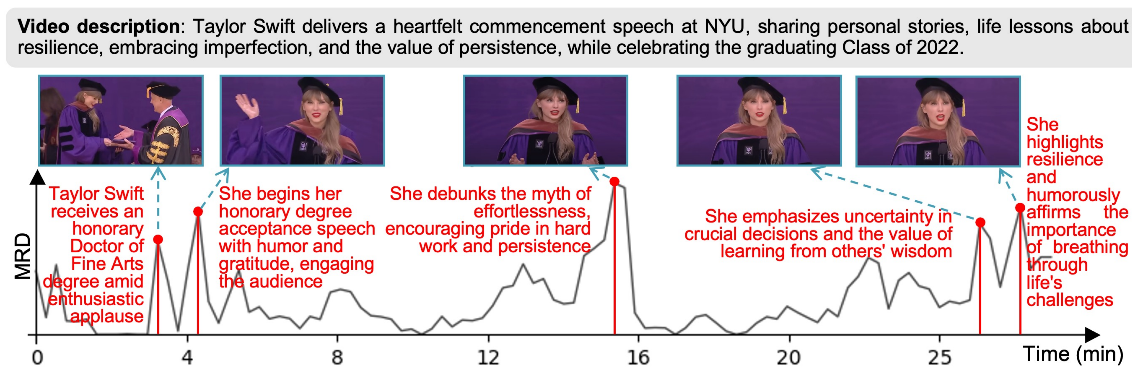

Fig. A10. MRD peaks align with critical moments in the speech [37] where the speaker delivers inspirational messages that resonate with the audience. Image source: [37].

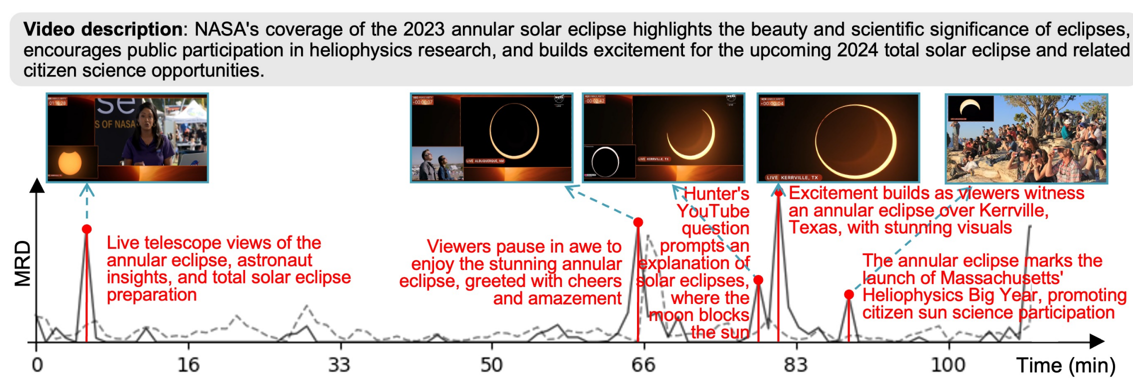

Fig. A11. MRD peaks align with critical moments in NASA's live broadcast [21] of the 2023 annular solar eclipse, showcasing awe-inspiring celestial visuals. In this video, YouTube live chat data is provided and plotted as a gray dotted line, with a cosine similarity of 0.59 between MRD and live chat frequency. The only peak around 66 minutes corresponds to the moment when the eclipse progressed to form a ring. Image source: [21].

B ADDITIONAL ANALYSIS OF VIDEO DATA USED IN USER STUDY 1

In this section, we further analyze the user responses per video data used in User Study 1 of the main paper. For each question on each video data, we conducted Friedman tests, and the results revealed significant differences ($p < 0.05$) among the three methods across all questions (Table B1). We performed post-hoc analyses using two-tailed Wilcoxon signed-rank tests with Holm-Bonferroni correction. Figures B12 to B16 visualize the results.

Table B1. Friedman statistics for Q1, Q2, and Q3 across videos.

	Golf [32]	Entertainment [40]	Tennis [41]	Documentary [19]	Soccer [31]
Q1: $\chi^2(2)$	24.35	8.96	22.51	25.47	12.70
Q2: $\chi^2(2)$	37.94	11.07	17.80	9.52	24.74
Q3: $\chi^2(2)$	33.13	12.14	23.91	21.31	15.60

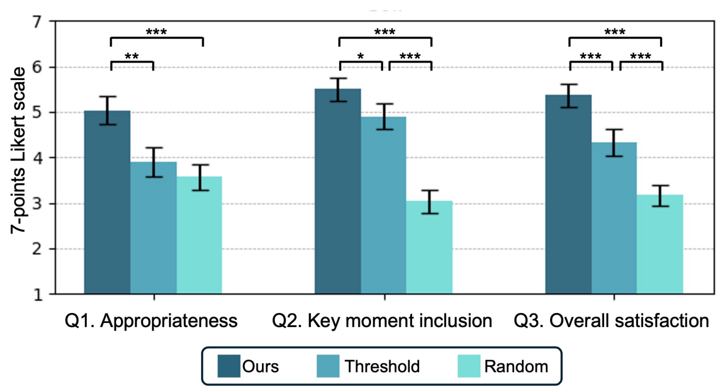

Fig. B12. 7-point Likert scale results of three methods for appropriateness (Q1), key moment inclusion (Q2), and overall satisfaction (Q3) of golf video [32]. Error bars represent the standard error of the mean. Asterisks show statistically significant differences between pairs of methods (*: $p < 0.05$;**: $p < 0.01$; ***: $p < 0.001$).

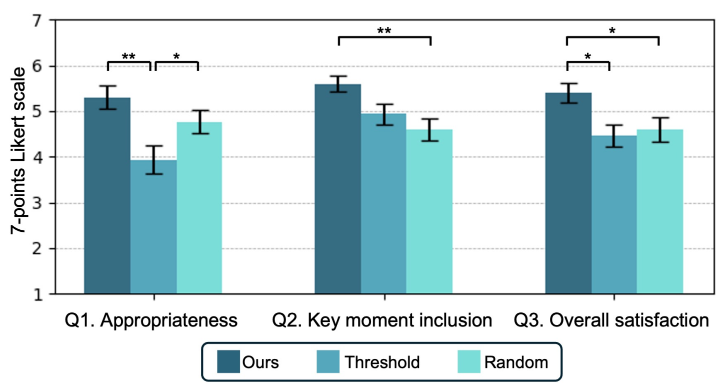

Fig. B13. 7-point Likert scale results of three methods for appropriateness (Q1), key moment inclusion (Q2), and overall satisfaction (Q3) of entertainment video [40]. Error bars represent the standard error of the mean. Asterisks show statistically significant differences between pairs of methods (*: $p < 0.05$;**: $p < 0.01$; ***: $p < 0.001$).

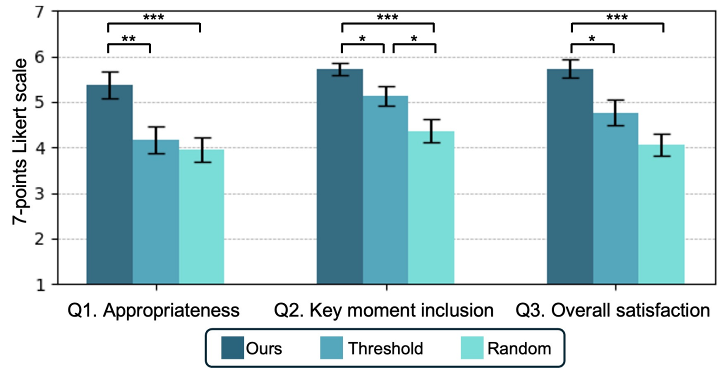

Fig. B14. 7-point Likert scale results of three methods for appropriateness (Q1), key moment inclusion (Q2), and overall satisfaction (Q3) of tennis video [41]. Error bars represent the standard error of the mean. Asterisks show statistically significant differences between pairs of methods (*: $p < 0.05$;**: $p < 0.01$; ***: $p < 0.001$).

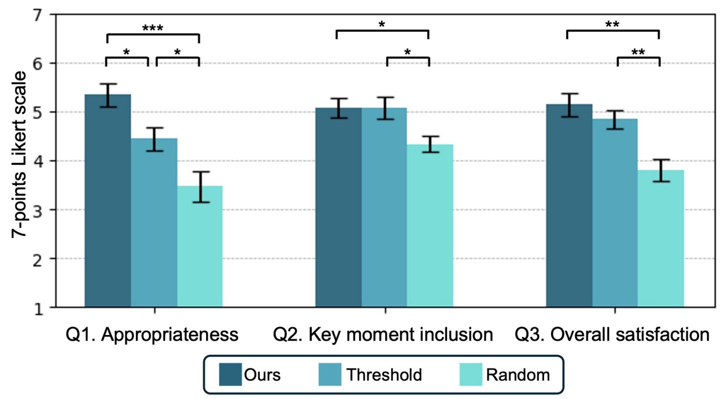

Fig. B15. 7-point Likert scale results of three methods for appropriateness (Q1), key moment inclusion (Q2), and overall satisfaction (Q3) of documentary video [19]. Error bars represent the standard error of the mean. Asterisks show statistically significant differences between pairs of methods (*: $p < 0.05$;**: $p < 0.01$; ***: $p < 0.001$).

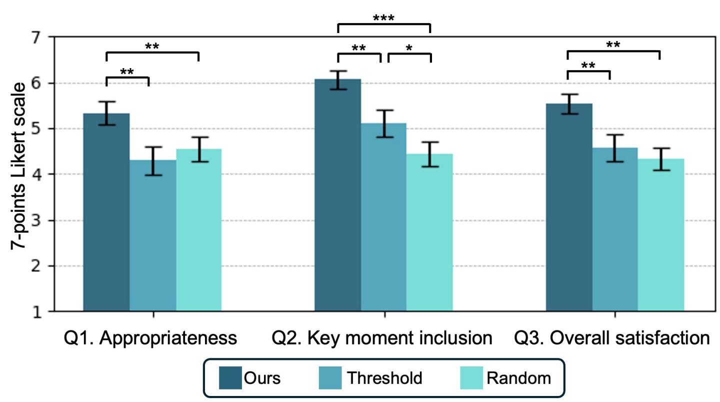

Fig. B16. 7-point Likert scale results of three methods for appropriateness (Q1), key moment inclusion (Q2), and overall satisfaction (Q3) of soccer video [31]. Error bars represent the standard error of the mean. Asterisks show statistically significant differences between pairs of methods (*: $p < 0.05$;**: $p < 0.01$; ***: $p < 0.001$).

C VIDEO SOURCE USED IN USER STUDY 2

Table C2 shows the original *YouTube* video sources and their corresponding human-edited highlights (*YouTube Shorts*) sources used in User Study 2 described in Section 6.2 of the main paper. The video index in the leftmost column corresponds to the data used in Figure 10 of the main paper.

Table C2. Video sources used in User Study 2.

Video	**Original video**	**Human-edited highlight**
A	[28]	[1]
B	[8]	[9]
C	[17]	[11]
D	[26]	[27]
E	[6]	[7]
F	[35]	[33]
G	[13]	[14]
H	[39]	[30]
I	[12]	[3]
J	[5]	[4]

D VISUAL EXAMPLES OF DEEP LEARNING-BASED MRD PREDICTION

We implemented a deep learning-based MRD prediction model [34] to check the feasibility of predicted MRD. An existing model [42] is trained to predict the importance scores for videos, using the MRD data (datasets of video and ground-truth MRD pairs are presented by Sul et al. [34]). Figures D17, D18, and D19 illustrate the ground truth MRD, the predicted MRD, and the smoothed predicted MRD. The data used in these figures are from the inference dataset in Sul et al. [34]. The smoothing was performed on the predicted MRD using median filtering with a kernel size of 9 after normalizing from 0 to 1. We calculated the cosine similarity between the ground truth and the smoothed MRD, obtaining high similarity scores of 0.91, 0.88, and 0.92, respectively. These indicate that the predicted MRD graphs closely resemble the ground truth. Additionally, we included a 30-second highlight created using our method based on the predicted MRD in the supplementary video[1].

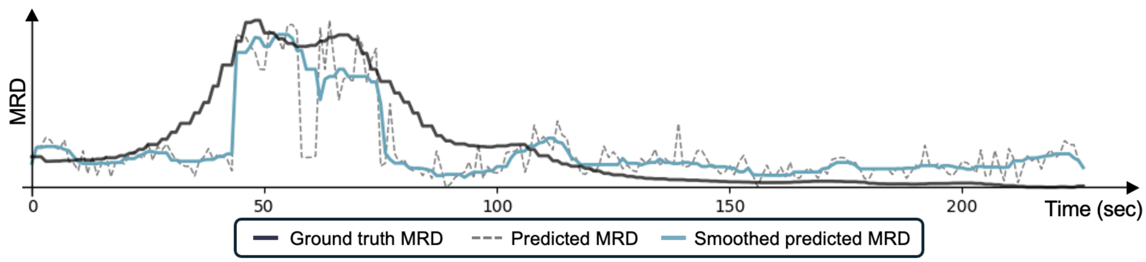

Fig. D17. The graphs of ground truth MRD, predicted MRD, and smoothed predicted MRD of [2].

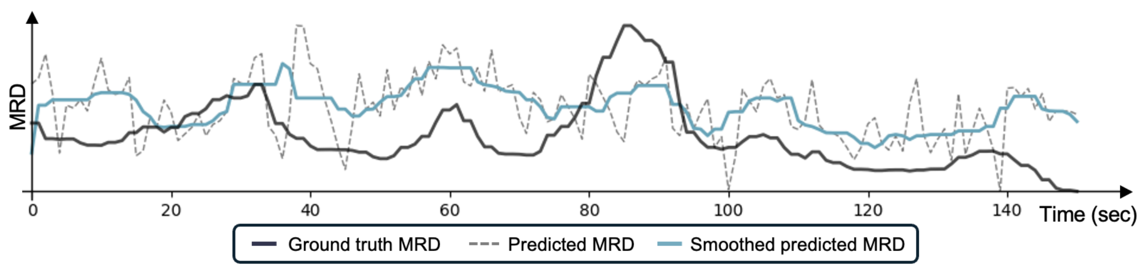

Fig. D18. The graphs of ground truth MRD, predicted MRD, and smoothed predicted MRD of [15].

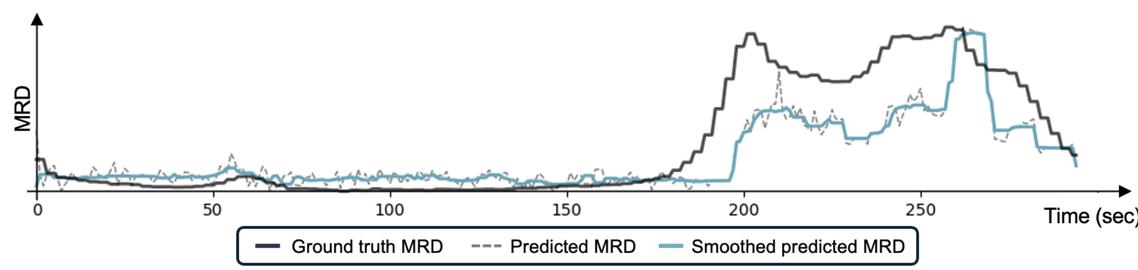

Fig. D19. The graphs of ground truth MRD, predicted MRD, and smoothed predicted MRD of [10].

[1] https://w-dlee.github.io/highlights

REFERENCES

[1] AGT. 2022. *Mayyas STUNS the judges.* https://www.youtube.com/shorts/spT0uxRiYm4 Accessed: 2024-09-13.

[2] Michelle Almendra. 2014. *MAIZENA E AÇÚCAR- Deixe seu cabelo LISO e SUPER HIDRATADO- (passo a passo).* https://www.youtube.com/watch?v=52B0NrF5ags Accessed: 2024-11-22.

[3] BogopangTV. 2023. *Jang Mi-ran Lifts the World | Iconic Sports Moment.* https://www.youtube.com/shorts/ssaqjp9VNI0 Accessed: 2024-09-13.

[4] The CHOOM. 2023. *As Soon as Hyojin Choi Appears, You Can't Help but Dance!* https://www.youtube.com/shorts/9sVRxaoDywg Accessed: 2024-09-13.

[5] The CHOOM. 2023. *[EN/JP] [SWF2/Episode 5] Deep N Dap Showcased the True Essence of Korean Style!* https://www.youtube.com/watch?v=bdxUg7Jxhh8 Accessed: 2024-09-13.

[6] The CHOOM. 2023. *[SWF2/Episode 5] Literally One Million with 100 Dancers The Jaw-Dropping 1M Mega Crew Behind-the-Scenes.* https://www.youtube.com/watch?v=cY04XjQEJyE Accessed: 2024-09-13.

[7] The CHOOM. 2023. *Warning: Too Cool! This Scene Represented the Korean Peninsula?!* https://www.youtube.com/shorts/MXTgQFvcMg4 Accessed: 2024-09-13.

[8] EBS Documentary. 2024. *A Starving Wild Wolf Enters the Human Territory I Guard... A Life-or-Death Battle Between the Wolf and the Dog | The Ghost-Seeing Dog | Mongolian Bankhar, Protecting Livestock All Night Without Sleep | DocuPrime.* https://www.youtube.com/watch?v=3SK5DFSoLlY Accessed: 2024-09-13.

[9] EBS Documentary. 2024. *Wolf vs Bankhar.* https://www.youtube.com/shorts/zkK9JY9CA94 Accessed: 2024-09-13.

[10] Contextual Electronics. 2013. *Introduction and KiCad Project Creation - "Getting To Blinky" KiCad Tutorial - Part 1.* https://www.youtube.com/watch?v=iTyi3RvNoB0 Accessed: 2024-11-22.

[11] Play Epik. 2022. *Neo vs Merovingian The Matrix Reloaded.* https://www.youtube.com/shorts/u7st1Ur-eTQ Accessed: 2024-09-13.

[12] MBC Sports Expedition. 2020. *All-Time Legend | The Olympics on MBC | Jang Mi-ran's Women's +75kg Weightlifting at the 2008 Beijing Olympics.* https://www.youtube.com/watch?v=5Xc5uTdSLV8 Accessed: 2024-09-13.

[13] SBS TV Animal Farm. 2020. *Alpaca Happens To Live In A City With Two Guys.* https://www.youtube.com/watch?v=NFJtBsB1pWI Accessed: 2024-09-13.

[14] SBS TV Animal Farm. 2022. *How an Alpaca Shows It's Eaten.* https://www.youtube.com/shorts/Vz7vBvnvRQ4 Accessed: 2024-09-13.

[15] FIFA. 2014. *Nigeria 0:3 Spain | FIFA Confederations Cup 2013 | Match Highlights.* https://www.youtube.com/watch?v=dLAJLppl5ts Accessed: 2024-11-22.

[16] FIFA. 2022. *FULL MATCH: France vs. Croatia | 2018 FIFA World Cup Final.* https://www.youtube.com/watch?v=SvV6aUki6LU Accessed: 2024-09-13.

[17] Flashback FM. 2020. *Neo vs Merovingian | The Matrix Reloaded [IMAX].* https://www.youtube.com/watch?v=uXGE0vuuaDo Accessed: 2024-09-13.

[18] National Geographic. 2019. *The Harp Seal's Race Against Time - Ep. 5 | Wildlife: The Big Freeze.* https://www.youtube.com/watch?v=BF2TZq-ntRQ Accessed: 2024-11-18.

[19] National Geographic. 2020. *The Sacrifice of Cassini | Cosmos: Possible Worlds.* https://www.youtube.com/watch?v=Je7FAXXNLAY Accessed: 2024-09-11.

[20] National Geographic. 2024. *Behind The First Summit of The World's Tallest Mountain (Full Episode) | Lost on Everest.* https://www.youtube.com/watch?v=gCr4Pg9mEA4 Accessed: 2024-11-18.

[21] NASA. 2023. *The Ring of Fire: 2023 Annular Solar Eclipse (Official NASA Broadcast).* https://www.youtube.com/watch?v=LlY79zjud-Q Accessed: 2024-11-18.

[22] NBA. 2023. *Warriors Lakers Instant Classic - 2021 Play-In Tournament| NBA Classic Game.* https://www.youtube.com/watch?v=LPDnemFoqVk Accessed: 2024-09-13.

[23] BBC News. 2024. *How next Trump administration may tackle economy, immigration and health | BBC News.* https://www.youtube.com/watch?v=VB5ZRVky0Yw Accessed: 2024-11-18.

[24] NBC News. 2021. *Watch: Jeff Bezos, Blue Origin Crew Launch Into Space | NBC News.* https://www.youtube.com/watch?v=4SSVr3EKTlM Accessed: 2024-11-18.

[25] Olymics. 2024. *RE-LIVE | Ice Hockey Men's Final.* https://www.youtube.com/watch?v=0PbKJJlUmPw Accessed: 2024-09-16.

[26] Olympics. 2016. *Yuna Kim - Short Program - Ladies' Figure Skating | Vancouver 2010.* https://www.youtube.com/watch?v=ecKnyyuMsmw Accessed: 2024-09-13.

[27] Olympics. 2022. *The Bond Girl who Thrilled Everyone with Just the Intro.* https://www.youtube.com/shorts/2HZBcC2BP00 Accessed: 2024-09-13.

[28] Talent Recap. 2022. *GOLDEN BUZZER! The Mayyas Prove What Arab Women Can Do With Inspirational Dance Act!* https://www.youtube.com/watch?v=UBA3OQSJU58 Accessed: 2024-09-14.

[29] N Reimers. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv preprint arXiv:1908.10084* (2019).

[30] Skydoo. 2022. *Spider Man Saves Train Scene.* https://www.youtube.com/shorts/j8gmjr4_sWU Accessed: 2024-09-13.

[31] Coupang Play Sports. 2024. *[2023 AFC Qatar Asian Cup] 2023 AFC Qatar Asian Cup Qatar VS Uzbekistan Full Highlights.* https://www.youtube.com/watch?v=lNpN_FH8_Ac Accessed: 2024-09-11.

[32] SBS Sports. 2019. *2016 Rio Olympics Women's Golf Final Round | Park In-bee.* https://www.youtube.com/watch?v=3GgxdPhm5cM Accessed: 2024-09-11.

[33] SSAULABI. 2024. *The Incredible Kicking Skills of Team Saurabi, Who Received the Golden Buzzer on BRITAIN'S GOT TALENT.* https://www.youtube.com/shorts/FH4YboYz_hE Accessed: 2024-09-13.

[34] Jinhwan Sul, Jihoon Han, and Joonseok Lee. 2024. Mr. HiSum: A Large-scale Dataset for Video Highlight Detection and Summarization. *Advances in Neural Information Processing Systems* 36 (2024).

[35] Britain's Got Talent. 2024. *Ssaulabi AMAZE us all with Golden Buzzer-winning performance | Auditions | BGT 2024.* https://www.youtube.com/watch?v=DzcH9eNFVto Accessed: 2024-09-13.

[36] TED. 2013. *The astounding athletic power of quadcopters | Raffaello D'Andrea.* https://www.youtube.com/watch?v=w2itwFJCgFQ Accessed: 2024-11-18.

[37] New York University. 2022. *NYU's 2022 Commencement Speaker Taylor Swift.* https://www.youtube.com/watch?v=OBG50aoUwlI Accessed: 2024-11-18.

[38] Veritasium. 2021. *The Simplest Math Problem No One Can Solve - Collatz Conjecture.* https://www.youtube.com/watch?v=094y1Z2wpJg Accessed: 2024-11-18.

[39] Voyage. 2020. *Spider-Man Stops A Train From Crashing | Spider-Man 2 | Voyage | With Captions.* https://www.youtube.com/watch?v=r_W6mXqzJNU Accessed: 2024-09-13.

[40] JTBC Voyage. 2020. *[Must Watch] "Ah.. Really!" National Treasure-level Center Bang-gu Seo Jang-hoon's Real Basketball Moments.* https://www.youtube.com/watch?v=8LwZsvWFmrY Accessed: 2024-09-11.

[41] Wimbledon. 2019. *Roger Federer vs Rafael Nadal Wimbledon 2019 semi-final highlights.* https://www.youtube.com/watch?v=J8__TwOgTY0 Accessed: 2024-09-11.

[42] Minghao Xu, Hang Wang, Bingbing Ni, Riheng Zhu, Zhenbang Sun, and Changhu Wang. 2021. Cross-category video highlight detection via set-based learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7970–7979.