

NOISYVIDEO: A COMPREHENSIVE BENCHMARK FOR ROBUST VIDEO UNDERSTANDING UNDER VISUAL NOISE

Anonymous authors

Paper under double-blind review

ABSTRACT

The recent progress of Video Multimodal Large Language Models (Video MLLMs) has greatly advanced multimodal understanding and reasoning in video analysis. However, the robustness of these models under diverse noise conditions that commonly occur in real-world scenarios remains largely unexplored, and existing research lacks systematic evaluations to comprehensively assess Video MLLM performance on question answering tasks under various noise conditions, both of which limit their reliability in practical deployments. To bridge this gap, we propose a thorough robustness benchmark encompassing 36 noise types in 8 categories, spanning diverse video categories and question types, resulting in 21,924 noise-corrupted test videos for comprehensive evaluation of Video MLLMs’ robustness. We evaluate 10 state-of-the-art MLLMs on this novel benchmark for the initial systematic evaluation from multiple perspectives. Our multi-faceted analysis uncovers existing bottlenecks and performance degradation under certain noise scenarios, particularly for tasks requiring fine-grained understanding and reasoning. Additionally, we also examine the effectiveness of current image restoration techniques in mitigating noise effects and discuss their limitations. By constructing this extensive benchmark, our work lays a foundation for the systematic evaluations of Video MLLMs, offering insightful findings for future research on robust video-language understanding.

1 INTRODUCTION

Video understanding Cho et al. (2025); Shaker et al. (2025); Tang et al. (2023) is a core capability for modern intelligent systems, underpinning a wide range of downstream tasks such as action recognition Abdelkawy et al. (2025), optical character recognition Nagaonkar et al. (2025), and scene graph generation Kim et al. (2025). Recent advances in large language models (LLMs) Touvron et al. (2023); Achiam et al. (2023) have fueled a paradigm shift in video understanding, where researchers have achieved remarkable progress by introducing visual mapping layers Liu et al. (2023) that project visual signals into the LLM processing space. This approach has led to the rapid development of Video Multimodal Large Language Models (Video MLLMs) Maaz et al. (2023); Lin et al. (2023); Li et al. (2024c), resulting in a diversity of available models for many tasks. To evaluate the performance of these models on video understanding, some benchmarks have been introduced, focusing more on some aspects, such as longer video contexts Tan et al. (2025); Zhou et al. (2024); Fang et al. (2024) and specific application domains Xu et al. (2024b); Gao et al. (2025); Xie et al. (2025).

Despite the significant progress brought by these studies, a pervasive challenge in practical video analysis (i.e., the presence of noise corruption) may be overlooked and remain largely underexplored. Noise is known to impact the performance of LLMs remarkably Chen et al. (2023). Thus, the above motivations lead to our study: *Do Video MLLMs remain robust when faced with various types of noise corruption that commonly occur in practical deployment scenarios?*

In practice, visual noise can be occasionally introduced by imperfect filming techniques, hardware limitations, or transmission errors Boncelet (2009), and may be unnoticed in real-world situations Wang & Zhu (1998). Consequently, visual noise remains hidden in videos, potentially undermining the performance of Video MLLMs Schiappa et al. (2022). Previous works also show that MLLMs are

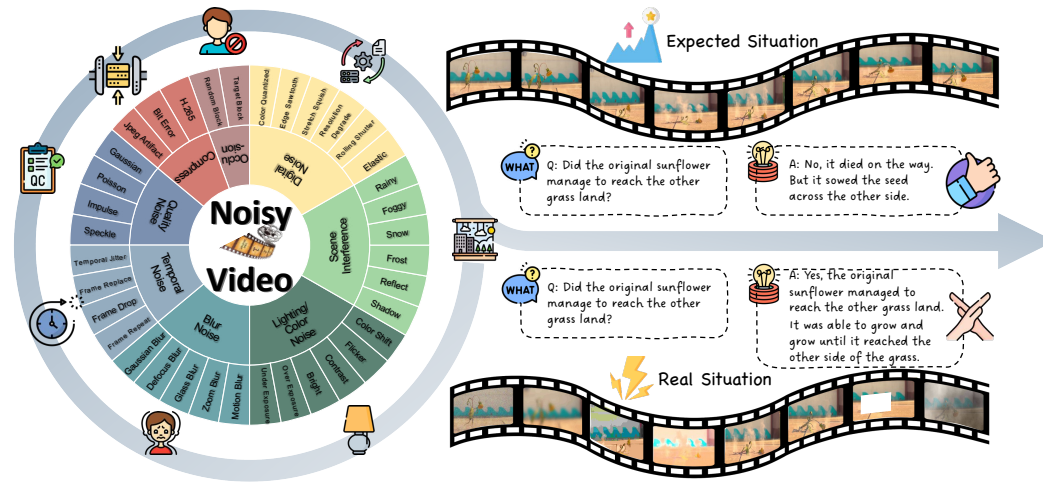


Figure 1: Overview of the eight major categories of noise included in our benchmark, with detailed noise types under each category. When applying Video MLLMs for video understanding, we expect to obtain reliable answers. However, in practice, due to the impact of visual noise, Video MLLMs may produce incorrect answers.

Table 1: Comparisons between our *NoisyVideo* benchmark and peer benchmarks about visual noise. Icons represent different noise types: Quality, Temporal, Blur, Lighting/Color, Scene Interference, Digital, Occlusion, and Compression.

Model									Total	Year	Type	Task/Model
Hendrycks et al. Hendrycks & Dietterich (2019)	3	—	4	2	3	2	—	1	15	2019	Image	Image Classification
Schiappa et al. Schiappa et al. (2022)	4	5	3	—	—	—	—	3	18	2023	Video	Text-Language Model
Sun et al. Sun et al. (2023)	3	—	4	2	3	2	—	1	15	2023	Image	Video-Language Model
Chen et al. Chen et al. (2023)	4	1	5	2	3	4	—	1	20	2023	Image	Image-Language Model
Shirmin et al. Shirmin et al. (2024)	—	—	1	—	2	2	—	—	5	2024	Image	Vision-Language Model
Zeng et al. Zeng et al. (2024)	—	2	1	1	—	—	1	—	5	2024	Video	Temporal Action Detection
Xie et al. Xie et al. (2025)	—	2	2	3	5	1	—	2	15	2025	Video	Autonomous Driving
NoisyVideo (Ours)	4	4	5	6	6	6	2	3	36	—	Video	Video Understanding

more sensitive to visual perturbations than textual perturbations Qiu et al. (2022). This implies that analyzing the robustness of Video MLLMs on visual noise is a crucial and impactful research direction. However, current benchmarks related to visual noise either do not support comprehensive video understanding tasks Hendrycks & Dietterich (2019); Sun et al. (2023), or cover a narrow set of visual noise types Schiappa et al. (2022); Zeng et al. (2024), as also shown in Table 1. We aim to explore further in this direction.

To fill this gap, we propose a comprehensive benchmark that systematically evaluates the robustness of Video MLLMs under a wide range of visually noisy conditions for video understanding tasks. Our benchmark offers several distinctive features. First, we construct an extensive and fine-grained noise set with 36 corruption types across 8 major categories (see Figure 1), covering degradations such as quality, blur, temporal disruptions, lighting and color variation, scene interference (e.g., rain or fog), digital noise, occlusion, and compression. This diversity enables rigorous and realistic testing. Second, we employ multiple evaluation metrics, including subjective GPT-based scores, objective SBERT-based semantic similarity for all question types, and conventional accuracy on True/False questions, ensuring a holistic assessment of model robustness. Third, we conduct fine-grained analysis across different model architectures, video genres, and question types, revealing nuanced patterns of vulnerability and strength, such as the sensitivity of perception-based versus reasoning-based questions, the unique challenges of specific video categories, and the influence of model structure on robustness to visual noise. Furthermore, we examine whether recent image restoration techniques can effectively recover the Video MLLMs’ understanding ability with visual noise.

Our analysis shows that pixel-level structural corruption substantially impairs the robustness of most models. Additionally, we observe that certain models are highly sensitive to specific video instances, noise types, or question categories, resulting in significant performance drops. While

denoising methods can partially restore performance on simple question answering tasks, they remain inadequate for recovering the fine-grained understanding required for complex reasoning. These findings highlight the significant challenges in achieving robust video understanding. Our work provides a practical and systematic benchmark for evaluating Video MLLMs and offers valuable insights to guide the development of more robust models.

Our contributions are summarized as follows:

- **A comprehensive visual noise robustness benchmark for Video MLLMs:** We propose *the first* systematic benchmark to evaluate the robustness of Video MLLMs, encompassing 8 broad categories and 36 noise types that represent various forms of video corruption relevant to practical deployment scenarios.
- **Multi-dimensional evaluation of robustness:** We consolidate holistic measurements that integrate subjective, objective, and accuracy-based metrics, showing a thorough quantitative evaluation of Video MLLMs’ performance under various noise conditions.
- **Fine-grained analysis across multiple perspectives:** We conduct detailed analysis with model architecture, video genre, and question type, revealing nuanced patterns among different Video MLLMs and providing insightful findings for future research.
- **Fully open-source benchmark and tools:** We make our benchmark and its tools for construction fully open-source to boost the research in video understanding. <https://anonymous.4open.science/r/NoisyVideo-1AD5>

2 NOISYVIDEO

In this section, we provide the detailed construction of our benchmark. Specifically, we elaborate on the criteria of the visual noise types and evaluation metrics in our benchmark. To show its effectiveness, we conduct the evaluation of representative Video MLLMs on this benchmark.

2.1 VIDEO DEGRADATION METHODOLOGY

We introduce a wide range of noise types that are representative of common video degradation patterns to evaluate the robustness of Video MLLMs. While evaluating on naturally occurring noisy videos would be ideal, most real-world noise conditions present significant challenges for systematic analysis: they typically contain complex mixtures of multiple degradation sources that cannot be easily disentangled, and lack corresponding clean reference videos for controlled comparison. Therefore, following established practices in robustness evaluation Schiappa et al. (2022); Xie et al. (2025), we adopt a hybrid approach that combines authentic reproduction methods with well-validated synthetic generation techniques. For degradations that can be faithfully reproduced using standard processing pipelines (such as compression artifacts), we employ real-world encoding workflows to generate authentic corruptions. For other noise types, we implement established synthetic generation methods documented in the literature to ensure controlled experimental conditions that enable systematic analysis.

To ensure comprehensive coverage of potential degradation scenarios, we harvest noise types from different sources in both image and video domains. From the literature Perry (2018); Ramesh et al. (2022); Pyatykh et al. (2012); Lim (2006); Boncelet (2009); Foi et al. (2008); Zlokolica et al. (2006); Amer & Dubois (2005); Amrani et al. (2021); Wang et al. (2019a); Zhang et al. (2021); Guo et al. (2024); Hendrycks & Dietterich (2019), we identify *36 types of different noise* due to capturing, processing, and saving. Each noise reflects a realistic degradation that may influence the performance of Video MLLMs in practical application scenarios.

For degradations that can be faithfully reproduced, we employ industry-grade encoders to generate the distorted videos, ensuring these corruptions closely match their real-world counterparts. For instance, to simulate H.265 compression artifacts, we utilize the same FFmpeg workflows used in industry practice, producing artifacts that are representative of real-world compression noise.

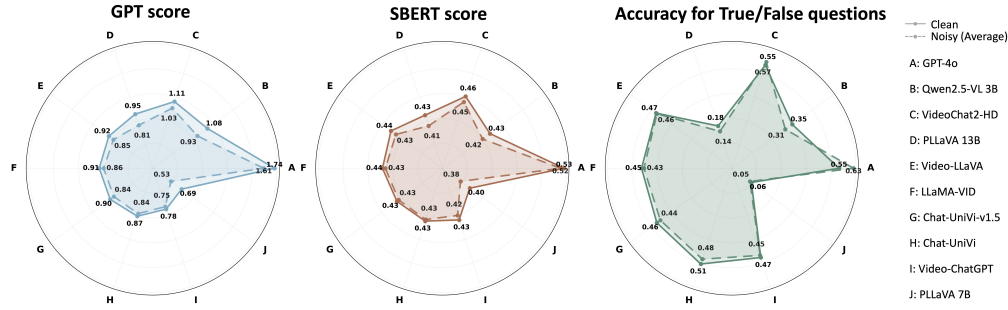


Figure 2: A preliminary study on different metrics in a controlled experiment.

2.2 NOISE TYPE

To better analyze the impacts of different noises, we further categorize them into 8 *groups* by their characteristics. These categories are well recognized in the existing studies Hendrycks & Dietterich (2019); Zeng et al. (2024); Wang et al. (2019b). Namely, they are distinct noises related to quality, temporality, blurring, lighting/color, scene interference, digitality, occlusion, and compression. Figure 1 depicts the relationships in this taxonomy.

We notice that some noise types have corruptions in the pixel continuity. More specifically, **Quality noise** (e.g., Gaussian or impulse noise) emulates sensor imperfections or external interference, which often occurs in low-light environments or with low-quality sensors. **Blur noise** like motion blur and defocus blur replicates scenarios where camera movement or improper focus impact video clarity, such as in handheld or dynamic recordings. **Temporal noise**, including frame drops, frame repeats, and jitter, represents irregularities in video playback or transmission. These are common in videos transmitted over unstable networks or in hardware-constrained systems. Alternatively, **compression** (e.g., H.265 compression and JPEG artifacts) simulates distortions caused by lossy encoding processes during video upload, download, or storage, which are frequently encountered on streaming platforms or in bandwidth-limited scenarios. **Occlusion** is another important type of noise, which simulates situations where the camera lens is partially blocked out by stains, or when the shooting angle causes the target object to be partly covered. **Digital distortions** in videos cover a wide range of artifacts. Some are related to geometric deformations, such as rolling shutter and elastic distortions, which alter the shape and structure of objects within the frame. Others involve changes in resolution, such as resolution degradation, which reduces visual clarity and detail. Additionally, certain distortions affect color properties, like color quantization. Although the abovementioned noises originate from different sources, it should be noted that from the perspective of their visual manifestations, most of them ultimately result in the destruction of continuity information within the frames.

The other types of noise do not damage the pixel integrity of objects in the frames. **Lighting/color noise**, such as flicker, overexposure, or underexposure, primarily simulates challenging illumination conditions rather than altering the underlying structure of objects. These lighting-related noises are commonly used in data augmentation to improve the robustness of models against varying lighting environments. **Scene interference**, such as rain, fog, and shadows, is designed to mimic environmental conditions that are frequently encountered during outdoor video capture, for instance, in autonomous driving or surveillance scenarios. These effects introduce additional visual complexity but do not necessarily disrupt the object’s structure.

The gathered noise types are carefully constructed with our best efforts to cover diverse and realistic scenarios, ensuring a comprehensive evaluation of Video MLLMs’ robustness. By thorough testing under these noise conditions, we aim to examine the strengths and limitations of Video MLLMs. Detailed descriptions and parameter settings for each noise type are provided in Appendix A.2.

2.3 EVALUATION METRICS

We find that existing studies generally rely on an LLM judge to evaluate the performance of Video MLLMs Fang et al. (2024); Maaz et al. (2023); Xu et al. (2024b). We have conducted a preliminary study with three measurements on video understanding with all visual noise as shown in Figure 2.

It is worth noticing that different models exhibit varying changes in performance across different metrics (i.e., columns A, B, and C), implying that relying on limited metrics could be biased. Therefore, we tend to extend the benchmark with more measurements to improve the fairness and comprehensiveness.

We follow the existing study Fang et al. (2024) to include a traditional measurement in our benchmark. Specifically, the metric is scored by conventional LLMs, also known as *GPT score*. However, since this scoring is subjective to single models, we aim to include alternative measurements. From the literature, we identify the *SBERT score* Reimers & Gurevych (2019), which is objective and commonly used in many LLM-related works Supriya et al. (2023); Zhong et al. (2020); Moon et al. (2020). We include this metric in our benchmark.

To further enhance the comprehensiveness of the measurements, we consider the statistical measurements, besides the predictive ones (i.e., GPT score) and the semantic ones (i.e., SBERT score). More specifically, the *accuracy* is measured on the selection-based questions (i.e., True or False questions in our benchmark) with concrete correct answers. It provides a new angle to assess the robustness of Video MLLMs under noise conditions.

In summary, our evaluation framework incorporates:

- ✓ **GPT Score:** Subjective assessment following established practices in Video MLLM evaluation, where GPT-4o OpenAI (2024) evaluates response quality based on relevance, correctness, and completeness.
- ✓ **SBERT Score:** Objective semantic similarity measurement that computes cosine similarity between model responses and ground truth answers using sentence embeddings, providing reproducible and bias-free evaluation.
- ✓ **Accuracy:** Concrete performance measurement on selection-based questions (e.g., True/False, multiple-choice) with definitive correct answers, offering clear and interpretable results for fundamental understanding assessment.

The NoisyVideo benchmark measures the abovementioned three metrics individually on both settings with and without noise presence. The difference between the two measurements reveals the intrinsic robustness of Video MLLMs, eliminating the other factors in a controlled experiment manner.

2.4 EVALUATION SUBJECTS

To demonstrate the effectiveness of our benchmark and answer the research question in our study, we conduct a comprehensive evaluation on various subjects.

We adopted 609 videos exhaustively from MMBench-video Fang et al. (2024), which consists of 19 video types and 9 question types to cover diverse practical application scenarios. All 36 types of noise are applied to these videos individually, producing an exhaustive set of 21,924 videos as our evaluation subjects.

For the Video MLLM subjects, since a full evaluation of all constructed video tasks would require substantial computational resources, we focus our assessment on representative models. We select models from two perspectives: *general MLLMs* and *specialized Video MLLMs*.

General MLLMs, such as GPT-4o OpenAI (2024) and Qwen2.5-VL 3B Yang et al. (2024), are chosen because they are widely adopted as base models for a variety of tasks, including video understanding Li et al. (2024a); Chen et al. (2025). Although these MLLMs are not specifically designed for video, their broad applicability makes them valuable references. For specialized Video MLLMs, we specifically include two spatio-temporal enhanced models (i.e., Chat-UniVi Jin et al. (2024) and Video-ChatGPT Maaz et al. (2023)), both of which focus on effective integration of temporal dynamics and spatial features. Four more state-of-the-art Video MLLMs (i.e., PLLaVA Xu et al. (2024a), VideoChat2 Li et al. (2023), Video-LLaVA Lin et al. (2023), and LLaMA-VID Li et al. (2024c)) are selected due to their popularity and recognition in related works. We also consider the impact of model size on the video understanding capabilities. The aforementioned specialized Video MLLMs are 7B-parameter models, which are widely adopted as the baselines in the MLLMs Li et al. (2024a); Liu et al. (2023). Moreover, we include the 13B variant of PLLaVA and the enhanced data

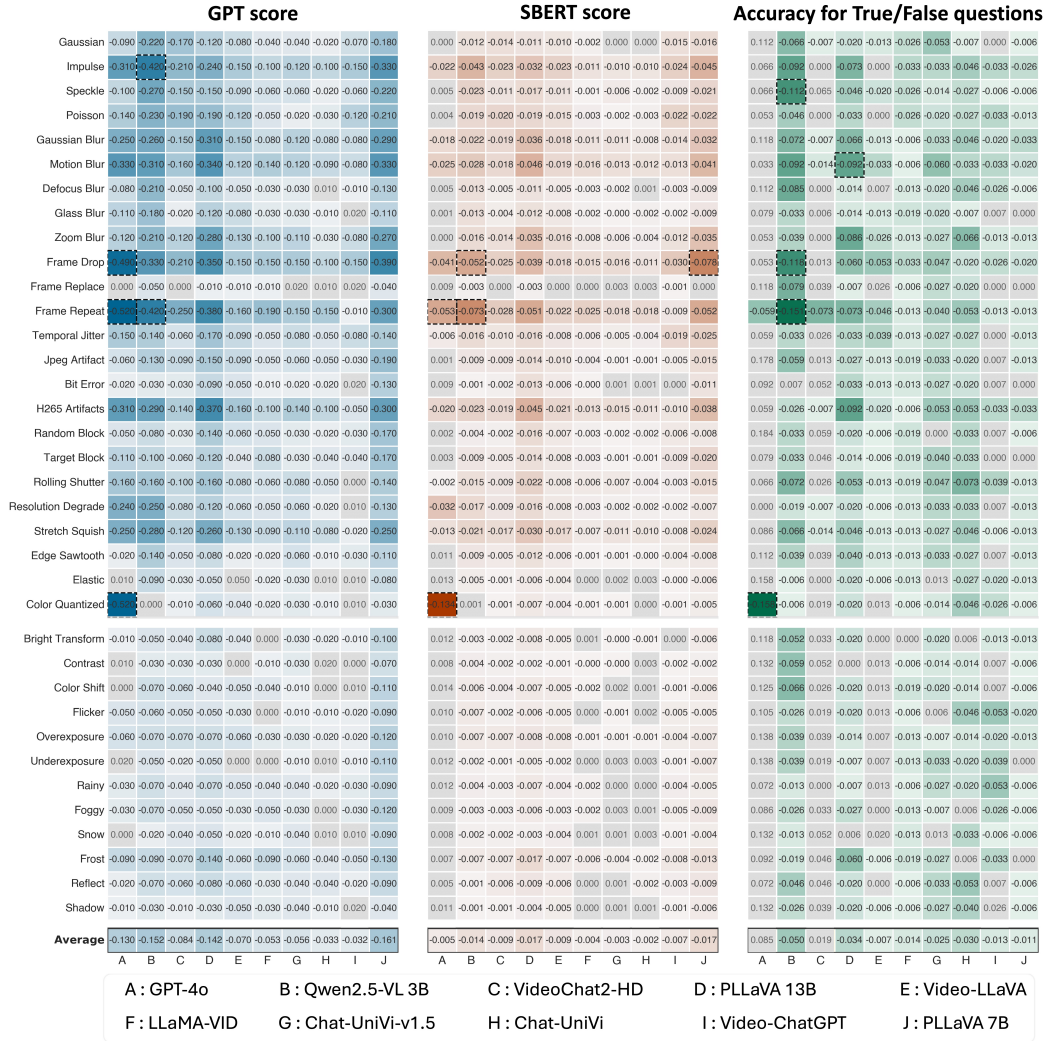


Figure 3: Heatmap of performance drops for general MLLMs and the specialized Video MLLMs under different noise conditions.

version of Chat-UniVi-v1.5 to enrich the evaluation, thanks to their public availability. The public links can be seen in Appendix A.3.

This setup enables us to analyze the variance across video scenarios, model architectures, model size, and training data, contributing to the video understanding robustness.

3 EXPERIMENTS

This section presents a comprehensive evaluation of Video MLLMs’ robustness under various noise conditions using our NoisyVideo benchmark. We systematically analyze the performance degradation patterns across different dimensions to understand the vulnerability and resilience characteristics of current Video MLLMs. Additionally, we investigate whether existing video denoising methods can restore noisy videos to achieve comparable video understanding performance as clean videos, providing valuable insights for practical deployment scenarios. Due to space constraints, more experimental results are presented in the Appendix, with all visual results (figures and charts) relocated there to maintain focus on the core findings and analysis in the main text.

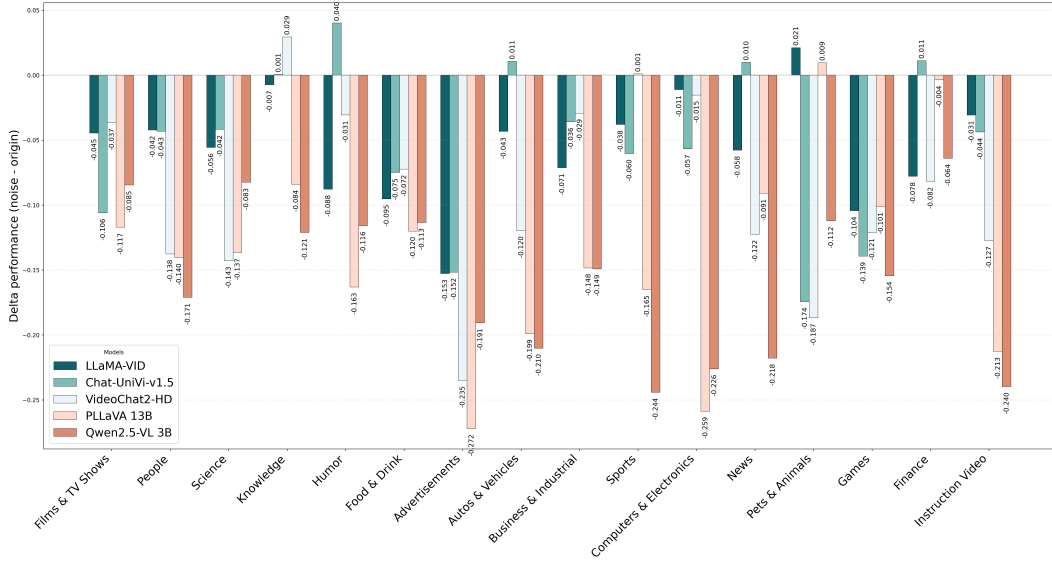


Figure 4: Delta performance (GPT Score) of general and Video MLLMs across different video categories. More detailed SBERT and accuracy results are provided in the Appendix A.7.

3.1 EVALUATION OF VIDEO MLLMs UNDER NOISE CONDITIONS

We evaluate the impacts of different visual noise types on video understanding tasks and assess the robustness of Video MLLMs under these noise conditions. Specifically, given the original performance of subject Video MLLMs on quality-assured videos (i.e., without noise in the videos), we examine the performance of the same subject models on the noisy version of the same videos. Then, we perform the comparison study on the two acquired performances to evaluate the impacts of noise. To ease our presentation, we refer to the quality-assured videos as *clean* videos and the noise-amended videos as *noisy* videos. We also perform a comparative analysis on different groups to investigate the potential factors related to robustness. We summarize the key results in Figure 3.

3.1.1 IMPACTS OF NOISE TYPES ON ROBUSTNESS OF VIDEO MLLMs

Figure 3 presents the delta performance of 10 models across 36 noise types. The columns of models are sorted by their original performance on clean videos in descending order (The general MLLMs and the specialized Video MLLMs are ranked separately). The rows of noise types are grouped by their influence on pixel continuity, as we show in Section 2.2. The last row summarizes the average impacts of different noises on the performance.

As illustrated in Figure 3, it is evident from all three metrics that noise generally exerts a detrimental effect on the performance of Video MLLMs. Out of 1,080 cases, 163, 44, and 873 cases show increases, retention, and drops in performance, respectively. The performance drops occupy 80.83% of cases in the experiments. Going deeper, general MLLMs (i.e., GPT-4o and Qwen2.5-VL 3B) suffer heavier performance drops than specialized Video MLLMs. These two general MLLMs receive better performance on clean videos but perform relatively worse on noisy videos on GPT score, which implies that *high video understanding performance does not ensure robustness*. Users may choose specialized Video MLLMs in the application scenarios with noise.

Another finding from this large-scale experiment is that *noises corrupting the pixel continuity generally have a greater negative impact on robustness*. For the two groups, the performance drops in the corruption group range from -0.20 to -0.04 for GPT Score (-0.12 ± 0.07), and from $-2.25e-2$ to $-4.10e-3$ for SBERT Score ($-1.23e-2 \pm 6.70e-3$). In contrast, in the other group, GPT Score drops range from -0.10 to -0.01 (-0.04 ± 0.03), and SBERT Score drops from $-6.80e-3$ to $9.60e-3$ ($-1.80e-3 \pm 4.80e-3$), which is significantly slighter than the former. Furthermore, as shown in the accuracy statistics, the mean delta performance of accuracy for the models under corruption group is -0.01 ± 0.04 . However, for the other group, the performance increases to $2.9e-3 \pm 0.04$. This indicates that, for these simple question-answering tasks, the impact of noise

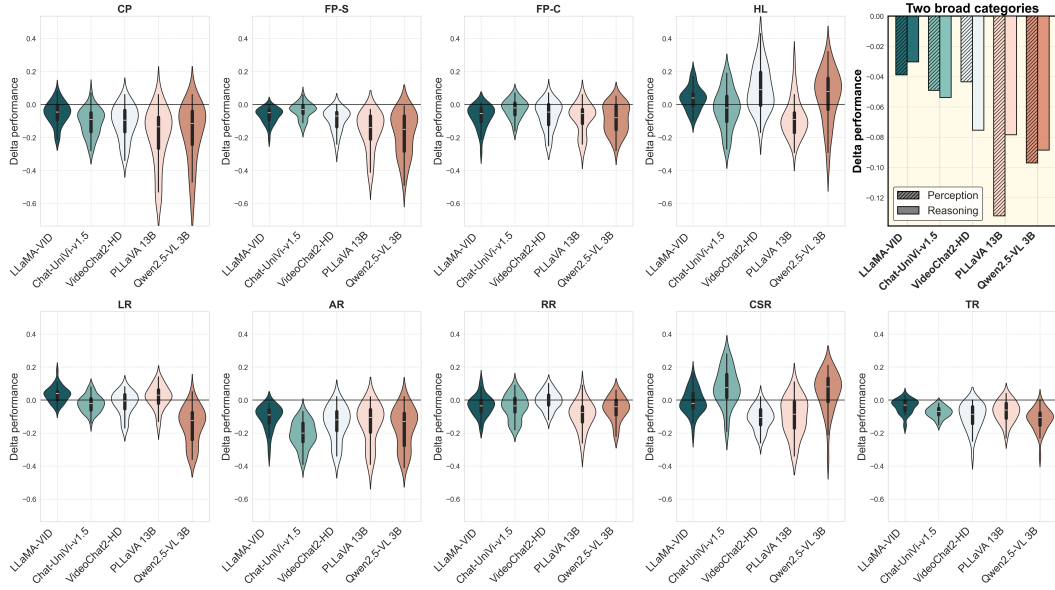


Figure 5: Delta performance (GPT Score) of general and Video MLLMs across different question categories. More detailed SBERT and accuracy results are provided in the Appendix A.7.

types related to color is minimal, and model performance remains largely unaffected. Overall, these results highlight the critical role of pixel continuity in the video understanding capabilities of Video MLLMs. Developers should therefore pay particular attention to noise that disrupts pixel continuity when designing and deploying such models in practical application scenarios.

3.1.2 ANALYSIS OF ROBUSTNESS OF VIDEO MLLMs ACROSS VIDEO TYPES

Figure 4 presents the delta performance of representative Video MLLMs on average in groups of different video types. Echoing the video categories mentioned in Section 2.4, we investigate the impact of noise on robustness in different video contents. 16 video categories are identified from the 609 videos¹. According to the performance in Figure 3, we focus on five representative models: two highly robust models (i.e., LLaMA-VID and Chat-UniVi-v1.5), one moderate model (i.e., VideoChat2-HD), and two less robust ones (i.e., PLLaVA 13B and Qwen2.5-VL 3B). Results show that robustness varies substantially by video categories and models. Overall, most models perform poorly on advertisements, suggesting that there may be limited effective training data in popular training corpora. Qwen2.5-VL 3B and PLLaVA 13B are especially non-robust in noisy sports and electronics videos, while Chat-UniVi-v1.5 remains robust on news content. This result shows that each model has its strengths and weaknesses across different video categories. Developers may evaluate and select the most suitable Video MLLM for their specific application, considering the video type factor.

3.1.3 ANALYSIS OF ROBUSTNESS OF VIDEO MLLMs ACROSS QUESTION TYPES

Figure 5 presents the delta performance of representative Video MLLMs in distributions and groups of different question types. The questions are divided into nine types within two broad categories: a) four perception questions, coarse perception (CP), fine-grained perception-single instance (FP-S), fine-grained perception-cross instance (FP-C), and hallucination (HL), and b) five reasoning questions, logic (LR), attribute (AR), relation (RR), common sense (CSR), and temporal reasoning (TR). The results show that model robustness varies notably by question type. For instance, PLLaVA’s GPT score declines by over 13% on perception questions when noise is present, but only drops about 8% on reasoning questions. This difference may be due to the fact that its descriptive answers are more resilient to noise in reasoning tasks. All models struggle significantly with Attribute

¹Three categories are filtered out due to insufficient videos for analysis, i.e., less than 10 videos.

Reasoning questions, exhibiting low robustness and high variance. This suggests that detailed attribute recognition is particularly sensitive to visual noise. Since these tasks require the identification of fine-grained properties such as color, shape, or emotion, noise can severely distort inputs and lead to a considerable decrease in performance. Overall, the effect of visual noise varies by question type, and fine-grained attribute reasoning continues to pose a major challenge for current Video MLLMs in noisy conditions.

3.2 DISCUSSION

Recent advances in image restoration have led to the development of effective methods for reducing visual noise in images. A key question is whether current Video MLLMs can achieve the same level of understanding when analyzing restored (denoised) videos compared to clean videos. To explore this, we employ two different approaches for video enhancement: the classic NAFNet model Chen et al. (2022) for CNN-based denoising and the SeedVR2-7B Wang et al. (2025) which is one diffusion model. We focus on three types of noise for our analysis: impulse noise, motion blur and H.265 which had a significant impact in our previous experiments.

As shown in Figure 6, we compare the results for each noise type across four settings: clean, noisy, and two types of restored videos. Our results reveal that NAFNet denoising maintains similar GPT and SBERT scores to noisy videos while improving accuracy, whereas SeedVR2-7B shows inconsistent performance across different noise types.

These findings suggest that CNN-based restoration enables the model to recover sufficient information for simple True/False questions, but fails to provide the finer details required for precise descriptions or complex reasoning. The diffusion-based approach demonstrates limitations in handling specific noise types, particularly motion blur. Both restoration methods indicate that video understanding tasks demand more detailed and nuanced visual information than simple question answering, and such fine-grained details remain challenging to fully restore with current techniques. This underscores a fundamental challenge: bridging the gap between perceptual visual quality improvement and robust, accurate video-language understanding.

4 CONCLUSION

We present a comprehensive evaluation of state-of-the-art Video MLLMs under various visual noise conditions. By examining both general-purpose and specialized models across diverse video categories and question types, we reveal significant robustness differences and specific challenges. Our results show that robustness largely depends on video domain and task type, with fine-grained reasoning especially sensitive to noise. While image restoration methods can partially recover performance on simple factual questions, they remain insufficient for restoring details needed for accurate descriptions and complex reasoning. These findings highlight the need for more robust Video MLLMs and improved restoration techniques that better preserve semantic and fine-grained information. Our benchmark and insights provide valuable guidance for evaluating and selecting Video MLLMs in practical application scenarios and support future progress in robust video understanding.

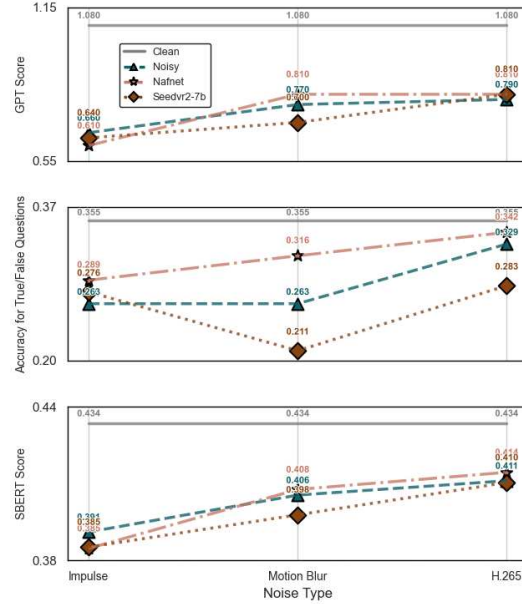


Figure 6: Impact of noise and denoising on Qwen2.5-VL 3B model performance across multiple metrics.

REFERENCES

- Ahmed Abdelkawy, Asem Ali, and Aly Farag. Epam-net: An efficient pose-driven attention-guided multimodal network for video action recognition. *Neurocomputing*, 633:129781, 2025.
- Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- Aishy Amer and Eric Dubois. Fast and reliable structure-oriented video noise estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(1):113–118, 2005.
- Elad Amrani, Rami Ben-Ari, Daniel Rotman, and Alex Bronstein. Noise estimation using density estimation for self-supervised multimodal learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 6644–6652, 2021.
- Charles Bonchelet. Image noise models. In *The essential guide to image processing*, pp. 143–167. Elsevier, 2009.
- Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pp. 17–33. Springer, 2022.
- Qiguang Chen, Libo Qin, Jinhao Liu, Dengyun Peng, Jiannan Guan, Peng Wang, Mengkang Hu, Yuhang Zhou, Te Gao, and Wanxiang Che. Towards reasoning era: A survey of long chain-of-thought for reasoning large language models. *arXiv preprint arXiv:2503.09567*, 2025.
- Shuo Chen, Jindong Gu, Zhen Han, Yunpu Ma, Philip Torr, and Volker Tresp. Benchmarking robustness of adaptation methods on pre-trained vision-language models. *Advances in Neural Information Processing Systems*, 36:51758–51777, 2023.
- Jang Hyun Cho, Andrea Madotto, Effrosyni Mavroudi, Triantafyllos Afouras, Tushar Nagarajan, Muhammad Maaz, Yale Song, Tengyu Ma, Shuming Hu, Suyog Jain, et al. Perceptionlm: Open-access data and models for detailed visual understanding. *arXiv preprint arXiv:2504.13180*, 2025.
- Xinyu Fang, Kangrui Mao, Haodong Duan, Xiangyu Zhao, Yining Li, Dahua Lin, and Kai Chen. Mmbench-video: A long-form multi-shot benchmark for holistic video understanding. *Advances in Neural Information Processing Systems*, 37:89098–89124, 2024.
- Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE transactions on image processing*, 17(10):1737–1754, 2008.
- Chaoyou Fu, Yuhang Dai, Yongdong Luo, Lei Li, Shuhuai Ren, Renrui Zhang, Zihan Wang, Chenyu Zhou, Yunhang Shen, Mengdan Zhang, et al. Video-mme: The first-ever comprehensive evaluation benchmark of multi-modal llms in video analysis. *arXiv preprint arXiv:2405.21075*, 2024.
- Hongcheng Gao, Jiashu Qu, Jingyi Tang, Baolong Bi, Yue Liu, Hongyu Chen, Li Liang, Li Su, and Qingming Huang. Exploring hallucination of large multimodal models in video understanding: Benchmark, analysis and mitigation. *arXiv preprint arXiv:2503.19622*, 2025.
- Noa Garcia, Mayu Otani, Chenhui Chu, and Yuta Nakashima. Knowit vqa: Answering knowledge-based questions about videos. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 10826–10834, 2020.
- Gemini Team. Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.
- Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *European conference on computer vision*, pp. 222–241. Springer, 2024.
- Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019.

- Peng Jin, Ryuichi Takanobu, Wancai Zhang, Xiaochun Cao, and Li Yuan. Chat-univi: Unified visual representation empowers large language models with image and video understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13700–13710, 2024.
- Kibum Kim, Kanghoon Yoon, Yeonjun In, Jaehyeong Jeon, Jinyoung Moon, Donghyun Kim, and Chanyoung Park. Weakly supervised video scene graph generation via natural language supervision. *arXiv preprint arXiv:2502.15370*, 2025.
- Bo Li, Yuanhan Zhang, Dong Guo, Renrui Zhang, Feng Li, Hao Zhang, Kaichen Zhang, Peiyuan Zhang, Yanwei Li, Ziwei Liu, et al. Llava-onevision: Easy visual task transfer. *arXiv preprint arXiv:2408.03326*, 2024a.
- Kunchang Li, Yinan Chen, Yi Wang, Yifeng Cong, Yixuan Chen, Liang Wang, Yizeng Yang, Wei Ran, Ziyi Hu, Yukang Zhu, et al. Videochat: Chat-centric video understanding. *arXiv preprint arXiv:2305.06355*, 2023.
- Kunchang Li, Yali Wang, Yinan He, Yizhuo Li, Yi Wang, Yi Liu, Zun Wang, Jilan Xu, Guo Chen, Ping Luo, et al. Mvbench: A comprehensive multi-modal video understanding benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22195–22206, 2024b.
- Linjie Li, Zhe Gan, and Jingjing Liu. A closer look at the robustness of vision-and-language pre-trained models. *arXiv preprint arXiv:2012.08673*, 2020.
- Yanwei Li, Chengyao Wang, and Jiaya Jia. Llama-vid: An image is worth 2 tokens in large language models. In *European Conference on Computer Vision*, pp. 323–340. Springer, 2024c.
- SukHwan Lim. Characterization of noise in digital photographs for image processing. In *Digital Photography II*, volume 6069, pp. 219–228. SPIE, 2006.
- Yang Lin, Shengqiong Jia, Xinjian Lin, Haotian Wang, Credelle Xu, Xiang Wang, Song Bai, Xinfeng Liu, Ming-Hsuan Yang, and Jindong Li. Video-llava: Learning united visual representation by alignment before projection. *arXiv preprint arXiv:2311.10122*, 2023.
- Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916, 2023.
- Muhammad Maaz, Abhinav Shrivastava, and Mahyar Najibi. Video-chatgpt: Towards detailed video understanding via large vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.
- Karttikeya Mangalam, Raiymbek Akshulakov, and Jitendra Malik. Egoschema: A diagnostic benchmark for very long-form video language understanding. *Advances in Neural Information Processing Systems*, 36:46212–46244, 2023.
- Jihyung Moon, Hyunchoo Cho, and Eunjeong L Park. Revisiting round-trip translation for quality estimation. *arXiv preprint arXiv:2004.13937*, 2020.
- Sankalp Nagaonkar, Augustya Sharma, Ashish Choithani, and Ashutosh Trivedi. Benchmarking vision-language models on optical character recognition in dynamic video environments. *arXiv preprint arXiv:2502.06445*, 2025.
- Munan Ning, Bin Zhu, Yujia Xie, Bin Lin, Jiayi Cui, Lu Yuan, Dongdong Chen, and Li Yuan. Video-bench: A comprehensive benchmark and toolkit for evaluating video-based large language models. *arXiv preprint arXiv:2311.16103*, 2023.
- OpenAI. Gpt-4v(ision) system card. Technical report, OpenAI, 2023.
- OpenAI. Hello gpt-4o. <https://openai.com/index/hello-gpt-4o>, 2024. Accessed: 2024-05-12.
- Stuart Perry. Image and video noise: an industry perspective. In *Denoising of photographic images and video: fundamentals, open challenges and new trends*, pp. 207–234. Springer, 2018.

- Stanislav Pyatykh, Jürgen Hesser, and Lei Zheng. Image noise level estimation by principal component analysis. *IEEE transactions on image processing*, 22(2):687–699, 2012.
- Jielin Qiu, Yi Zhu, Xingjian Shi, Florian Wenzel, Zhiqiang Tang, Ding Zhao, Bo Li, and Mu Li. Benchmarking robustness of multimodal image-text models under distribution shift. *arXiv preprint arXiv:2212.08044*, 2022.
- G Ramesh, J Logeshwaran, J Gowri, and Ajay Mathew. The management and reduction of digital noise in video image processing by using transmission based noise elimination scheme. *ICTACT Journal on Image & Video Processing*, 13(1), 2022.
- Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.
- Madeline Schiappa, Shruti Vyas, Hamid Palangi, Yogesh Rawat, and Vibhav Vineet. Robustness analysis of video-language models against visual and language perturbations. *Advances in Neural Information Processing Systems*, 35:34405–34420, 2022.
- Abdelrahman Shaker, Muhammad Maaz, Chenhui Gou, Hamid Rezaatofighi, Salman Khan, and Fahad Shahbaz Khan. Mobile-videogpt: Fast and accurate video understanding language model. *arXiv preprint arXiv:2503.21782*, 2025.
- Alexander Shirnin, Nikita Andreev, Sofia Potapova, and Ekaterina Artemova. Analyzing the robustness of vision & language models. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024.
- Shitong Sun, Jindong Gu, and Shaogang Gong. Benchmarking robustness of text-image composed retrieval. *arXiv preprint arXiv:2311.14837*, 6, 2023.
- Supriya, Jyoti Prakash Singh, and Gunjan Kumar. Identification of clickbait news articles using sbert and correlation matrix. *Social Network Analysis and Mining*, 13(1):153, 2023.
- Xichen Tan, Yuanjing Luo, Yunfan Ye, Fang Liu, and Zhiping Cai. Allvb: All-in-one long video understanding benchmark. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 7211–7219, 2025.
- Yunlong Tang, Jing Bi, Siting Xu, Luchuan Song, Susan Liang, Teng Wang, Daoan Zhang, Jie An, Jingyang Lin, Rongyi Zhu, et al. Video understanding with large language models: A survey. *arXiv preprint arXiv:2312.17432*, 2023.
- Makarand Tapaswi, Yukun Zhu, Rainer Stiefelhofen, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. Movieqa: Understanding stories in movies through question-answering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4631–4640, 2016.
- Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.
- Jianyi Wang, Zhijie Lin, Meng Wei, Yang Zhao, Ceyuan Yang, Chen Change Loy, and Lu Jiang. Seedvr: Seeding infinity in diffusion transformer towards generic video restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 2161–2172, 2025.
- Wei Wang, Xin Chen, Cheng Yang, Xiang Li, Xuemei Hu, and Tao Yue. Enhancing low light videos by exploring high sensitivity camera noise. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4111–4119, 2019a.
- Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 0–0, 2019b.
- Yao Wang and Qin-Fan Zhu. Error control and concealment for video communication: A review. *Proceedings of the IEEE*, 86(5):974–997, 1998.

- Frank Wilcoxon. Individual comparisons by ranking methods. In *Breakthroughs in statistics: Methodology and distribution*, pp. 196–202. Springer, 1992.
- Junbin Xiao, Nanxin Huang, Hangyu Qin, Dongyang Li, Yicong Li, Fengbin Zhu, Zhulin Tao, Jianxing Yu, Liang Lin, Tat-Seng Chua, et al. Videoqa in the era of llms: An empirical study. *International Journal of Computer Vision*, pp. 1–24, 2025.
- Shaoyuan Xie, Lingdong Kong, Yuhao Dong, Chonghao Sima, Wenwei Zhang, Qi Alfred Chen, Ziwei Liu, and Liang Pan. Are vlms ready for autonomous driving? an empirical study from the reliability, data, and metric perspectives. *arXiv preprint arXiv:2501.04003*, 2025.
- Jun Xu, Tao Mei, Ting Zeng, and Jinhui Yu. Videoqa: A comprehensive benchmark for video question answering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.
- Lin Xu, Yilin Zhao, Daquan Zhou, Zhijie Lin, See Kiong Ng, and Jiashi Feng. Pllava: Parameter-free llava extension from images to videos for video dense captioning. *arXiv preprint arXiv:2404.16994*, 2024a.
- Lu Xu, Sijie Zhu, Chunyuan Li, Chia-Wen Kuo, Fan Chen, Xinyao Wang, Guang Chen, Dawei Du, Ye Yuan, and Longyin Wen. Beyond raw videos: Understanding edited videos with large multimodal model. *arXiv preprint arXiv:2406.10484*, 2024b.
- An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, et al. Qwen2. 5 technical report. *arXiv preprint arXiv:2412.15115*, 2024.
- Runhao Zeng, Xiaoyong Chen, Jiaming Liang, Huisi Wu, Guangzhong Cao, and Yong Guo. Benchmarking the robustness of temporal action detection models against temporal corruptions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18263–18274, 2024.
- Hang Zhang, Kai Li, et al. Video-llama: An instruction-tuned audio-visual language model for video understanding. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023.
- Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4791–4800, 2021.
- Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, and Xuanjing Huang. Extractive summarization as text matching. *arXiv preprint arXiv:2004.08795*, 2020.
- Junjie Zhou, Yan Shu, Bo Zhao, Boya Wu, Shitao Xiao, Xi Yang, Yongping Xiong, Bo Zhang, Tiejun Huang, and Zheng Liu. Mlvu: A comprehensive benchmark for multi-task long video understanding. *arXiv preprint arXiv:2406.04264*, 2024.
- Vladimir Zlokolica, Aleksandra Pizurica, and Wilfried Philips. Noise estimation for video processing based on spatio-temporal gradients. *IEEE Signal Processing Letters*, 13(6):337–340, 2006.

A APPENDIX

This appendix provides supplementary information to our main paper. We first shows the related work, second we present the noise parameter settings used in NoisyVideo to facilitate reproducibility. Thirdly, we describe model implementation details, followed by a discussion of the limitations of our work. Then we declare the use of LLM in our benchmark. We also present the additional experimental results and visualizations not included in the main manuscript. Finally, we present the License of the our work.

A.1 RELATED WORK

A.1.1 VIDEO MLLMS FOR VIDEO UNDERSTANDING

In recent years, LLMs have rapidly evolved towards multimodal capabilities, exemplified by GPT OpenAI (2023), Gemini Gemini Team (2023), and Qwen Yang et al. (2024), which aim to provide general understanding ability across multiple modalities. To better address video-related tasks, a new line of research has emerged focusing on specialized LLMs for video understanding. Early works such as Video-ChatGPT Maaz et al. (2023) aligned frame-level visual features with language embeddings through linear projectors, while VideoChat Li et al. (2023) added a learnable Q-former structure. Video-LLaMA Zhang et al. (2023) integrated audio features, and Video-LLaVA Lin et al. (2023) learned from mixed image and video datasets. In this work, we construct a benchmark to evaluate the video understanding capabilities of both general MLLMs and video-specialized models under various visual noise conditions.

A.1.2 VIDEO UNDERSTANDING

Currently, video question answering (VideoQA) is the predominant kind of task for evaluating a model’s video understanding capabilities Xu et al. (2022); Xiao et al. (2025). Early benchmarks typically focused on a single type of video content Tapaswi et al. (2016); Garcia et al. (2020); Mangalam et al. (2023). In recent years, VideoQA tasks have expanded to cover a much broader range of video genres and question types Li et al. (2024b); Ning et al. (2023); Fu et al. (2024); Fang et al. (2024). Video-MME Fu et al. (2024) is the comprehensive evaluation benchmark for Video MLLM. MLVU Zhou et al. (2024) focuses on long videos. MMBench-Video Fang et al. (2024) covers 16 types of videos and 26 categories of questions. Due to the complexity and diversity of question types in current VideoQA datasets, including the challenging type of open-ended questions, LLMs such as GPT are usually employed as evaluation methods Maaz et al. (2023). Unlike the existing benchmarks, we focus on visual noise in videos and conduct a comprehensive and multi-faceted evaluation for video understanding under noise conditions.

A.1.3 VISUAL MODEL ROBUSTNESS

Recently, robustness analysis in multimodal tasks has emerged, which considers the real-world situations and takes a further step towards a more reliable system. Yet, this field is still in its infancy. Li et al. Li et al. (2020) take the first step to analyze the robustness in a multimodal task, Visual Question Answering (VQA), against 4 generic corruptions, including linguistic variation and visual content manipulation. Schiappa et al. Schiappa et al. (2022) introduce naturally corrupted visual and textual benchmarks on text-to-video retrieval. Chen et al. Chen et al. (2023) assess the robustness of 11 widely-used adaptation methods across 4 vision-language datasets under mixed corruptions. Xu et al. Xu et al. (2024b) propose a benchmark to evaluate the model robustness based on edited videos. Zeng et al. Zeng et al. (2024) analyze the robustness of seven leading temporal action detection (TAD) methods. However, few of these benchmarks have systematically analyzed the robustness of Video MLLMs to visual noise. Our benchmark fills this gap by exploring a comprehensive evaluation in this area.

A.2 NOISE IMPLEMENTATION DETAILS

Our benchmark categorizes noise into 8 major types, and we employ PSNR and SSIM as control metrics to ensure fair comparison across different categories. We implement a two-level control strategy: first, we normalize perturbations within each noise category to achieve similar PSNR

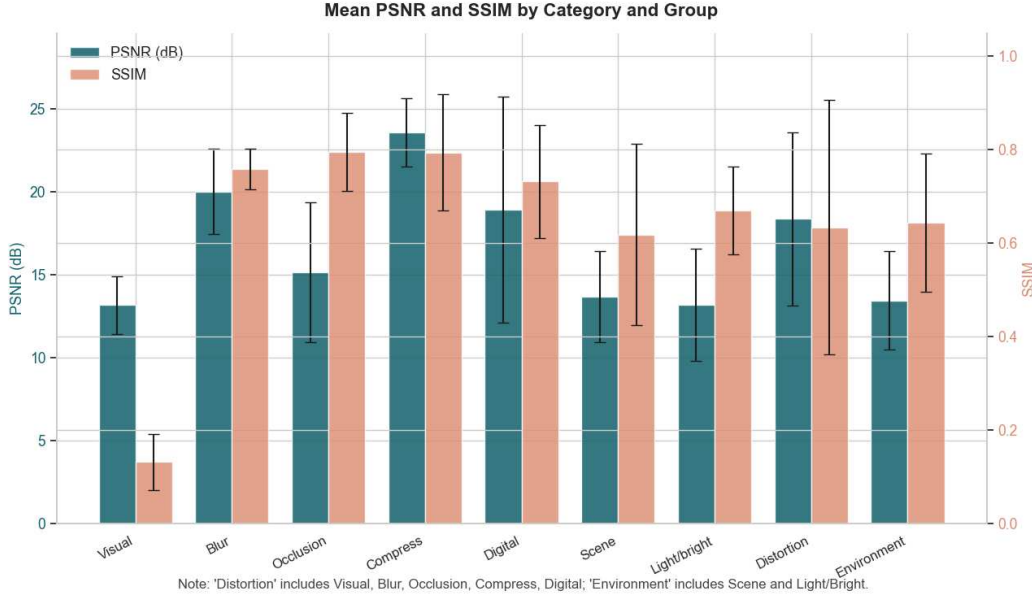


Figure 7: PSNR and SSIM for one video demo

and SSIM values; second, we further classify these 8 noise categories into two higher-level groups based on whether they introduce corruptions in pixel continuity. Specifically, we categorize noises that disrupt pixel continuity as “distortions”, including Quality noise, Blur noise, Temporal noise, Compression, Occlusion, and Digital distortions. The remaining categories—Lighting/color noise and Scene interference—are classified as “environment noise” as they preserve pixel continuity. Since our classification criterion focuses on pixel continuity, we particularly emphasize controlling SSIM values to be nearly identical between the distortion and environment noise groups, while also maintaining PSNR differences within reasonable ranges. This controlled approach ensures that we mitigate the effects of different noise strengths, allowing for fair comparison across categories.

Note that some references use 3 Zeng et al. (2024) or 5 levels Hendrycks & Dietterich (2019) for each perturbation and report averaged results across these levels, while we directly control perturbation strength through PSNR and SSIM metrics to achieve more precise normalization. Both methodologies serve the same fundamental purpose: controlling variables to enable fair comparison across different perturbations. We use one video to show the demo PSNR/SSIM value, as shown in Figure 7. We detailed noise setting is shown as below:

Gaussian noise: Additive noise sampled from a zero-mean normal distribution with standard deviation 100 is added to each frame, and results are clamped to the valid pixel range [0, 255].

Impulse noise: A uniform mask is sampled for each frame. With default probability 0.7, pixels with mask values below 0.35 are set to black (0), those between 0.35 and 0.7 are set to white (255), and the rest remain unchanged.

Speckle noise: Multiplicative noise with intensity 0.7 is applied. The noise map is generated at full resolution for single-pixel grains or upsampled from a lower resolution patch for larger grains. Noisy pixels are computed as $I_{\text{noisy}} = I \times (1 + \text{noise})$, with results clamped to [0, 255].

Poisson noise: Each frame is scaled by gain (default 0.01), used as the rate parameter for Poisson sampling, then divided by gain to restore scale, and the results are clamped to [0, 255].

Gaussian blur: A 2D Gaussian kernel of size 101 with $\sigma = 20$ is generated and applied via depth-wise convolution across all channels for each affected frame, smoothing image details according to the specified kernel.

Motion blur: A 2D motion kernel of size 101 oriented at 45° is generated and applied via depth-wise convolution across all channels for each frame, averaging pixel values along the specified direction to simulate directional motion blur.

Defocus blur: A disk-shaped convolution kernel is generated by thresholding a circular mask of radius r_s (with $r_s \in \{3, 4, 6, 8, 10\}$ for severity levels $s = 1, \dots, 5$) and applying Gaussian anti-aliasing with standard deviation $\sigma_s \in \{0.1, 0.5, 0.5, 0.5, 0.5\}$. This smoothed disk kernel is then applied per channel via 2D filtering to simulate an out-of-focus effect, with pixel values clamped to the valid range $[0, 255]$.

Glass blur: A two-stage Gaussian smoothing (with $\sigma \in \{0.7, 0.9, 1.0, 1.1, 1.5\}$ for severity levels 1–5) is applied before and after a local pixel shuffle (max displacement $\Delta \in \{1, 2, 2, 3, 4\}$ pixels over $N \in \{2, 1, 3, 2, 2\}$ iterations), creating a refractive “frosted glass” effect.

Zoom blur: For severity $s \in \{1, \dots, 5\}$, define the zoom factors $c_s = \{1.0, 1.0 + \Delta_s, \dots, 1.0 + 0.1s\}$ with $\Delta_s \in \{0.01, 0.01, 0.02, 0.02, 0.03\}$ for $s = 1, \dots, 5$; for each $f \in c_s$, crop the central $\frac{H}{f} \times \frac{W}{f}$ region, bilinearly resize it to $H \times W$, sum these zoomed frames with the original, divide by $|c_s| + 1$, and clamp to $[0, 255]$ to produce the radial zoom-blur effect.

Frame drop: A random subset of frames is replaced with blank frames by setting every pixel channel to 0, simulating the complete loss of those frames.

Frame replace: A subset of frames is selected according to the noise ratio; those frames are randomly permuted among themselves, and the shuffled frames are reinserted in place to simulate out-of-order frame delivery.

Frame repeat: A subset of frames is selected as reference frames, and each remaining frame is replaced by its temporally nearest reference frame, producing repeated-frame artifacts throughout the video.

Temporal jitter: A random split of the noise ratio ρ is computed by sampling $\alpha = \rho \cdot U(0, 1)$ and setting $\beta = 1 - \alpha$; then frame replacement (with fraction α) and frame drop (with fraction β) are applied in sequence according to the specified execution order (default `replace_first`), yielding a mixture of shuffled and omitted frames.

JPEG artifact: A subset of frames selected by the noise ratio is encoded to JPEG at quality level $q \in [1, 100]$ (default $q = 1$) and then decoded, introducing quantization and blocking distortions characteristic of low-quality JPEG compression.

Bit error: A subset of frames (controlled by the noise ratio) is selected, and in each a random rectangular region covering between $\frac{1}{4}$ and $\frac{1}{2}$ of the frame is chosen; within that region vertical stripes of width $w \in \{15, 10, 5\}$ for severity levels {low, medium, high} are created by copying pixels from randomly sampled columns, yielding a streaked stripe-artifact effect.

H.265 artifacts: A two-pass H.265 encoder is applied with the quantization parameter $\text{CRF} \in \{32, 38, 45, 51\}$ and bitrate $\in \{400\text{k}, 200\text{k}, 100\text{k}, 50\text{k}\}$ for low-extreme severity levels, where frames are first downsampled to $\frac{1}{4}$ resolution before reencoding at full size, producing pronounced blocking, ringing, and other compression artifacts.

Random block: For each selected frame, a rectangular region of height $h \in [0.3H, 0.7H]$ and width $w \in [0.3W, 0.7W]$ is sampled at a random position and all its pixel values are set to zero, simulating partial occlusion.

Target block: For each selected frame, a pre-trained detection model predicts the bounding box of the primary object (x_1, y_1, x_2, y_2) , and all pixels within that region are set to zero to occlude the main subject.

Rolling shutter: A subset of frames (ratio ρ) is chosen and for each selected frame at index t , a random orientation (horizontal or vertical) and direction are picked; then for each column x (or row y), a temporal offset

$$\delta = \min\left(\left\lfloor \frac{x}{W-1} d \min(t, B) \right\rfloor, B\right)$$

(with width W or height H , delay factor d , and buffer size B) is computed, and the pixel strip at x (or y) is sampled from frame $t - \delta$, yielding a characteristic rolling-shutter smear.

Resolution degrade: Each selected frame is downsampled via bicubic interpolation to $(\lfloor sH \rfloor, \lfloor sW \rfloor)$ with scale factor s (default $s = 0.1$) and then upsampled back to (H, W) via bicubic interpolation, with pixel values clamped to the valid range $[0, 255]$.

Stretch/squish: A random orientation (horizontal or vertical) is selected, then each chosen frame is downsampled via bicubic interpolation along that axis to size $(H, \lfloor sW \rfloor)$ or $(\lfloor sH \rfloor, W)$ with scale factor $s = \frac{1}{30}$, and finally upsampled back to (H, W) via bicubic interpolation, with pixel values clamped to $[0, 255]$.

Edge sawtooth: Edges are detected via Canny with thresholds $(T_{\text{low}} = 50, T_{\text{high}} = 150)$, dilated by a 2×2 kernel, then a surround mask of neighboring pixels is sampled with probability $\rho = 0.3$, and all masked pixels are replaced with uniformly random RGB values to produce serrated, noisy edges.

Color quantization: Each selected frame’s pixel values are uniformly quantized to 2^b levels by computing

$$\tilde{I} = \text{round}\left(\frac{I}{\Delta}\right) \times \Delta, \quad \Delta = \frac{255}{2^b - 1} \quad (b = 3 \text{ by default}),$$

and then clamping $\tilde{I}$ to the valid range $[0, 255]$.

Elastic transform: For severity $s \in \{1, \dots, 5\}$, generate random displacement fields Δ_x, Δ_y by sampling uniformly in $[-\delta_{\text{max}}, \delta_{\text{max}}]$ with $\delta_{\text{max}} = 0.005 \min(H, W)$, smoothing each via a Gaussian filter with $\sigma = 0.01 \min(H, W)$, and scaling by $\alpha_s \in \{12.5, 16.25, 21.25, 25, 30\}$; the warped image is then obtained by remapping pixels from (x, y) to $\lfloor x + \Delta_x(x, y), y + \Delta_y(x, y) \rfloor$ using bilinear interpolation, and clamped to $[0, 255]$.

Brightness transform: For severity levels $s \in \{1, \dots, 5\}$, let $c_s = 0.1s$; each selected frame is converted to HSV (for color), its V channel is increased by c_s (clipped to $[0, 1]$) before converting back to RGB, or for grayscale frames the intensity is simply offset by c_s (clipped to $[0, 1]$), then rescaled to $[0, 255]$.

Contrast transform: For each selected frame, compute its mean intensity μ , sample a factor $\alpha \sim \mathcal{U}(m, M)$ with $(m, M) = (-5, 5)$, replace α by $-1/\alpha$ if $\alpha < 0$, and apply

$$I_{\text{out}} = \text{clip}(\alpha(I - \mu) + \mu, 0, 255),$$

thereby symmetrically scaling pixel deviations around the original mean.

Color shift: For each selected frame, sample multiplicative gains $\alpha_k \sim \mathcal{U}(1-s, 1+s)$ for $k = 1, 2, 3$ (or a single α if `per_channel=false`), multiply each channel by its α_k , and clamp the result to $[0, 255]$, where s is the shift magnitude (default $s = 1$).

Flicker: For each frame, sample $\alpha \sim \mathcal{N}(1, \rho^2)$ with ρ equal to the flicker ratio (default 0.2), multiply all pixel values by α , and clamp to $[0, 255]$, producing random frame-level brightness fluctuations.

Overexposure: For each selected frame, sample a brightness offset $b \sim \mathcal{U}(0.1, 0.3)$ and a gamma exponent $\gamma \sim \mathcal{U}(1.1, 1.4)$; first add b to normalized intensity $(I/255)$, then apply inverse gamma correction $(I')^{1/\gamma}$, and rescale to $[0, 255]$, yielding locally blown-out highlights.

Underexposure: For each selected frame, sample a brightness offset $b \sim \mathcal{U}(-0.3, -0.1)$ and a gamma exponent $\gamma \sim \mathcal{U}(0.6, 0.9)$; first add b to normalized intensity $(I/255)$, then apply inverse gamma correction $(I')^{1/\gamma}$, and rescale to $[0, 255]$, producing dimmed, murky shadows.

Rain. For severity level $s \in \{1, \dots, 5\}$, define (m_s, L_s, ρ_s) as the s th row of

$$\begin{pmatrix} 0.05 & 8 & 1 \\ 0.07 & 10 & 2 \\ 0.09 & 15 & 3 \\ 0.11 & 18 & 4 \\ 0.13 & 22 & 5 \end{pmatrix},$$

where the three columns correspond to m_s, L_s and ρ_s , respectively. Blur this density map with a Gaussian filter ($\rho = 0.5$), then blend the original image of each frame I and the rain layer R by

$$I_{\text{rain}} = \text{clip}(I(1 - 0.3R) + 200Rb_s, 0, 255),$$

using the brightness–contrast pairs (b_s, c_s) from

$$\{(0.20, 1.03), (0.25, 1.04), (0.30, 1.05), (0.35, 1.06), (0.40, 1.07)\},$$

and finally add the fog weighted by $f_s \in \{0.05, 0.07, 0.10, 0.13, 0.15\}$.

Fog: For severity levels $s \in \{1, \dots, 5\}$, generate a plasma fractal map F of size $2^{\lceil \log_2(\max(H, W)) \rceil}$ via the diamond-square algorithm with decay parameter $d_s \in \{2, 2, 1.7, 1.5, 1.4\}$ for each s , normalize $F \in [0, 1]$ and crop to (H, W) , then for each selected frame normalized to $I \in [0, 1]$, compute

$$I_{\text{fog}} = \text{clip}\left(\frac{I + c_s F}{1 + c_s}, 0, 1\right) \times 255,$$

where $c_s \in \{1.5, 2.0, 2.5, 2.5, 3.0\}$, producing a depth-varying haze effect.

Snow. For severity level $s \in \{1, \dots, 5\}$, the seven-tuple $(\mu_s, \sigma_s, r_s, t_s, R_s, \Sigma_s, \beta_s)$ is defined by the sth row of

$$\begin{pmatrix} 0.10 & 0.30 & 3.0 & 0.50 & 10 & 4 & 0.80 \\ 0.20 & 0.30 & 2.0 & 0.50 & 12 & 4 & 0.70 \\ 0.55 & 0.30 & 4.0 & 0.90 & 12 & 8 & 0.70 \\ 0.55 & 0.30 & 4.5 & 0.85 & 12 & 8 & 0.65 \\ 0.55 & 0.30 & 2.5 & 0.85 & 12 & 12 & 0.55 \end{pmatrix},$$

where columns correspond to $\mu_s, \sigma_s, r_s, t_s, R_s, \Sigma_s, \beta_s$. Generate a Gaussian snow layer $S \sim \mathcal{N}(\mu_s, \sigma_s^2)$, clip values below t_s , zoom by factor r_s , and apply a 1D motion blur with $K = 2\lceil \sigma_s \rceil + 1$ and $\rho = R_s$ at a random angle. Finally blend via

$$I_{\text{snow}} = \text{clip}\left(\beta_s I + (1 - \beta_s) \max(I, 1.5 \text{ gray}(I) + 0.5) + S + \text{rot90}(S, 2), 0, 1\right) \times 255.$$

Frost: For severity levels $s \in \{1, \dots, 5\}$, let

$$(\lambda_s, \mu_s) \in \{(1, 0.4), (0.8, 0.6), (0.7, 0.7), (0.65, 0.7), (0.6, 0.75)\}$$

respectively; a frost texture F is randomly selected, resized, and tiled to cover the frame, and the output is

$$I_{\text{frost}} = \text{clip}(\lambda_s I + \mu_s F, 0, 255),$$

blending original content with ice-crystal patterns.

Specular reflection: For each selected frame, compute a highlight mask by thresholding the mean intensity per pixel above 100, apply Gaussian blur in depth (kernel size 15, $\sigma = 5$) to the frame, blend original and blurred via $1.5 \times \text{orig} + 0.5 \times \text{blur}$ in highlight regions, then combine with a pregenerated Perlin noise layer via $0.9 \times \text{frame} + 0.1 \times \text{noise}$ and clamp to $[0, 255]$ to simulate mirror-like reflections.

Shadow: For each selected frame of size $H \times W$, sample

$$x_0 \sim \mathcal{U}\left(\frac{W}{4}, \frac{3W}{4}\right), \quad y_0 \sim \mathcal{U}\left(\frac{H}{4}, \frac{3H}{4}\right), \quad a \sim \mathcal{U}\left(\frac{W}{4}, \frac{W}{2}\right), \quad b \sim \mathcal{U}\left(\frac{H}{4}, \frac{H}{2}\right), \quad \theta \sim \mathcal{U}(0, 180^\circ).$$

sample an elliptical mask M defined by center (x_0, y_0) , axes (a, b) , and rotation θ , blur M with a Gaussian kernel of size 51×51 to obtain $\widetilde{M}$, and compute

$$I_{\text{shadow}} = \text{clip}(I(1 - 0.7 \widetilde{M}), 0, 255),$$

producing soft, occluding shadows.

Since video models process multiple frames as input, we conducted a preliminary experiment to investigate how the proportion of noisy frames affects model performance. Specifically, we varied the percentage of frames corrupted by noise and observed that model accuracy decreased most significantly when noise was applied to 90% of the frames (see Figure 8). Based on this observation, and to better amplify the impact of noise, we set the severe noise level in our benchmark to correspond to corrupting 90% of the input frames.

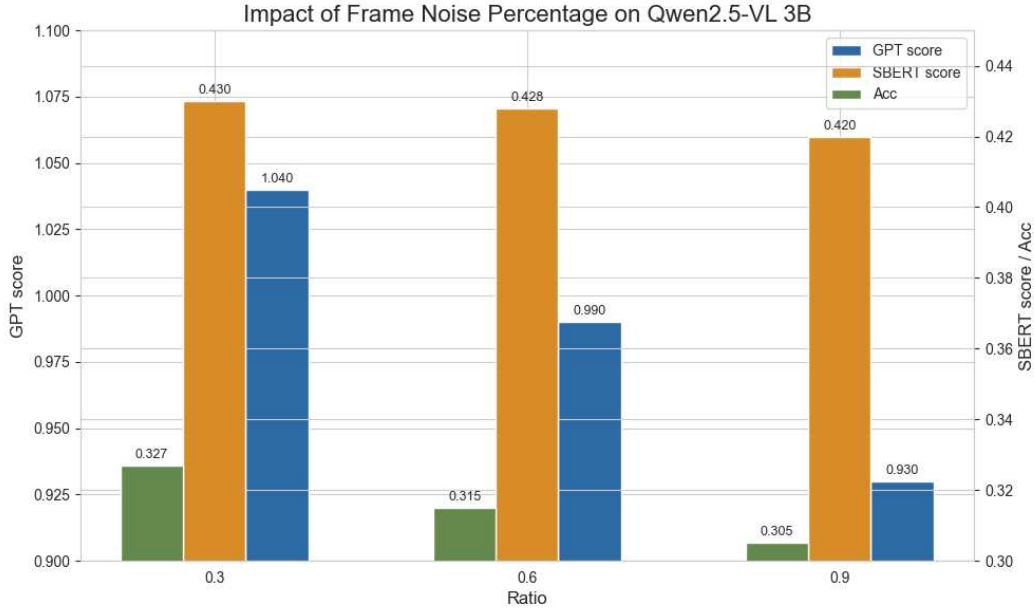


Figure 8: Impact of Frame Noise Percentage

A.3 MODEL IMPLEMENTATION DETAILS

We evaluated a diverse set of state-of-the-art video-language models, each with distinct architectural designs and input image resolutions:

- **Video-LLaVA-7B**: Extends the LLaVA architecture to videos by processing multiple frames using a CLIP ViT-L/14 vision encoder and aligning visual features with a Vicuna language model, enabling strong zero-shot performance on video understanding tasks. [Code]
- **VideoChat2-HD**: Following the BLIP2 framework, this model utilizes a QFormer to bridge vision and language, supporting high-resolution video inputs for enhanced video-language understanding and dialogue. [Code]
- **Chat-UniVi-7B / 7B-v1.5**: Employs a unified visual representation with dynamic visual tokens, using a shared visual encoder and Vicuna LLM for both image and video tasks, thus achieving versatility across different visual modalities. [Code]
- **LLaMA-VID-7B**: Utilizes a token compression mechanism to efficiently process long video sequences, equipped with an EVA-G visual encoder and a Vicuna-7B language backbone for improved scalability and efficiency in video analysis. [Code]
- **Video-ChatGPT**: Incorporates both spatial and temporal pooling after CLIP-based visual encoding, followed by a Vicuna v1.1 language backbone, enabling robust video-language interaction and reasoning. [Code]
- **PLLaVA-7B / 13B**: A parameter-free extension of LLaVA that introduces temporal pooling on top of a CLIP vision encoder, designed to enhance the robustness and efficiency of video captioning without increasing model parameters. [Code]
- **Qwen2.5-VL-3B-Instruct**: Qwen’s multimodal instruction-following model, equipped with a vision transformer encoder and a 3B-parameter language model, designed for strong performance on vision-language tasks. [Code]
- **GPT-4o**: The latest flagship multimodal model from OpenAI, capable of high-performance visual and language understanding across both images and videos, featuring a proprietary architecture and dynamic input resolution. (Closed-source)

All models were evaluated on 8 NVIDIA A100 GPUs with Intel(R) Xeon(R) Platinum 8358 CPUs. For fair comparison, we set the number of input frames to 8 for all models. Our benchmark also allows flexible frame number selection, enabling users to specify the frame count as needed.

A.4 LIMITATIONS

We develop a benchmark to evaluate the robustness of video-language models (video-LLMs) against visual noise. The main limitations of our benchmark are as follows: it is challenging to obtain real-world noisy videos, so we simulate visual noise to approximate real-world conditions. Since we use GPT-4o as the judge for our benchmark, there are associated costs. Due to budget constraints, we selected only a few representative video-LLMs for comparison. Our benchmark includes a large number of noise types. Running the full benchmark is equivalent to running 36 visual clean benchmarks, which limits the duration and quantity of clean videos in our selected dataset. However, our codebase allows users to freely choose any clean dataset according to their needs and budget.

A.5 DECLARATION OF LLM USAGE

Our benchmark follows previous video understanding benchmarks by using GPT-4o OpenAI (2024) as a judge to evaluate the performance of Video MLLMs when faced with visual noise. Through our experiments, we found that relying solely on GPT as a judge can introduce bias. Therefore, we designed our evaluation metrics to assess model performance from both subjective and objective perspectives. The details of metrics are listed in Section 2.3. Additionally, we also describe the Video MLLMs in Section 2.4 and Appendix A.3.

A.6 MODEL FACTORS ON ROBUSTNESS OF VIDEO MLLMs

We observe that *the model architecture has strong connections with the robustness in video understanding*. We intensively include the spatio-temporal enhanced models (i.e., Chat-UniVi models and Video-ChatGPT). Generally, they achieve the smallest absolute delta GPT scores on average. This likely benefits from their architectural designs that explicitly decouple temporal and spatial information. In contrast, PLLaVA models perform the worst in specialized Video MLLMs. This may be due to PLLaVA employing adaptive pooling to synchronize the reduction of both temporal and spatial dimensions.

Based on Figure 3, we further compare the PLLaVA 13B (column D) with its 7B version (column J) and find out that *increasing the model size does not necessarily improve the robustness*. For many cases in these two columns, the performance drops in column D are larger than those in column J, suggesting that model size is not the key bottleneck towards robustness.

Comparing the data-enhanced version of Chat-UniVi-v1.5 (column G) with its vanilla version (i.e., column H of Chat-UniVi), we find that *training with more data does not contribute to higher robustness*. We conduct the Wilcoxon signed-rank test Wilcoxon (1992) (paired test) on the two sets of values and cannot conclude that there is a significant difference at the confidence level of 0.05.

A.7 OTHER EXPERIMENT RESULTS

In this section, we provide the full experimental results corresponding to the analyses presented in Section 4 of the main text, specifically the “Analysis of Robustness of Video MLLMs across Video Types” and the “Analysis of Robustness of Video MLLMs across Question Types.” Our evaluation covers 10 representative models; users interested in other models are encouraged to conduct additional tests as needed.

Tables 2–11 report robustness results across different video types, while Tables 12–21 summarize the results for different question types. Differing from the main text, which focuses on delta performance, the tables here display the raw output scores for each model. For clarity, the lowest value in each column—corresponding to the greatest performance drop under a particular noise type—is highlighted in gray.

We select five models and analyze each type of noise as shown in Figure 9–Figure 11. Across all three metrics (**Acc**, **SBERT Score**, and **GPT Score**), we observe that the impact of noise types on model

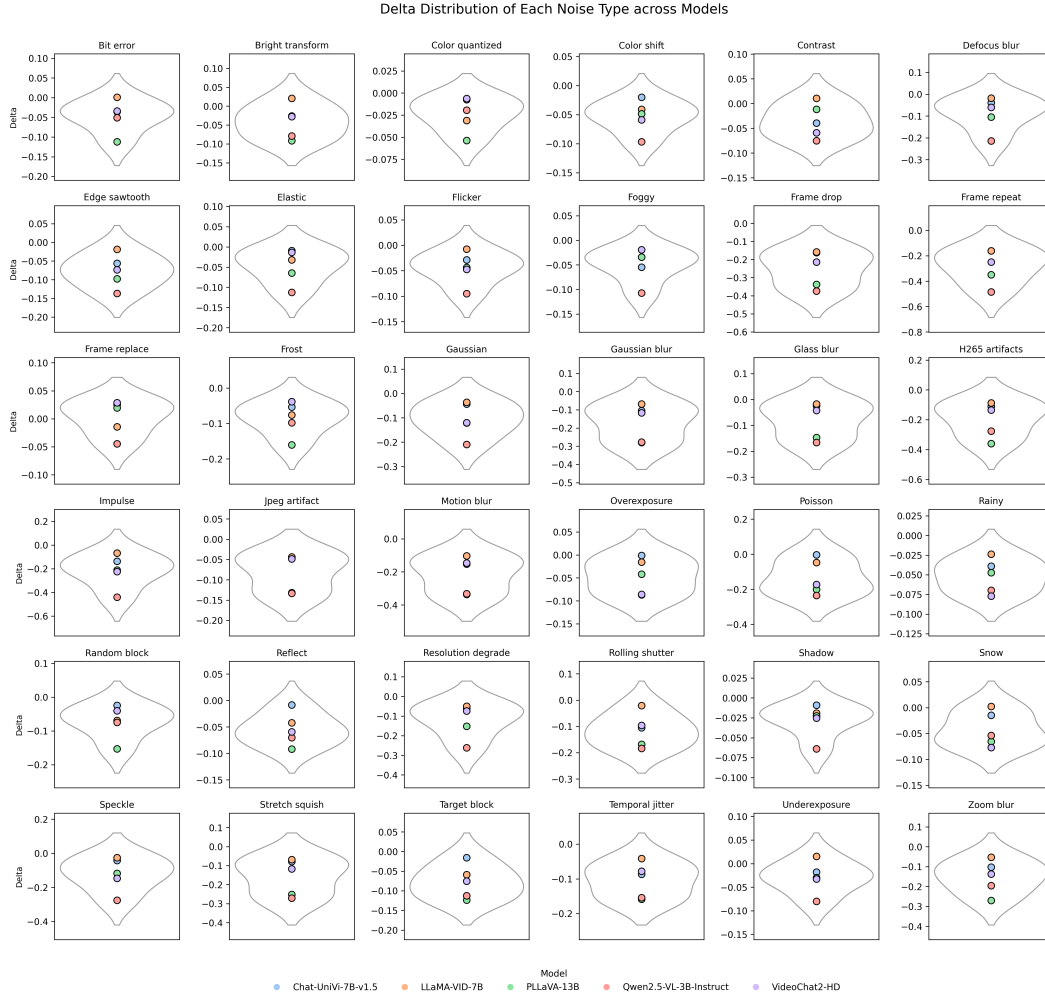


Figure 9: Delta Distribution of Each Noise Type across Models (GPT Score).

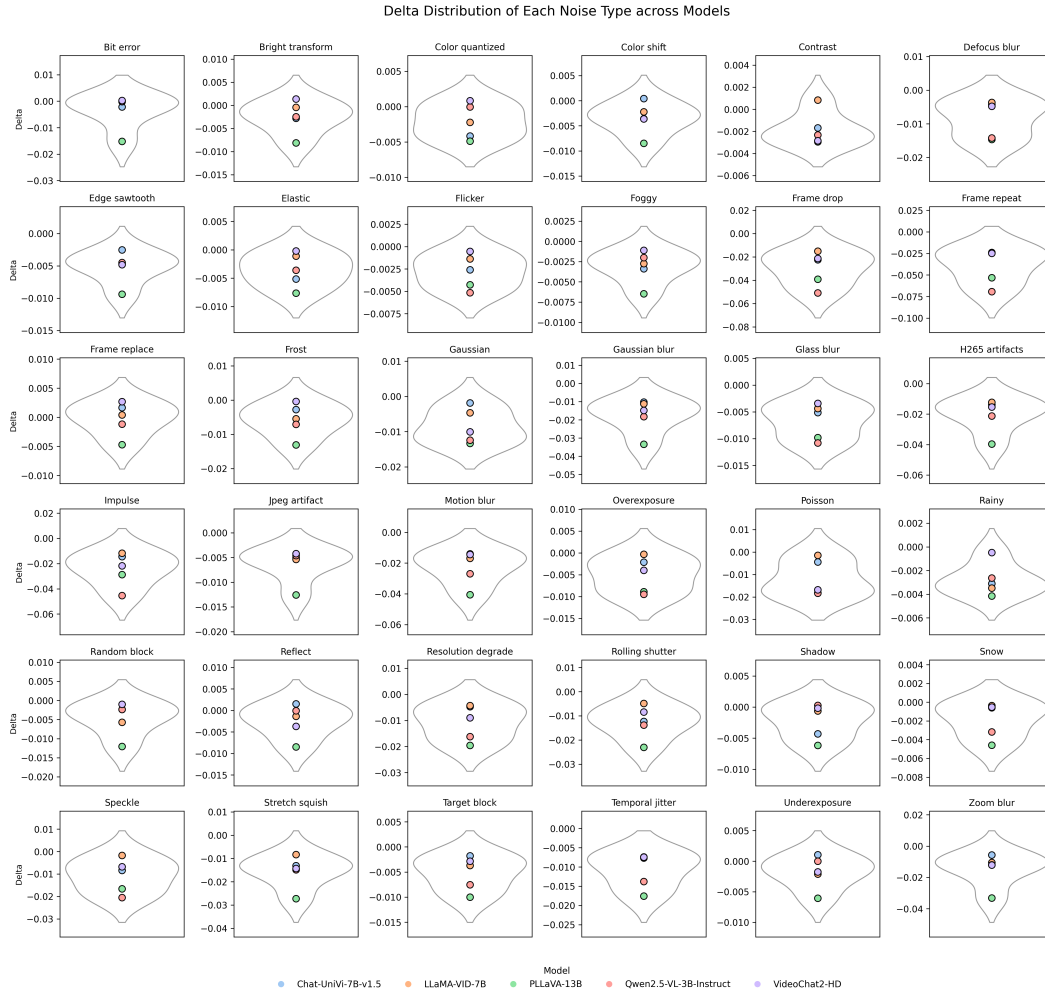


Figure 10: Delta Distribution of Each Noise Type across Models (SBERT Score).

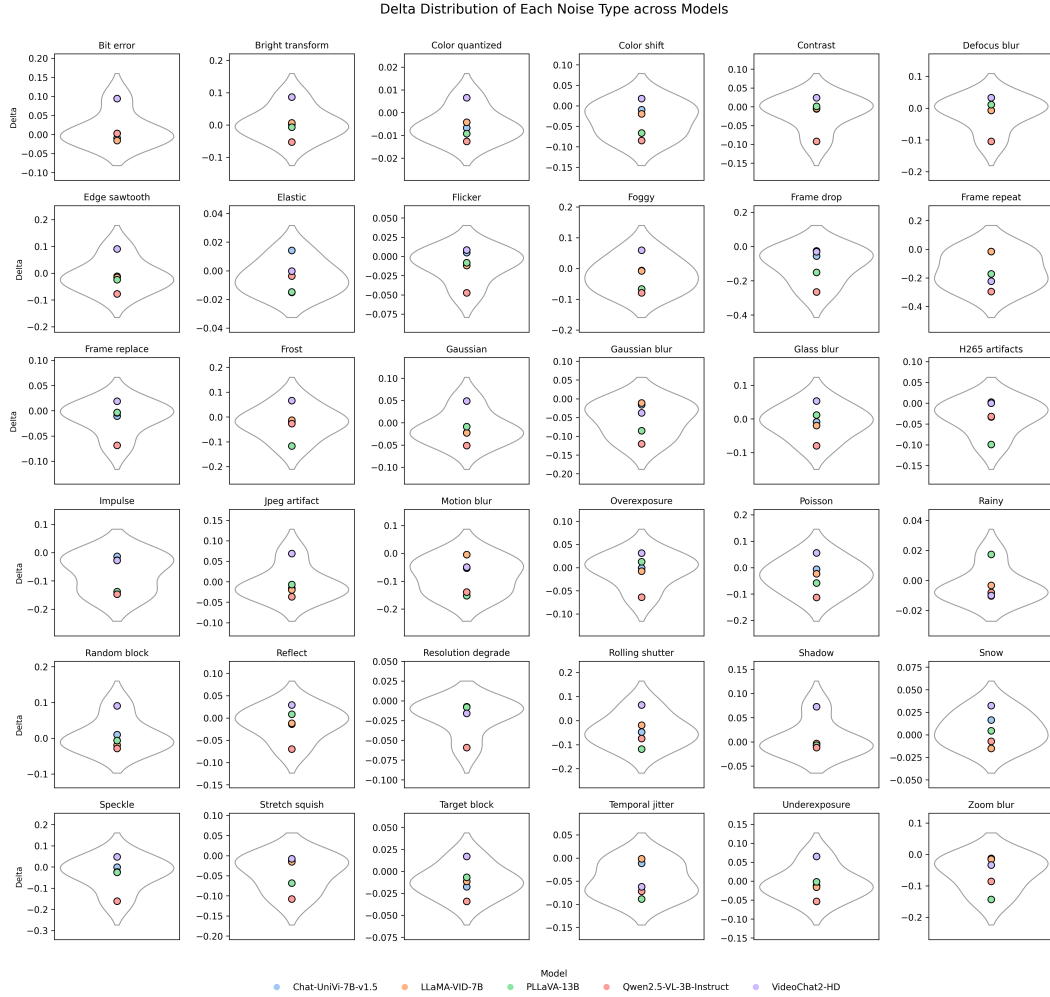


Figure 11: Delta Distribution of Each Noise Type across Models (Acc Score).

robustness is both metric- and model-dependent, revealing a number of nuanced behaviors. For blur-related corruptions such as *Gaussian blur* and *Motion blur*, all models suffer noticeable performance degradation, but the effect is most pronounced on the **GPT Score** metric, where PLLaVA-13B’s delta can reach as low as -0.35 , surpassing the drops observed in **Acc** (around -0.15). In contrast, *VideoChat2-HD* consistently demonstrates the strongest resilience, with deltas often 30% less negative than the next best model. This pattern suggests that while all models are sensitive to spatial blur, the severity is amplified under semantic or generative evaluation (**GPT Score**), and that *VideoChat2-HD* is inherently more robust to such distortions.

Temporal disruptions, such as *Frame drop* and *Frame repeat*, expose differences in temporal modeling capability. PLLaVA-13B is particularly vulnerable to these noises, as evidenced by its sharply lower delta values in both **Acc** and **GPT Score** metrics (-0.20 and -0.30 , respectively), while other models, though affected, can better preserve their performance. This suggests that certain architectures may lack effective mechanisms for handling discontinuities or redundancies in the video stream.

For compression artifacts like *H265 artifacts*, the gap between models is striking. On the **GPT Score** metric, LLaMA-VID shows a dramatic delta drop, approaching -0.65 , whereas *VideoChat2-HD* and *Qwen2.5-VL 3B* experience much milder degradation. This finding highlights a specific weakness of LLaMA-VID when dealing with heavy compression, possibly due to its handling of blocky or low-frequency corrupted frames.

Geometric distortions present another interesting case: under *Stretch squish* noise, PLLaVA 13B’s performance sharply deteriorates, especially on the **GPT Score** metric (delta near -0.40), while other models remain relatively stable. This outlier behavior points to a lack of geometric invariance in PLLaVA 13B, which could be addressed by targeted data augmentation or architectural changes.

Some noise types, such as *Impulse* and *Jpeg artifact*, yield much smaller and more uniform delta values across all models and metrics, indicating that these models are well-adapted to cope with minor pixel-level corruptions. Meanwhile, for extreme exposure conditions (*Overexposure* and *Underexposure*), all models exhibit similar levels of performance decrease, suggesting that this is a common weakness among current architectures, regardless of their design or training.

A.8 THE VISUALIZATION OF THE QA OVER DIFFERENT VISUAL NOISE

In this section, we present eight demonstration cases to illustrate the impact of different visual noises on model performance, as shown in Figure 12. For each noise category, we randomly selected one type of noise and applied it to a randomly chosen video. Both the original clean video and its noisy counterpart are shown for comparison. To simulate the typical process of frame extraction in video models, we uniformly sampled 8 frames from each video and applied noise to 90% of these frames. For consistency, all visualizations are generated using the Qwen2.5-VL 3B model.

These demonstrations clearly show that visual noise can severely degrade the model’s ability to discern fine details. On clean frames, the model is able to accurately recognize objects and count the number of instances within the video. However, when noise is present, such as in the color shift and motion blur examples, the model frequently fails to distinguish between similar objects or to count them correctly. Notably, in cases where noise severely degrades information (e.g., with resolution degradation), the model may produce hallucinated or irrelevant answers due to the lack of useful visual cues.

A.9 LICENSE OF THE USED ASSETS

MMBench-video is under CC BY-NC 4.0 license. Imagecorruption is under Apache-2.0 license.



(a) The visualization of color shift noise

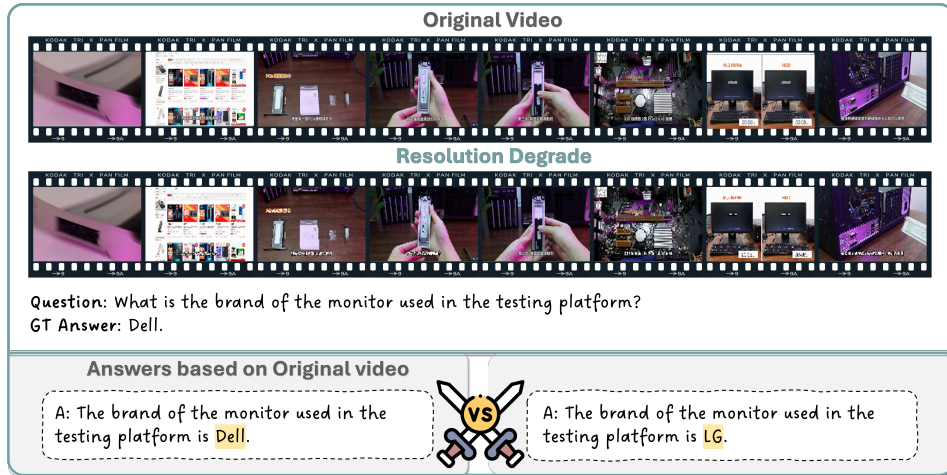


(b) The visualization of frost noise

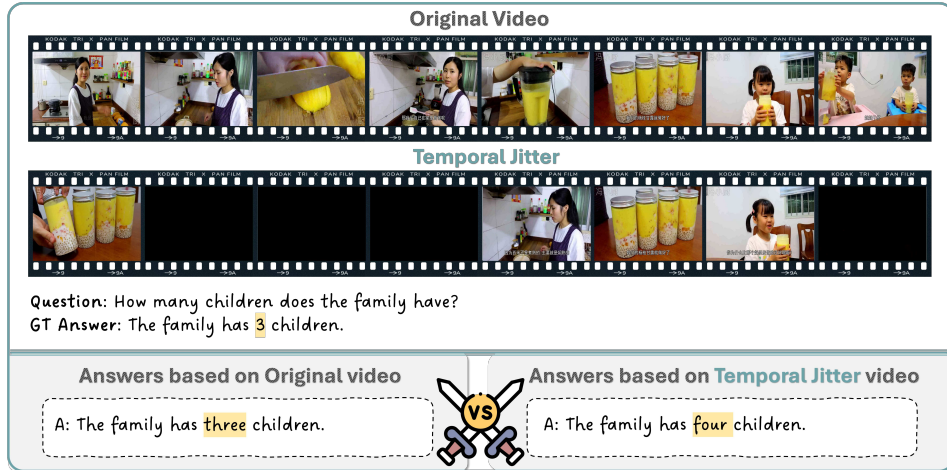


(c) The visualization of H.265 noise

Figure 12: Visualization of different types of noise (a)–(c)



(d) The visualization of resolution degrade noise

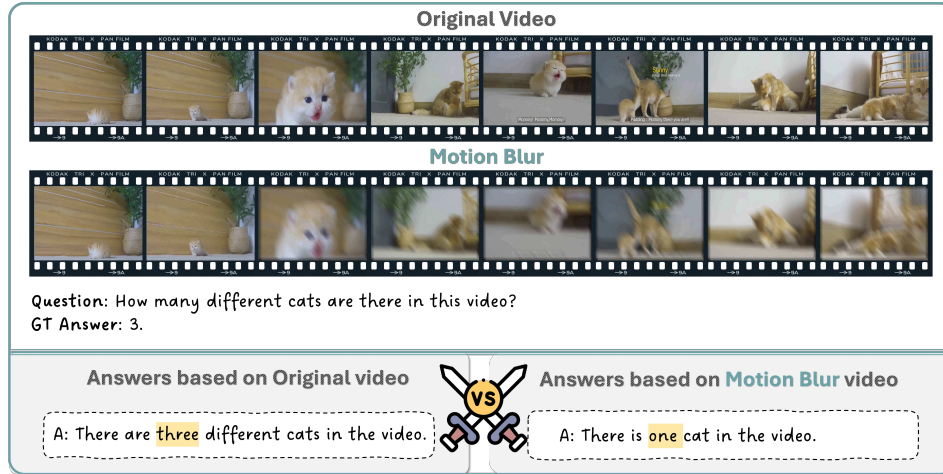


(e) The visualization of temporal jitter noise



(f) The visualization of impluse noise

Figure 12: Visualization of different types of noise (d)–(f)



(g) The visualization of motion blur



(h) The visualization of target block

Figure 12: Visualization of different types of noise (g)–(h)

Table 2: Chat-UniVi-v1.5 performance across different video types

Noise Type	A			B			C			D			E			F			G			H			I					
	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc			
Clean	0.976	0.416	0.167	0.100	0.357	-	1.143	0.423	0.846	0.827	0.346	0.278	0.827	0.344	0.300	1.250	0.420	0.847	0.479	0.340	0.938	0.462	0.333	1.147	0.524	0.812	0.811	0.340	1.000	
Gaussian	0.959	0.413	0.167	0.875	0.355	-	1.170	0.434	0.846	0.827	0.345	0.222	0.712	0.338	0.500	1.250	0.421	-	0.634	0.479	0.340	0.938	0.462	0.333	1.147	0.523	0.688	0.700	0.324	1.000
Impulse	0.732	0.393	0.167	0.750	0.298	-	1.152	0.430	0.846	0.827	0.346	0.222	0.712	0.334	0.500	1.250	0.415	-	0.667	0.474	0.250	1.000	0.461	0.333	1.039	0.526	0.756	0.678	0.316	1.000
Speckle	0.764	0.393	0.167	0.750	0.298	-	1.170	0.422	0.923	0.755	0.343	0.278	0.654	0.339	0.500	1.250	0.419	-	0.677	0.474	0.250	1.000	0.461	0.333	1.039	0.518	0.750	0.711	0.326	1.000
Poisson	0.902	0.408	0.167	0.750	0.306	-	1.277	0.432	0.923	0.918	0.353	0.222	0.865	0.338	0.500	1.250	0.415	-	0.753	0.473	0.312	0.925	0.457	0.333	1.047	0.521	0.750	0.622	0.310	1.000
Gaussian Blur	0.683	0.383	0.167	0.625	0.342	-	1.195	0.428	0.846	0.809	0.353	0.222	0.644	0.338	0.500	1.250	0.425	-	0.604	0.458	0.231	0.912	0.466	0.333	1.047	0.520	0.688	0.652	0.302	1.000
Motion Blur	0.674	0.389	0.167	0.625	0.342	-	1.170	0.428	0.846	0.809	0.353	0.222	0.644	0.338	0.500	1.250	0.425	-	0.604	0.458	0.231	0.912	0.466	0.333	1.047	0.520	0.688	0.652	0.302	1.000
Distortion	0.674	0.389	0.167	0.625	0.342	-	1.170	0.428	0.846	0.809	0.353	0.222	0.644	0.338	0.500	1.250	0.425	-	0.604	0.458	0.231	0.912	0.466	0.333	1.047	0.520	0.688	0.652	0.302	1.000
Zoom Blur	0.943	0.418	0.167	0.375	0.320	-	1.205	0.430	0.923	0.745	0.363	0.278	0.837	0.344	0.500	1.000	0.401	-	0.723	0.474	0.281	1.025	0.467	0.333	1.178	0.532	0.812	0.767	0.328	1.000
Frame Drop	0.512	0.353	0.167	0.375	0.314	-	1.134	0.427	0.846	0.817	0.338	0.222	0.733	0.348	0.500	1.250	0.437	-	0.668	0.470	0.312	0.800	0.459	0.333	1.097	0.529	0.750	0.622	0.306	1.000
Frame Replace	0.919	0.419	0.167	0.875	0.372	-	1.096	0.438	0.846	0.773	0.344	0.222	0.923	0.349	0.500	1.250	0.430	-	0.906	0.483	0.281	0.863	0.464	0.333	1.163	0.531	0.750	0.800	0.334	1.000
Frame Repeat	0.618	0.355	0.167	0.375	0.276	-	1.054	0.414	0.846	0.773	0.341	0.278	0.798	0.348	0.500	1.250	0.430	-	0.906	0.483	0.281	0.863	0.464	0.333	1.163	0.531	0.750	0.800	0.334	1.000
Temporal Jitter	0.756	0.390	0.167	0.625	0.325	-	1.161	0.430	0.846	0.827	0.342	0.222	0.644	0.338	0.500	1.250	0.406	-	0.775	0.473	0.312	0.938	0.473	0.333	1.163	0.531	0.750	0.800	0.334	1.000
Jpeg Artifact	0.829	0.397	0.167	0.625	0.353	-	1.214	0.439	0.846	0.900	0.351	0.222	0.837	0.352	0.500	1.250	0.401	-	0.804	0.477	0.281	1.075	0.481	0.333	1.155	0.535	0.750	0.733	0.320	1.000
Random Noise	0.772	0.397	0.167	0.250	0.308	-	1.170	0.430	0.846	0.827	0.355	0.278	0.731	0.334	0.500	1.250	0.469	-	0.600	0.462	0.281	0.675	0.432	0.333	1.162	0.532	0.688	0.678	0.320	1.000
H265 Artifacts	0.829	0.411	0.167	0.250	0.308	-	1.098	0.431	0.846	0.827	0.352	0.278	0.731	0.334	0.500	1.250	0.469	-	0.762	0.481	0.281	0.675	0.432	0.333	1.162	0.532	0.688	0.678	0.320	1.000
Random Block	0.789	0.401	0.167	0.750	0.343	-	1.205	0.425	0.846	0.755	0.347	0.222	0.798	0.344	0.500	1.250	0.425	-	0.783	0.482	0.281	1.012	0.468	0.333	1.152	0.540	0.812	0.689	0.327	1.000
Target Block	0.707	0.389	0.167	0.500	0.323	-	1.107	0.431	0.846	0.809	0.347	0.222	0.721	0.330	0.500	1.000	0.422	-	0.753	0.481	0.281	1.012	0.459	0.333	0.977	0.522	0.750	0.578	0.305	1.000
Rolling Shutter	0.821	0.405	0.167	0.500	0.319	-	1.161	0.431	0.846	0.809	0.357	0.222	0.721	0.336	0.500	1.000	0.422	-	0.706	0.479	0.281	0.988	0.477	0.333	1.132	0.528	0.750	0.744	0.315	1.000
Resolution Degradation	0.756	0.385	0.167	1.000	0.342	-	1.179	0.424	0.846	0.727	0.344	0.278	0.663	0.321	0.500	1.250	0.425	-	0.638	0.470	0.312	0.872	0.465	0.333	1.154	0.527	0.750	0.733	0.315	1.000
Edge Sawtooth	0.886	0.418	0.167	1.000	0.342	-	1.277	0.425	0.846	0.773	0.335	0.222	0.712	0.346	0.500	1.250	0.410	-	0.532	0.470	0.250	1.000	0.458	0.333	1.162	0.527	0.750	0.656	0.307	1.000
Color Shift	0.886	0.418	0.167	1.000	0.342	-	1.277	0.425	0.846	0.773	0.335	0.222	0.712	0.346	0.500	1.250	0.410	-	0.532	0.470	0.250	1.000	0.458	0.333	1.162	0.527	0.750	0.656	0.307	1.000
Flicker	0.886	0.418	0.167	1.000	0.342	-	1.277	0.425	0.846	0.773	0.335	0.222	0.712	0.346	0.500	1.250	0.410	-	0.532	0.470	0.250	1.000	0.458	0.333	1.162	0.527	0.750	0.656	0.307	1.000
Overexposure	0.894	0.440	0.167	1.000	0.355	-	1.215	0.439	0.846	0.782	0.349	0.222	0.760	0.346	0.500	1.000	0.440	-	0.855	0.486	0.344	0.963	0.464	0.333	0.984	0.518	0.750	0.700	0.320	1.000
Underexposure	0.894	0.440	0.167	1.000	0.355	-	1.215	0.439	0.846	0.782	0.349	0.222	0.760	0.346	0.500	1.000	0.440	-	0.855	0.486	0.344	0.963	0.464	0.333	0.984	0.518	0.750	0.700	0.320	1.000
Rainy	0.919	0.414	0.167	0.750	0.333	-	1.170	0.427	0.846	0.791	0.356	0.278	0.760	0.346	0.500	1.000	0.440	-	0.855	0.486	0.344	0.963	0.464	0.333	0.984	0.518	0.750	0.700	0.320	1.000
Foggy	0.870	0.417	0.167	0.750	0.333	-	1.170	0.427	0.846	0.791	0.356	0.278	0.760	0.346	0.500	1.000	0.440	-	0.855	0.486	0.344	0.963	0.464	0.333	0.984	0.518	0.750	0.700	0.320	1.000
Contrast	0.829	0.447	0.167	0.375	0.332	-	1.201	0.439	0.846	0.809	0.351	0.222	0.760	0.340	0.500	1.000	0.457	-	0.800	0.480	0.250	1.025	0.457	0.333	1.116	0.525	0.750	0.678	0.321	1.000
Color Shift	0.829	0.447	0.167	0.375	0.332	-	1.201	0.439	0.846	0.809	0.351	0.222	0.760	0.340	0.500	1.000	0.457	-	0.800	0.480	0.250	1.025	0.457	0.333	1.116	0.525	0.750	0.678	0.321	1.000
Flicker	0.829	0.447	0.167	0.375	0.332	-	1.201	0.439	0.846	0.809	0.351	0.222	0.760	0.340	0.500	1.000	0.457	-	0.800	0.480	0.250	1.025	0.457	0.333	1.116	0.525	0.750	0.678	0.321	1.000
Overexposure	0.829	0.447	0.167	0.375	0.332	-	1.201	0.439	0.846	0.809	0.351	0.222	0.760	0.340	0.500	1.000	0.457	-	0.800	0.480	0.250	1.025	0.457	0.333	1.116	0.525	0.750	0.678	0.321	1.000
Underexposure	0.829	0.447	0.167	0.375	0.332	-	1.201	0.439	0.846	0.809	0.351	0.222	0.760	0.340	0.500	1.000	0.457	-	0.800	0.480	0.250	1.025	0.457	0.333	1.116	0.525	0.750	0.678	0.321	1.000
Rainy	0.797	0.398	0.167	1.000	0.355	-	1.134	0.424	0.846	0.727	0.346	0.222	0.750	0.345	0.500	1.250	0.457	-	0.685	0.478	0.281	0.938	0.477	0.333	1.109	0.538	0.750	0.778	0.319	1.000
Foggy	0.927	0.410	0.167	1.000	0.343	-	1.161	0.442	0.846	0.773	0.344	0.222	0.750	0.345	0.500	1.250	0.457	-	0.685	0.478	0.281	0.938	0.477	0.333	1.109	0.538	0.750	0.778	0.319	1.000
Contrast	0.837	0.419	0.167	1.250	0.314	-	1.054	0.430	0.846	0.773	0.333	0.222	0.817	0.347	0.500	1.000	0.457	-	0.804	0.485	0.312	0.875	0.461	0.333	1.062	0.530	0.812	0.611	0.314	1.000
Shadow	0.837	0.419	0.167	1.250	0.314	-	1.054	0.430	0.846	0.773	0.333	0.222	0.817	0.347	0.500	1.000	0.457	-	0.804	0.485	0.312	0.875	0.461	0.333	1.062	0.530	0.812	0.611	0.314	1.000

Noise Type	K			L			M			N			O			P			Q			R			S			T			
	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	
Clean	0.453	0.333	0.667	0.364	0.395	-	1.000	0.721	0.506	0.357	0.758	0.321	0.600	0.538	0.342	0.000	0.817	0.406	0.263	0.120	0.412	0.000	0.741	0.544	0.400	1.311	0.475	0.371	0.897	0.430	0.461
Gaussian	0.511	0.340	0.667	1.273	0.398	-	1.000	0.782	0.511	0.357	0.811	0.321	0.600	0.615	0.315	0.000	0.780	0.412	0.158	0.136	0.405	0.000	0.741	0.544	0.400	1.311	0.475	0.371			

Note: A: Advertisements, B: Algorithm & Models, C: Autos & Vehicles, D: Business & Industrial, E: Computers & Electronics, F: Fairy Tale, G: Films & TV Shows, H: Finance, I: Food & Drink, J: Games, K: Humor, L: Instruction Video (how to), M: Knowledge, N: News, O: Others, P: People, Q: Pets & Animals, R: Science, S: Sports, T: Overall

1512
1513
1514
1515
1516
1517
1518
1519
1520
1522
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565

1512
1513
1514
1515
1516
1517
1518
1519
1520
1522
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565

1512
1513
1514
1515
1516
1517
1518
1519
1520
1522
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565

Table 4: GPT-4o performance across different video types

Noise Type	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Clean	1.577	1.508	1.617	1.715	1.406	1.955	1.523	1.385	1.855	1.597	1.722	1.865	1.508	1.000	1.000	1.250	1.000	1.000	1.000	1.000
Gaussian	1.545	1.512	1.607	1.720	1.436	1.795	1.496	1.462	1.709	1.597	1.722	1.865	1.508	1.000	1.000	1.250	1.000	1.000	1.000	1.000
Impulse	1.244	1.469	1.617	1.215	1.427	1.348	1.475	1.462	1.364	1.588	1.611	1.683	1.519	1.504	1.000	1.250	1.000	1.000	1.000	1.000
Speckle	1.512	1.509	1.617	1.200	1.400	1.768	1.523	1.462	1.336	1.588	1.611	1.683	1.519	1.504	1.000	1.250	1.000	1.000	1.000	1.000
Poisson	1.439	1.495	1.617	1.200	1.400	1.768	1.523	1.462	1.336	1.588	1.611	1.683	1.519	1.504	1.000	1.250	1.000	1.000	1.000	1.000
Gaussian Blur	1.728	1.474	1.333	1.200	1.427	1.661	1.505	1.538	1.418	1.577	1.611	1.606	1.511	1.500	1.000	1.250	1.000	1.000	1.000	1.000
Motion Blur	1.517	1.469	1.333	1.200	1.427	1.661	1.505	1.538	1.418	1.577	1.611	1.606	1.511	1.500	1.000	1.250	1.000	1.000	1.000	1.000
Dark Noise	1.517	1.469	1.333	1.200	1.427	1.661	1.505	1.538	1.418	1.577	1.611	1.606	1.511	1.500	1.000	1.250	1.000	1.000	1.000	1.000
Glitch	1.447	1.495	1.333	1.200	1.427	1.661	1.505	1.538	1.418	1.577	1.611	1.606	1.511	1.500	1.000	1.250	1.000	1.000	1.000	1.000
Zoom Blur	1.512	1.495	1.333	1.200	1.427	1.661	1.505	1.538	1.418	1.577	1.611	1.606	1.511	1.500	1.000	1.250	1.000	1.000	1.000	1.000
Frame Drop	1.967	1.443	1.617	1.125	1.354	1.321	1.439	1.308	1.287	1.523	1.556	1.337	1.478	1.500	1.250	1.000	1.000	1.000	1.000	1.000
Frame Replace	1.602	1.524	1.500	1.200	1.430	1.723	1.511	1.538	1.664	1.577	1.611	1.981	1.510	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Temporal Jitter	1.520	1.494	1.333	1.200	1.427	1.500	1.495	1.462	1.391	1.584	1.611	1.683	1.519	1.504	1.000	1.250	1.000	1.000	1.000	1.000
Jpeg Artifact	1.529	1.506	1.617	1.200	1.427	1.714	1.505	1.538	1.382	1.572	1.611	1.606	1.511	1.500	1.000	1.250	1.000	1.000	1.000	1.000
Color Shift	1.489	1.454	1.333	1.200	1.427	1.661	1.498	1.462	1.391	1.584	1.611	1.683	1.519	1.504	1.000	1.250	1.000	1.000	1.000	1.000
Random Block	1.472	1.500	1.617	1.200	1.427	1.661	1.498	1.462	1.391	1.584	1.611	1.683	1.519	1.504	1.000	1.250	1.000	1.000	1.000	1.000
Target Block	1.528	1.509	1.617	1.200	1.427	1.661	1.498	1.462	1.391	1.584	1.611	1.683	1.519	1.504	1.000	1.250	1.000	1.000	1.000	1.000
Rolling Shutter	1.488	1.490	1.333	1.200	1.427	1.571	1.482	1.462	1.391	1.584	1.611	1.683	1.519	1.504	1.000	1.250	1.000	1.000	1.000	1.000
Resolution Degradation	1.463	1.476	1.500	1.200	1.430	1.509	1.470	1.462	1.207	1.552	1.611	1.433	1.493	1.500	1.000	1.250	1.000	1.000	1.000	1.000
Stretch Squish	1.374	1.477	1.500	1.200	1.430	1.402	1.469	1.385	1.200	1.552	1.611	1.433	1.493	1.500	1.000	1.250	1.000	1.000	1.000	1.000
Edge Sawtooth	1.553	1.516	1.500	1.200	1.430	1.705	1.508	1.538	1.618	1.585	1.611	1.865	1.522	1.000	1.000	1.250	1.000	1.000	1.000	1.000
Color Shift	1.520	1.494	1.333	1.200	1.427	1.661	1.498	1.462	1.391	1.584	1.611	1.683	1.519	1.504	1.000	1.250	1.000	1.000	1.000	1.000
Contrast	1.431	1.510	1.617	1.200	1.427	1.661	1.498	1.462	1.391	1.584	1.611	1.683	1.519	1.504	1.000	1.250	1.000	1.000	1.000	1.000
Bright Transform	1.569	1.510	1.617	1.200	1.427	1.732	1.527	1.495	1.421	1.518	1.589	1.678	1.904	1.511	1.000	1.250	1.000	1.000	1.000	1.000
Color Shift	1.650	1.522	1.500	1.200	1.430	1.821	1.524	1.462	1.207	1.552	1.611	1.827	1.511	1.000	1.250	1.000	1.000	1.000	1.000	1.000
Flicker	1.577	1.506	1.333	1.200	1.427	1.812	1.524	1.462	1.207	1.552	1.611	1.827	1.511	1.000	1.250	1.000	1.000	1.000	1.000	1.000
Overexposure	1.462	1.523	1.500	1.200	1.430	1.795	1.502	1.538	1.645	1.590	1.778	1.888	1.524	1.000	1.000	1.250	1.000	1.000	1.000	1.000
Underexposure	1.486	1.516	1.500	1.200	1.430	1.752	1.529	1.538	1.673	1.602	1.656	1.897	1.522	1.000	1.000	1.250	1.000	1.000	1.000	1.000
Rainy	1.675	1.506	1.333	1.200	1.427	1.795	1.521	1.538	1.609	1.576	1.722	1.817	1.508	1.000	1.250	1.000	1.000	1.000	1.000	1.000
Peggy	1.488	1.508	1.617	1.200	1.427	1.705	1.513	1.385	1.691	1.585	1.667	1.808	1.527	1.000	1.250	1.000	1.000	1.000	1.000	1.000
Reflection	1.537	1.514	1.617	1.200	1.397	1.812	1.531	1.308	1.600	1.576	1.500	1.827	1.528	1.000	1.250	1.000	1.000	1.000	1.000	1.000
Shadow	1.642	1.508	1.500	1.200	1.434	1.750	1.515	1.538	1.827	1.604	1.667	1.750	1.531	1.000	1.250	1.000	1.000	1.000	1.000	1.000

Noise Type	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Clean	1.350	1.468	1.667	1.242	1.590	1.769	1.574	1.786	1.905	1.518	1.600	1.231	1.400	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Gaussian	1.007	1.366	1.667	1.136	1.427	1.667	1.565	1.571	1.786	1.863	1.518	1.600	1.231	1.400	1.000	1.000	1.000	1.000	1.000	1.000
Impulse	1.759	1.399	1.000	1.833	1.550	1.000	1.367	1.550	1.714	1.568	1.491	1.600	1.231	1.400	1.000	1.000	1.000	1.000	1.000	1.000
Speckle	1.109	1.434	1.667	1.136	1.571	1.000	1.612	1.560	1.786	1.958	1.537	1.000	1.077	1.462	1.000	1.000	1.000	1.000	1.000	1.000
Poisson	1.927	1.428	1.333	1.970	1.584	1.000	1.571	1.561	1.786	1.905	1.535	1.000	1.385	1.449	1.000	1.000	1.000	1.000	1.000	1.000
Gaussian Blur	1.876	1.424	1.667	1.727	1.536	1.000	1.313	1.528	1.786	1.632	1.492	1.600	1.231	1.443	1.000	1.000	1.000	1.000	1.000	1.000
Motion Blur	1.078	1.414	1.667	1.127	1.525	1.000	1.667	1.574	1.786	1.905	1.537	1.000	1.077	1.462	1.000	1.000	1.000	1.000	1.000	1.000
Dark Noise	1.209	1.417	1.000	1.076	1.595	1.000	1.667	1.574	1.786	1.905	1.537	1.000	1.077	1.462	1.000	1.000	1.000	1.000	1.000	1.000
Glitch	1.044	1.433	1.000	1.182	1.578	1.000	1.667	1.574	1.786	1.905	1.537	1.000	1.077	1.462	1.000	1.000	1.000	1.000	1.000	1.000
Zoom Blur	1.745	1.401	1.000	1.682	1.535	1.000	1.667	1.574	1.786	1.905	1.537	1.000	1.077	1.462	1.000	1.000	1.000	1.000	1.000	1.000
Frame Drop	1.102	1.443	1.333	1.218	1.575	1.000	1.687	1.575	1.786	1.874	1.529	1.000	1.077	1.462	1.000	1.000	1.000	1.000	1.000	1.000
Frame Replace	1.672	1.378	1.333	1.591	1.477	1.000	1.340	1.530	1.714	1.558	1.449	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Temporal Jitter	1.190	1.432	1.667	1.242	1.598	1.000	1.667	1.565	1.571	1.786	1.863	1.518	1.600	1.231	1.400	1.000	1.000	1.000	1.000	1.000
Jpeg Artifact	1.978	1.410	1.667	1.121	1.582	1.000	1.667	1.574	1.786	1.905	1.537	1.000	1.077	1.462	1.000	1.000	1.000	1.000	1.000	1.000
Color Shift	1.068	1.409	1.667	1.139	1.565	1.000	1.667	1.574	1.786	1.905	1.537	1.000	1.077	1.462	1.000	1.000	1.000	1.000	1.000	1.000
H265 Artifacts	1.709	1.409	1.667	1.139	1.565	1.000	1.667	1.574	1.786	1.905	1.537	1.000	1.077	1.462	1.000	1.000	1.000	1.000	1.000	1.000
Random Block	1.109	1.442	1.667	1.030	1.568	1.000	1.626	1.565	1.786	1.800	1.506	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Target Block	1.964	1.430	1.333	1.227	1.594	1.000	1.769	1.577	1.786	1.832	1.536	1.000	1.154	1.437	1.000	1.000	1.000	1.000	1.000	1.000
Rolling Shutter	1.007	1.441	1.667	1.190	1.554	1.000	1.667	1.531	1.578	1.786	1.905	1.511	1.600	1.231	1.402	1.000	1.000	1.000	1.000	1.000
Resolution Degradation	1.044	1.412	1.667	1.203	1.546	1.000	1.388	1.486	1.571	1.653	1.505	1.000	1.154	1.431	1.000	1.000	1.000	1.000	1.000	1.000
Stretch Squish	1.869	1.420	1.333	1.818	1.505	1.000	1.524	1.557	1.786	1.495	1.474	1.000	1.154	1.431	1.000	1.000	1.000	1.000	1.000	1.000
Edge Sawtooth	1.219	1.458	1.667	1.182	1.581	1.000	1.762	1.588	1.920	1.800	1.516	1.000	1.231	1.446	1.000	1.000	1.000	1.000	1.000	1.000
Color Shift	1.166	1.428	1.667	1.182	1.581	1.000	1.762	1.588	1.920	1.800	1.516	1.000	1.231	1.446	1.000	1.000	1.000	1.000	1.000	1.000
Contrast	1.015	1.399	1.333	1.142	1.596	1.000	1.667	1.574	1.786	1.905	1.537	1.000	1.077	1.462	1.000	1.000	1.000	1.000	1.000	1.000
Bright Transform	1.241	1.451	1.000	1.227	1.611	1.000	1.667	1												

Table 5: LLAMA-VID performance across different video types

Noise Type	A			B			C			D			E			F			G			H			I			J																												
	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc																										
Clean	0.992	0.417	0.167	0.625	0.341	1.188	0.440	0.923	0.927	0.363	0.728	0.798	0.359	0.500	0.500	0.478	0.770	0.481	0.281	0.038	0.486	0.333	0.264	0.544	0.875	0.756	0.326	1.000	0.992	0.417	0.167	0.625	0.341	1.188	0.440	0.923	0.927	0.363	0.728	0.798	0.359	0.500	0.500	0.478	0.770	0.481	0.281	0.038	0.486	0.333	0.264	0.544	0.875	0.756	0.326	1.000
Gaussian	0.854	0.401	0.167	0.625	0.341	1.134	0.443	0.923	0.900	0.367	0.722	0.827	0.348	0.500	0.500	0.486	0.728	0.481	0.281	0.038	0.472	0.333	0.264	0.544	0.875	0.756	0.326	1.000	0.854	0.401	0.167	0.625	0.341	1.134	0.443	0.923	0.900	0.367	0.722	0.827	0.348	0.500	0.500	0.486	0.728	0.481	0.281	0.038	0.472	0.333	0.264	0.544	0.875	0.756	0.326	1.000
Impulse	0.772	0.397	0.167	1.000	0.341	1.134	0.442	0.923	0.764	0.354	0.722	0.692	0.337	0.500	0.500	0.482	0.651	0.479	0.281	0.038	0.471	0.333	0.185	0.525	0.750	0.578	0.326	1.000	0.772	0.397	0.167	1.000	0.341	1.134	0.442	0.923	0.764	0.354	0.722	0.692	0.337	0.500	0.500	0.482	0.651	0.479	0.281	0.038	0.471	0.333	0.185	0.525	0.750	0.578	0.326	1.000
Speckle	0.837	0.407	0.167	1.125	0.345	1.134	0.448	0.923	0.909	0.367	0.722	0.808	0.356	0.500	0.500	0.496	0.753	0.486	0.281	0.038	0.471	0.333	0.185	0.525	0.750	0.578	0.326	1.000	0.837	0.407	0.167	1.125	0.345	1.134	0.448	0.923	0.909	0.367	0.722	0.808	0.356	0.500	0.500	0.496	0.753	0.486	0.281	0.038	0.471	0.333	0.185	0.525	0.750	0.578	0.326	1.000
Poisson	0.780	0.401	0.167	0.750	0.337	1.152	0.446	0.923	0.764	0.355	0.722	0.808	0.356	0.500	0.500	0.482	0.736	0.477	0.281	0.038	0.471	0.333	0.185	0.525	0.750	0.578	0.326	1.000	0.780	0.401	0.167	0.750	0.337	1.152	0.446	0.923	0.764	0.355	0.722	0.808	0.356	0.500	0.500	0.482	0.736	0.477	0.281	0.038	0.471	0.333	0.185	0.525	0.750	0.578	0.326	1.000
Gaussian Blur	0.675	0.383	0.133	1.000	0.331	1.071	0.439	0.923	0.706	0.353	0.722	0.774	0.347	0.500	0.500	0.479	0.624	0.471	0.281	0.038	0.470	0.333	0.185	0.525	0.750	0.578	0.326	1.000	0.675	0.383	0.133	1.000	0.331	1.071	0.439	0.923	0.706	0.353	0.722	0.774	0.347	0.500	0.500	0.479	0.624	0.471	0.281	0.038	0.470	0.333	0.185	0.525	0.750	0.578	0.326	1.000
Motion Blur	0.910	0.415	0.167	0.625	0.341	1.125	0.440	0.923	0.936	0.367	0.728	0.673	0.347	0.500	0.500	0.478	0.745	0.477	0.281	0.038	0.485	0.333	0.203	0.540	0.875	0.756	0.326	1.000	0.910	0.415	0.167	0.625	0.341	1.125	0.440	0.923	0.936	0.367	0.728	0.673	0.347	0.500	0.500	0.478	0.745	0.477	0.281	0.038	0.485	0.333	0.203	0.540	0.875	0.756	0.326	1.000
Distortion	0.829	0.405	0.167	0.750	0.321	1.250	0.443	0.923	0.918	0.365	0.722	0.702	0.345	0.500	0.500	0.486	0.723	0.482	0.281	0.038	0.479	0.333	0.248	0.544	0.875	0.678	0.326	1.000	0.829	0.405	0.167	0.750	0.321	1.250	0.443	0.923	0.918	0.365	0.722	0.702	0.345	0.500	0.500	0.486	0.723	0.482	0.281	0.038	0.479	0.333	0.248	0.544	0.875	0.678	0.326	1.000
Zoom Blur	0.837	0.399	0.167	1.125	0.334	1.205	0.447	0.923	0.682	0.351	0.722	0.606	0.338	0.500	0.500	0.486	0.638	0.462	0.281	0.038	0.471	0.333	0.199	0.520	0.750	0.578	0.326	1.000	0.837	0.399	0.167	1.125	0.334	1.205	0.447	0.923	0.682	0.351	0.722	0.606	0.338	0.500	0.500	0.486	0.638	0.462	0.281	0.038	0.471	0.333	0.199	0.520	0.750	0.578	0.326	1.000
Frame Drop	0.740	0.395	0.167	0.500	0.317	1.018	0.434	0.900	0.901	0.358	0.722	0.712	0.345	0.500	0.500	0.491	0.809	0.484	0.281	0.038	0.471	0.333	0.124	0.531	0.875	0.756	0.326	1.000	0.740	0.395	0.167	0.500	0.317	1.018	0.434	0.900	0.901	0.358	0.722	0.712	0.345	0.500	0.500	0.491	0.809	0.484	0.281	0.038	0.471	0.333	0.124	0.531	0.875	0.756	0.326	1.000
Frame Replace	0.993	0.364	0.167	0.750	0.333	0.938	0.412	0.923	0.900	0.360	0.728	0.942	0.365	0.500	0.500	0.497	0.809	0.484	0.281	0.038	0.471	0.333	0.124	0.531	0.875	0.756	0.326	1.000	0.993	0.364	0.167	0.750	0.333	0.938	0.412	0.923	0.900	0.360	0.728	0.942	0.365	0.500	0.500	0.497	0.809	0.484	0.281	0.038	0.471	0.333	0.124	0.531	0.875	0.756	0.326	1.000
Temporal Jitter	0.821	0.401	0.167	0.750	0.332	1.036	0.437	0.923	0.900	0.368	0.722	0.740	0.355	0.500	0.500	0.493	0.719	0.478	0.281	0.038	0.471	0.333	0.163	0.540	0.812	0.756	0.326	1.000	0.821	0.401	0.167	0.750	0.332	1.036	0.437	0.923	0.900	0.368	0.722	0.740	0.355	0.500	0.500	0.493	0.719	0.478	0.281	0.038	0.471	0.333	0.163	0.540	0.812	0.756	0.326	1.000
Jpeg Artifact	0.846	0.406	0.167	0.750	0.348	1.134	0.441	0.923	0.945	0.360	0.728	0.913	0.357	0.500	0.500	0.496	0.762	0.481	0.281	0.038	0.472	0.333	0.225	0.546	0.875	0.667	0.337	1.000	0.846	0.406	0.167	0.750	0.348	1.134	0.441	0.923	0.945	0.360	0.728	0.913	0.357	0.500	0.500	0.496	0.762	0.481	0.281	0.038	0.472	0.333	0.225	0.546	0.875	0.667	0.337	1.000
H265 Artifacts	0.740	0.386	0.133	0.875	0.342	1.045	0.444	0.923	0.882	0.354	0.722	0.802	0.339	0.500	0.500	0.484	0.664	0.466	0.281	0.038	0.471	0.333	0.109	0.525	0.812	0.633	0.318	1.000	0.740	0.386	0.133	0.875	0.342	1.045	0.444	0.923	0.882	0.354	0.722	0.802	0.339	0.500	0.500	0.484	0.664	0.466	0.281	0.038	0.471	0.333	0.109	0.525	0.812	0.633	0.318	1.000
Random Block	0.691	0.403	0.167	0.625	0.342	1.034	0.444	0.923	0.918	0.361	0.722	0.894	0.364	0.500	0.500	0.489	0.791	0.485	0.281	0.038	0.471	0.333	0.132	0.531	0.812	0.667	0.337	1.000	0.691	0.403	0.167	0.625	0.342	1.034	0.444	0.923	0.918	0.361	0.722	0.894	0.364	0.500	0.500	0.489	0.791	0.485	0.281	0.038	0.471	0.333	0.132	0.531	0.812	0.667	0.337	1.000
Target Block	0.813	0.407	0.167	0.625	0.334	1.098	0.440	0.923	0.936	0.366	0.722	0.740	0.364	0.500	0.500	0.487	0.689	0.478	0.281	0.038	0.471	0.333	0.171	0.532	0.812	0.667	0.337	1.000	0.813	0.407	0.167	0.625	0.334	1.098	0.440	0.923	0.936	0.366	0.722	0.740	0.364	0.500	0.500	0.487	0.689	0.478	0.281	0.038	0.471	0.333	0.171	0.532	0.812	0.667	0.337	1.000
Rolling Shutter	0.602	0.380	0.167	1.500	0.370	1.170	0.436	0.923	0.936	0.366	0.722	0.740	0.364	0.500	0.500	0.487	0.689	0.478	0.281	0.038	0.471	0.333	0.171	0.532	0.812	0.667	0.337	1.000	0.602	0.380	0.167	1.500	0.370	1.170	0.436	0.923	0.936	0.366	0.722	0.740	0.364	0.500	0.500	0.487	0.689	0.478	0.281	0.038	0.471	0.333	0.171	0.532	0.812	0.667	0.337	1.000
Resolution Degrad	0.886	0.413	0.167	0.750	0.333	1.188	0.447	0.923	0.791	0.358	0.722	0.692	0.341	0.500	0.500	0.476	0.766	0.480	0.281	0.038	0.471	0.333	0.194	0.544	0.875	0.611	0.328	1.000	0.886	0.413	0.167	0.750	0.333	1.188	0.447	0.923	0.791	0.358	0.722	0.692	0.341	0.500	0.500	0.476	0.766	0.480	0.281	0.038	0.471	0.333	0.194	0.544	0.875	0.611	0.328	1.000
Stretch Squish	0.813	0.401	0.167	0.750	0.333	1.152	0.436	0.923	0.782	0.353	0.722	0.654	0.346	0.500	0.500	0.503	0.766	0.480	0.281	0.038	0.471	0.333	0.194	0.544	0.875	0.611	0.328	1.000	0.813	0.401	0.167	0.750	0.333	1.152	0.436	0.923	0.782	0.353	0.722	0.654	0.346	0.500	0.500	0.503	0.766	0.480	0.281	0.038	0.471	0.333	0.194	0.544	0.875	0.611	0.328	1.000
Edge Sawtooth	0.919	0.413	0.167	1.125	0.352	1.152	0.442	0.923	0.918	0.363	0.722	0.769	0.352	0.500	0.500	0.494	0.723	0.483	0.281	0.038	0.471	0.333	0.163	0.543	0.875	0.678	0.334	1.000	0.919	0.413	0.167	1.125	0.352	1.152	0.442	0.923	0.918	0.363	0.722	0.769	0.352	0.500	0.500	0.494	0.723	0.483	0.281	0.038	0.471	0.333	0.163	0.543	0.875	0.678	0.334	1.000
Color Shift	0.972	0.417	0.167	0.625	0.350	1.134	0.444	0.923	0.873	0.365	0.722	0.885	0.357	0.500	0.500	0.492	0.787	0.481	0.281	0.038	0.471	0.333	0.178	0.543	0.875	0.678	0.334	1.000	0.972	0.417	0.167	0.625	0.350	1.134	0.444	0.923	0.873	0.365	0.722	0.885	0.357	0.500	0.500	0.492	0.787	0.481	0.281	0.038	0.471	0.333	0.178	0.543	0.875	0.678	0.334	1.000
Crop	0.911	0.417	0.167	0.625	0.330	1.230	0.449	0.92																																																

Note: A: Advertisements, B: Games, K: Humor, L: Instruction Video (how to), M: Knowledge, N: News, O: Others, P: People, Q: Pets & Animals, R: Science, S: Sports, T: Overall

Table 6: PLLaVA 13B performance across different video types

Noise Type	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T												
	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc											
Clean	0.984	0.414	0.167	0.125	0.345	-	0.188	0.435	0.385	0.955	0.743	0.000	0.096	0.371	0.000	0.000	0.000	0.279	-	0.800	0.469	0.188	0.163	0.470	0.500	0.961	0.490	0.250	0.756	0.326	1.000	
Gaussian	0.675	0.394	0.040	0.075	0.305	-	0.973	0.425	0.385	0.809	0.332	0.000	0.096	0.371	0.000	0.000	0.000	0.279	-	0.719	0.455	0.156	0.121	0.475	0.500	0.961	0.490	0.250	0.756	0.326	1.000	
Impulse	0.504	0.374	0.000	0.750	0.309	-	0.875	0.402	0.385	0.809	0.332	0.000	0.096	0.371	0.000	0.000	0.000	0.279	-	0.536	0.414	0.062	0.108	0.444	0.444	0.500	0.744	0.481	0.188	0.722	0.319	1.000
Speckle	0.715	0.371	0.167	0.125	0.299	-	1.036	0.433	0.385	0.809	0.332	0.000	0.096	0.371	0.000	0.000	0.000	0.279	-	0.626	0.445	0.094	0.136	0.482	0.444	0.500	0.822	0.481	0.188	0.644	0.308	1.000
Poisson	0.724	0.349	0.167	0.375	0.326	-	1.000	0.430	0.385	0.809	0.332	0.000	0.096	0.371	0.000	0.000	0.000	0.279	-	0.604	0.451	0.125	0.121	0.470	0.500	0.721	0.478	0.188	0.567	0.312	1.000	
Gaussian Blur	0.520	0.349	0.167	0.000	0.314	-	0.768	0.404	0.385	0.627	0.297	0.000	0.067	0.342	0.000	0.000	0.000	0.287	-	0.485	0.427	0.094	0.075	0.449	0.333	0.690	0.463	0.125	0.589	0.301	1.000	
Motion Blur	0.882	0.440	0.167	0.750	0.263	-	0.643	0.384	0.385	0.609	0.345	0.000	0.096	0.371	0.000	0.000	0.000	0.279	-	0.428	0.428	0.094	0.075	0.449	0.333	0.690	0.463	0.125	0.589	0.301	1.000	
Distortion	0.882	0.440	0.167	0.750	0.263	-	0.643	0.384	0.385	0.609	0.345	0.000	0.096	0.371	0.000	0.000	0.000	0.279	-	0.428	0.428	0.094	0.075	0.449	0.333	0.690	0.463	0.125	0.589	0.301	1.000	
Zoom Blur	0.813	0.399	0.167	0.375	0.320	-	0.938	0.419	0.385	1.009	0.345	0.000	0.096	0.371	0.000	0.000	0.000	0.279	-	0.732	0.450	0.156	0.117	0.482	0.500	0.845	0.485	0.250	0.633	0.316	1.000	
Frame Drop	0.585	0.361	0.167	0.500	0.299	-	0.866	0.397	0.385	0.609	0.320	0.000	0.096	0.371	0.000	0.000	0.000	0.279	-	0.523	0.429	0.031	0.163	0.462	0.333	0.628	0.457	0.125	0.567	0.305	1.000	
Frame Replace	0.374	0.339	0.000	1.000	0.323	-	0.875	0.407	0.385	0.555	0.320	0.000	0.065	0.333	0.000	0.000	0.000	0.286	-	0.455	0.416	0.188	0.187	0.432	0.167	0.589	0.455	0.188	0.600	0.295	1.000	
Temporal Jitter	0.928	0.410	0.167	0.250	0.299	-	1.071	0.431	0.385	0.855	0.336	0.000	0.096	0.371	0.000	0.000	0.000	0.271	-	0.898	0.473	0.188	0.288	0.472	0.500	1.000	0.491	0.250	0.722	0.312	1.000	
Frame Repeat	0.202	0.310	0.000	0.875	0.303	-	0.991	0.434	0.231	0.682	0.334	0.000	0.064	0.324	0.000	0.000	0.000	0.294	-	0.438	0.402	0.156	0.081	0.433	0.333	0.519	0.444	0.250	0.533	0.281	1.000	
Temporal Shift	0.561	0.360	0.167	0.750	0.308	-	1.062	0.431	0.462	0.818	0.348	0.000	0.070	0.356	0.000	0.000	0.000	0.269	-	0.638	0.448	0.121	0.107	0.452	0.333	0.706	0.484	0.188	0.700	0.314	1.000	
Jpeg Artifact	0.683	0.390	0.167	0.750	0.308	-	1.062	0.431	0.462	0.818	0.348	0.000	0.070	0.356	0.000	0.000	0.000	0.269	-	0.638	0.448	0.121	0.107	0.452	0.333	0.706	0.484	0.188	0.700	0.314	1.000	
Color Shift	0.472	0.360	0.167	0.625	0.295	-	0.998	0.386	0.231	0.673	0.317	0.000	0.065	0.304	0.000	0.000	0.000	0.319	-	0.181	0.420	0.031	0.073	0.451	0.333	0.535	0.453	0.188	0.489	0.206	1.000	
H265 Artifacts	0.634	0.363	0.167	0.500	0.322	-	1.045	0.423	0.385	0.855	0.347	0.000	0.096	0.371	0.000	0.000	0.000	0.279	-	0.702	0.446	0.125	0.107	0.452	0.333	0.628	0.476	0.312	0.656	0.306	1.000	
Random Block	0.675	0.390	0.167	0.375	0.343	-	0.982	0.419	0.385	0.909	0.342	0.000	0.096	0.371	0.000	0.000	0.000	0.279	-	0.740	0.446	0.188	0.162	0.462	0.500	0.953	0.488	0.312	0.800	0.322	1.000	
Target Block	0.528	0.366	0.167	0.375	0.319	-	0.875	0.410	0.385	0.736	0.324	0.000	0.070	0.344	0.000	0.000	0.000	0.279	-	0.655	0.447	0.094	0.121	0.462	0.500	0.868	0.488	0.250	0.689	0.306	1.000	
Rolling Shutter	0.764	0.388	0.167	0.500	0.296	-	1.071	0.420	0.385	0.718	0.318	0.000	0.064	0.324	0.000	0.000	0.000	0.271	-	0.757	0.455	0.156	0.110	0.469	0.500	0.922	0.487	0.250	0.733	0.317	1.000	
Resolution Degradation	0.512	0.362	0.167	0.375	0.279	-	0.812	0.400	0.462	0.545	0.302	0.000	0.068	0.329	0.000	0.000	0.000	0.321	-	0.617	0.443	0.094	0.107	0.467	0.333	0.674	0.470	0.188	0.633	0.304	1.000	
Stretch Squish	0.805	0.397	0.167	0.875	0.350	-	1.080	0.427	0.385	0.991	0.354	0.000	0.086	0.355	0.000	0.000	0.000	0.308	-	0.787	0.458	0.125	0.137	0.458	0.333	0.907	0.491	0.188	0.622	0.311	1.000	
Edge Sawtooth	0.780	0.396	0.167	0.125	0.326	-	0.991	0.426	0.385	0.839	0.345	0.000	0.084	0.353	0.000	0.000	0.000	0.274	-	0.957	0.452	0.156	0.113	0.480	0.500	0.977	0.498	0.250	0.682	0.312	1.000	
Color Quantization	0.780	0.396	0.167	0.125	0.326	-	0.991	0.426	0.385	0.839	0.345	0.000	0.084	0.353	0.000	0.000	0.000	0.274	-	0.957	0.452	0.156	0.113	0.480	0.500	0.977	0.498	0.250	0.682	0.312	1.000	
Bright Transform	0.772	0.394	0.167	0.625	0.329	-	0.982	0.425	0.385	0.827	0.343	0.000	0.096	0.371	0.000	0.000	0.000	0.249	-	0.737	0.458	0.125	0.130	0.471	0.500	0.922	0.487	0.250	0.711	0.318	1.000	
Contrast	1.033	0.404	0.167	0.125	0.324	-	1.143	0.439	0.462	0.809	0.343	0.000	0.096	0.371	0.000	0.000	0.000	0.243	-	0.728	0.464	0.188	0.275	0.480	0.500	1.023	0.496	0.250	0.656	0.314	1.000	
Color Shift	0.805	0.416	0.167	0.125	0.334	-	1.107	0.432	0.385	0.955	0.351	0.000	0.081	0.362	0.000	0.000	0.000	0.257	-	0.796	0.463	0.188	0.200	0.464	0.500	0.915	0.485	0.250	0.644	0.317	1.000	
Flicker	0.845	0.416	0.167	0.875	0.316	-	1.099	0.436	0.385	0.864	0.349	0.000	0.096	0.371	0.000	0.000	0.000	0.250	-	0.719	0.456	0.156	0.125	0.477	0.500	0.977	0.497	0.250	0.700	0.296	1.000	
Overexposure	0.805	0.401	0.167	0.875	0.311	-	1.134	0.436	0.385	0.900	0.353	0.000	0.077	0.367	0.000	0.000	0.000	0.259	-	0.753	0.462	0.156	0.123	0.478	0.500	0.946	0.488	0.250	0.700	0.306	1.000	
Underexposure	0.797	0.401	0.167	0.875	0.311	-	1.134	0.436	0.385	0.900	0.353	0.000	0.077	0.367	0.000	0.000	0.000	0.259	-	0.753	0.462	0.156	0.123	0.478	0.500	0.946	0.488	0.250	0.700	0.306	1.000	
Rainy	0.884	0.403	0.167	0.750	0.325	-	1.027	0.427	0.385	0.900	0.348	0.000	0.096	0.371	0.000	0.000	0.000	0.270	-	0.770	0.453	0.125	0.135	0.485	0.500	0.952	0.497	0.250	0.711	0.315	1.000	
Foggy	0.992	0.419	0.167	0.750	0.326	-	1.036	0.431	0.462	0.955	0.354	0.000	0.086	0.360	0.000	0.000	0.000	0.270	-	0.783	0.470	0.188	0.212	0.484	0.500	1.031	0.501	0.250	0.711	0.314	1.000	
Freeze	0.675	0.378	0.167	0.375	0.330	-	0.991	0.425	0.385	0.873	0.342	0.000	0.096	0.371	0.000	0.000	0.000	0.268	-	0.655	0.440	0.031	0.123	0.474	0.500	0.837	0.486	0.188	0.656	0.324	1.000	
Reflection	0.870	0.405	0.167	0.375	0.313	-	1.107	0.435	0.385	0.873	0.342	0.000	0.096	0.371	0.000	0.000	0.000	0.281	-	0.736	0.460	0.125	0.138	0.474	0.500	0.868	0.496	0.250	0.600	0.315	1.000	
Shadow	0.886	0.405	0.167	0.125	0.321	-	1.089	0.433	0.385	0.945	0.355	0.000	0.089	0.366	0.000	0.000	0.000	0.251	-	0.843	0.463	0.125	0.130	0.474	0.500	0.915	0.491	0.250	0.722	0.325	1.000	

Noise Type	K	L	M	N	O	P	Q	R	S	T																					
	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc										
Clean	0.693	0.368	0.000	0.291	0.434	-	0.000	0.816	0.496	0.143	0.832	0.344	0.000	1.000	0.342	0.000	0.919	0.326	0.105	0.978	0.412	0.000	0.877	0.532	0.000	1.179	0.423	0.286	0.948	0.429	0.178
Gaussian	0.508	0.332	0.000	0.424	0.418	-	0.000	0.762	0.487	0.143	0.789	0.323	0.000	1.000	0.342	0.000	0.919	0.326	0.105	0.978	0.412	0.000	0.877	0.532	0.000	1.179	0.423	0.286	0.948	0.429	0.178
Impulse	0.358	0.322	0.000	1.136	0.379	-	0.333	0.592	0.478	0.071	0.621	0.310	0.000	0.923	0.338	0.000	0.726	0.389	0.000	0.967	0.410	0.000	0.967	0.388	0.000						

Table 7: PLLaVA 7B performance across different video types

Noise Type	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T												
Clean	0.569	0.372	0.0000	0.75	0.301	-	0.046	0.406	0.385	0.573	0.322	0.0000	0.625	0.320	0.0000	0.250	0.263	-	0.533	0.417	0.031	0.938	0.449	0.333	0.783	0.456	0.0000	0.467	0.298	0.000		
Gaussian	0.374	0.340	0.0000	0.75	0.301	-	0.670	0.386	0.385	0.573	0.322	0.0000	0.394	0.324	0.0000	0.250	0.373	-	0.340	0.387	0.0000	0.875	0.446	0.333	0.504	0.458	0.0000	0.467	0.298	0.000		
Impulse	0.216	0.338	0.0000	0.125	0.260	-	0.464	0.338	0.231	0.309	0.303	0.0000	0.298	0.302	0.0000	0.000	0.335	-	0.234	0.332	0.0000	0.767	0.422	0.167	0.326	0.417	0.0000	0.278	0.266	0.000		
Poisson	0.236	0.338	0.0000	0.000	0.301	-	0.589	0.370	0.385	0.373	0.305	0.0000	0.288	0.302	0.0000	0.000	0.335	-	0.128	0.393	0.0000	0.787	0.438	0.167	0.256	0.447	0.0000	0.289	0.294	0.000		
Gaussian Blur	0.244	0.333	0.0000	0.375	0.267	-	0.571	0.372	0.154	0.382	0.307	0.0000	0.385	0.312	0.0000	0.250	0.306	-	0.319	0.404	0.0000	0.812	0.434	0.333	0.504	0.454	0.0000	0.411	0.285	0.000		
Motion Blur	0.405	0.320	0.0000	0.175	0.263	-	0.609	0.384	0.325	0.266	0.329	0.0000	0.472	0.305	0.0000	0.000	0.318	-	0.255	0.381	0.0000	0.662	0.432	0.0000	0.426	0.454	0.0000	0.300	0.281	0.000		
Distortion	0.405	0.320	0.0000	0.175	0.263	-	0.609	0.384	0.325	0.266	0.329	0.0000	0.472	0.305	0.0000	0.000	0.318	-	0.255	0.381	0.0000	0.662	0.432	0.0000	0.426	0.454	0.0000	0.300	0.281	0.000		
Zoom Blur	0.431	0.365	0.0000	0.500	0.284	-	0.795	0.400	0.385	0.682	0.329	0.0000	0.519	0.317	0.0000	0.250	0.296	-	0.291	0.465	0.0000	0.633	0.427	0.167	0.401	0.445	0.0000	0.222	0.280	0.000		
Zoom Frame	0.341	0.365	0.0000	0.750	0.302	-	0.589	0.371	0.231	0.345	0.301	0.0000	0.375	0.292	0.0000	0.250	0.306	-	0.451	0.403	0.0000	0.963	0.454	0.333	0.581	0.451	0.0000	0.433	0.295	0.000		
Frame Drop	0.138	0.273	0.0000	0.375	0.246	-	0.446	0.303	0.231	0.336	0.249	0.0000	0.288	0.262	0.0000	0.250	0.321	-	0.260	0.378	0.0000	0.700	0.423	0.333	0.333	0.419	0.0000	0.300	0.254	0.000		
Frame Replace	0.472	0.308	0.0000	0.625	0.329	-	0.902	0.407	0.385	0.645	0.323	0.0000	0.577	0.326	0.0000	0.000	0.321	-	0.498	0.414	0.0000	0.877	0.432	0.333	0.279	0.384	0.0000	0.222	0.241	0.000		
Temporal Jitter	0.203	0.378	0.0000	0.500	0.280	-	0.518	0.346	0.231	0.301	0.273	0.0000	0.375	0.280	0.0000	0.250	0.306	-	0.187	0.319	0.0000	0.700	0.411	0.333	0.357	0.430	0.0000	0.300	0.241	0.000		
Temporal Jitter	0.423	0.325	0.0000	0.375	0.330	-	0.732	0.379	0.231	0.309	0.291	0.0000	0.471	0.310	0.0000	0.250	0.306	-	0.449	0.386	0.0000	0.688	0.421	0.167	0.488	0.458	0.0000	0.411	0.289	0.000		
Jpeg, Artifact	0.407	0.362	0.0000	0.375	0.330	-	0.750	0.383	0.308	0.545	0.337	0.0000	0.475	0.310	0.0000	0.250	0.306	-	0.332	0.402	0.0000	0.688	0.421	0.167	0.488	0.458	0.0000	0.411	0.289	0.000		
Jpeg, Artifact	0.423	0.325	0.0000	0.375	0.330	-	0.732	0.379	0.231	0.309	0.291	0.0000	0.471	0.310	0.0000	0.250	0.306	-	0.449	0.386	0.0000	0.688	0.421	0.167	0.488	0.458	0.0000	0.411	0.289	0.000		
H265 Artifacts	0.276	0.315	0.0000	0.250	0.307	-	0.571	0.347	0.184	0.327	0.380	0.0000	0.500	0.301	0.0000	0.000	0.284	-	0.226	0.376	0.0000	0.912	0.432	0.167	0.349	0.431	0.0000	0.389	0.258	0.000		
Random Block	0.341	0.358	0.0000	0.250	0.307	-	0.741	0.383	0.308	0.600	0.382	0.0000	0.481	0.331	0.0000	0.000	0.277	-	0.374	0.402	0.0000	0.912	0.432	0.167	0.349	0.431	0.0000	0.389	0.258	0.000		
Rolling Shutter	0.520	0.355	0.0000	0.375	0.288	-	0.723	0.367	0.308	0.482	0.311	0.0000	0.587	0.309	0.0000	0.000	0.263	-	0.383	0.401	0.0000	0.838	0.413	0.333	0.473	0.437	0.0000	0.356	0.292	0.000		
Resolution Degrad	0.293	0.334	0.0000	0.250	0.321	-	0.714	0.377	0.231	0.345	0.321	0.0000	0.588	0.322	0.0000	0.750	0.299	-	0.400	0.404	0.0000	0.988	0.437	0.500	0.465	0.455	0.0000	0.456	0.277	0.000		
Stretch Squish	0.244	0.339	0.0000	0.375	0.288	-	0.839	0.391	0.385	0.518	0.319	0.0000	0.490	0.329	0.0000	0.250	0.293	-	0.434	0.405	0.0000	0.838	0.457	0.167	0.628	0.458	0.0000	0.478	0.297	0.000		
Edge Sawtooth	0.512	0.358	0.0000	0.375	0.302	-	0.839	0.383	0.385	0.636	0.331	0.0000	0.567	0.330	0.0000	0.250	0.344	-	0.455	0.407	0.0000	0.838	0.445	0.167	0.597	0.459	0.0000	0.422	0.283	0.000		
Color Quantized	0.423	0.366	0.0000	0.625	0.328	-	0.804	0.443	0.385	0.577	0.322	0.0000	0.590	0.329	0.0000	0.250	0.294	-	0.532	0.440	0.0000	0.950	0.459	0.333	0.571	0.459	0.0000	0.482	0.297	0.000		
Bright Transform	0.455	0.361	0.0000	0.500	0.299	-	0.812	0.395	0.385	0.609	0.328	0.0000	0.558	0.330	0.0000	0.000	0.293	-	0.468	0.416	0.0000	0.887	0.447	0.167	0.636	0.451	0.0000	0.389	0.289	0.000		
Color Shift	0.480	0.352	0.0000	0.625	0.315	-	0.777	0.397	0.385	0.691	0.328	0.0000	0.596	0.343	0.0000	0.000	0.302	-	0.519	0.417	0.0000	0.875	0.457	0.167	0.698	0.458	0.0000	0.433	0.297	0.000		
Flicker	0.480	0.352	0.0000	0.625	0.315	-	0.777	0.397	0.385	0.691	0.328	0.0000	0.596	0.343	0.0000	0.000	0.302	-	0.519	0.417	0.0000	0.875	0.457	0.167	0.698	0.458	0.0000	0.433	0.297	0.000		
Overexposure	0.504	0.362	0.0000	0.625	0.297	-	0.688	0.377	0.308	0.555	0.326	0.0000	0.654	0.344	0.0000	0.500	0.307	-	0.451	0.413	0.0000	0.963	0.447	0.167	0.667	0.458	0.0000	0.489	0.312	0.000		
Underexposure	0.512	0.362	0.0000	0.250	0.320	-	0.723	0.387	0.308	0.645	0.332	0.0000	0.538	0.326	0.0000	0.250	0.289	-	0.434	0.400	0.0000	0.928	0.449	0.500	0.612	0.451	0.0000	0.500	0.302	0.000		
Rainy	0.439	0.355	0.0000	0.375	0.311	-	0.809	0.380	0.388	0.673	0.325	0.0000	0.548	0.328	0.0000	0.250	0.320	-	0.451	0.399	0.0000	0.938	0.449	0.500	0.612	0.451	0.0000	0.500	0.302	0.000		
Peggy	0.585	0.373	0.0000	0.125	0.314	-	0.786	0.403	0.308	0.673	0.325	0.0000	0.625	0.333	0.0000	0.000	0.298	-	0.431	0.409	0.0000	0.887	0.447	0.167	0.580	0.458	0.0000	0.467	0.306	0.000		
Frost	0.390	0.358	0.0000	0.750	0.290	-	0.777	0.384	0.231	0.491	0.322	0.0000	0.529	0.324	0.0000	0.250	0.258	-	0.409	0.400	0.0000	0.912	0.453	0.167	0.605	0.449	0.0000	0.533	0.300	0.000		
Reflect	0.350	0.346	0.0000	0.750	0.318	-	0.812	0.387	0.385	0.645	0.325	0.0000	0.519	0.321	0.0000	0.000	0.293	-	0.460	0.409	0.0000	0.912	0.453	0.167	0.605	0.449	0.0000	0.544	0.289	0.000		
Shadow	0.520	0.363	0.0000	0.625	0.321	-	0.955	0.405	0.308	0.664	0.320	0.0000	0.577	0.325	0.0000	0.250	0.266	-	0.502	0.412	0.0000	1.000	0.453	0.333	0.597	0.452	0.0000	0.489	0.294	0.000		
Clean	0.460	0.342	0.0000	0.348	0.413	-	0.000	0.551	0.454	0.0000	0.632	0.332	0.0000	0.692	0.311	0.0000	0.774	0.391	-	0.000	0.890	0.387	0.0000	0.896	0.590	0.143	0.687	0.396	0.053	0.053		
Gaussian	0.336	0.322	0.0000	0.076	0.376	-	0.376	0.330	0.408	0.438	0.0000	0.621	0.317	0.0000	0.466	0.372	0.0000	0.478	-	0.000	0.692	0.374	0.0000	0.420	0.489	0.0000	0.642	0.378	0.0000	0.507	0.381	0.053
Impulse	0.161	0.300	0.0000	0.924	0.355	-	0.333	0.408	0.438	0.0000	0.263	0.264	0.0000	0.231	0.211	0.0000	0.368	0.344	-	0.000	0.495	0.350	0.0000	0.370	0.473	0.0000	0.387	0.346	0.0000	0.356	0.351	0.053
Poisson	0.226	0.316	0.0000	0.955	0.374	-	0.333	0.272	0.439	0.0000	0.516	0.314	0.0000	0.615	0.290	0.0000	0.075	0.368	-	0.000	0.484	0.362	0.0000	0.480	0.480	0.0000	0.500	0.385	0.143	0.315	0.375	0.053
Gaussian Blur	0.139	0.301	0.0000	0.864	0.340	-	0.000	0.354	0.424	0.0000	0.526	0.304	0.0000	0.581	0.372	0.0000	0.788	0.481	-	0.000	0.458	0.361	0.0000	0.401	0.477	0.0000	0.481	0.359	0.143	0.483	0.374	0.046
Motion Blur	0.340	0.333	0.0000	0.864	0.340	-	0.333	0.503	0.442	0.0000	0.316	0.280	0.0000	0.615	0.326	0.0000	0.312	0.351	-	0.000	0.626	0.356	0.0000	0.481	0.492	0.0000	0.396	0.350	0.143	0.401	0.364	0.026
Distortion	0.327	0.337	0.0000	0.864	0.340	-	0.333	0.374	0.438	0.0000	0.316	0.280	0.0000	0.615	0.326	0.0000	0.312	0.351	-	0.000	0.626	0.356	0.0000	0.481	0.492	0.0000	0.396	0.350	0.143	0.401	0.364	0.026

Table 8: Qwen2.5-VL 3B performance across different video types

Noise Type	A		B		C		D		E		F		G		H		I		J									
	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT								
Clean	1.000	0.414	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Gaussian	0.667	0.371	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Impulse	0.439	0.345	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Speckle	0.740	0.369	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Poisson	0.683	0.379	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Gaussian Blur	0.667	0.380	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Motion Blur	0.751	0.395	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Zoom Blur	0.707	0.386	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Frame Drop	0.602	0.340	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Frame Repeat	0.935	0.414	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Temporal Jitter	0.829	0.390	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Jpeg Artifact	0.951	0.404	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
H265 Artifacts	0.650	0.384	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Random Block	0.780	0.402	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Target Block	0.951	0.404	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Rolling Shutter	0.707	0.380	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Resolution Degradation	0.691	0.375	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Stretch Squish	0.935	0.411	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Edge Sawtooth	1.000	0.409	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Color Quantized	0.919	0.402	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Bright Transform	0.878	0.400	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Contrast	0.919	0.407	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Color Shift	0.911	0.398	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Flicker	0.862	0.398	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Overexposure	0.919	0.410	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Underexposure	0.902	0.407	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Rainy	0.902	0.407	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Foggy	0.941	0.410	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Zoom In	0.878	0.404	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Zoom Out	0.902	0.395	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Reflect	1.000	0.414	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000
Shadow	1.016	0.416	0.167	0.375	0.384	1.000	0.433	0.151	0.964	0.371	0.117	0.154	0.367	0.500	0.500	0.316	0.851	0.470	0.375	0.288	0.481	0.333	0.962	0.511	0.500	0.789	0.332	1.000

Noise Type	K		L		M		N		O		P		Q		R		S		T											
	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT										
Clean	0.991	0.351	0.667	0.773	0.412	1.000	0.014	0.511	0.214	0.179	0.351	0.600	0.923	0.345	1.000	0.011	0.409	0.105	0.187	0.418	1.000	0.068	0.550	0.200	0.600	0.459	0.571	1.075	0.434	0.355
Gaussian	0.453	0.335	0.667	0.773	0.398	1.000	0.803	0.497	0.286	0.874	0.313	0.600	0.923	0.345	1.000	0.011	0.409	0.105	0.187	0.418	1.000	0.068	0.550	0.200	0.600	0.459	0.571	1.075	0.434	0.355
Impulse	0.358	0.310	0.667	0.773	0.398	1.000	0.803	0.497	0.286	0.874	0.313	0.600	0.923	0.345	1.000	0.011	0.409	0.105	0.187	0.418	1.000	0.068	0.550	0.200	0.600	0.459	0.571	1.075	0.434	0.355
Speckle	0.473	0.321	0.333	1.52	0.366	1.000	0.850	0.501	0.357	0.747	0.314	0.200	0.923	0.345	1.000	0.011	0.409	0.105	0.187	0.418	1.000	0.068	0.550	0.200	0.600	0.459	0.571	1.075	0.434	0.355
Poisson	0.372	0.320	0.333	1.52	0.366	1.000	0.850	0.501	0.357	0.747	0.314	0.200	0.923	0.345	1.000	0.011	0.409	0.105	0.187	0.418	1.000	0.068	0.550	0.200	0.600	0.459	0.571	1.075	0.434	0.355
Gaussian Blur	0.365	0.316	0.667	0.773	0.398	1.000	0.803	0.497	0.286	0.874	0.313	0.600	0.923	0.345	1.000	0.011	0.409	0.105	0.187	0.418	1.000	0.068	0.550	0.200	0.600	0.459	0.571	1.075	0.434	0.355
Motion Blur	0.450	0.360	0.667	0.773	0.398	1.000	0.803	0.497	0.286	0.874	0.313	0.600	0.923	0.345	1.000	0.011	0.409	0.105	0.187	0.418	1.000	0.068	0.550	0.200	0.600	0.459	0.571	1.075	0.434	0.355
Zoom Blur	0.450	0.360	0.667	0.773	0.398	1.000	0.803	0.497	0.286	0.874	0.313	0.600	0.923	0.345	1.000	0.011	0.409	0.105	0.187	0.418	1.000	0.068	0.550	0.200	0.600	0.459	0.571	1.075	0.434	0.355
Frame Drop	0.474	0.338	0.667	0.773	0.398	1.000	0.803	0.497	0.2																					

Note: A: Advertisements, B: Algorithm & Models, C: Autos & Vehicles, D: Business & Industrial, E: Computers & Electronics, F: Fairy Tale, G: Films & TV Shows, H: Finance, I: Food & Drink, J: Games, K: Humor, L: Instruction Video (how to), M: Knowledge, N: News, O: Others, P: People, Q: Pets & Animals, R: Science, S: Sports, T: Overall

Table 3. Video ChatGPT-1 performance across different video types

Table 3. Video ChatGPT-1 performance across different video types

Table 3. Video ChatGPT-1 performance across different video types

Table 10: Video-LLaVA performance across different video types

Noise Type	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT
	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT
	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc
Clean	0.846	0.416	0.167	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Gaussian	0.780	0.395	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Impulse	0.732	0.385	0.333	0.125	0.301	0.297	0.093	0.434	0.923	0.855	0.354	0.333	0.788	0.355	0.500	1.000	0.429	0.480	0.489	0.290
Speckle	0.797	0.395	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Poisson	0.789	0.395	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Gaussian Blur	0.618	0.378	0.333	0.375	0.302	0.312	0.091	0.440	0.923	0.882	0.357	0.352	0.222	0.808	0.356	0.500	1.000	0.429	0.480	0.290
Motion Blur	0.772	0.404	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Zoom Blur	0.813	0.404	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Frame Drop	0.724	0.392	0.333	0.125	0.301	0.297	0.093	0.434	0.923	0.855	0.354	0.333	0.788	0.355	0.500	1.000	0.429	0.480	0.489	0.290
Frame Replace	0.780	0.404	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Temporal Jitter	0.752	0.395	0.333	0.375	0.302	0.312	0.091	0.440	0.923	0.882	0.357	0.352	0.222	0.808	0.356	0.500	1.000	0.429	0.480	0.290
Jpeg Artifact	0.772	0.404	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Color Shift	0.752	0.395	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Overexposure	0.752	0.395	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Underexposure	0.752	0.395	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Rainy	0.752	0.395	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Foggy	0.752	0.395	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Freeze	0.752	0.395	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Reflect	0.752	0.395	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293
Shadow	0.752	0.395	0.333	0.250	0.310	0.316	0.152	0.464	0.846	0.809	0.815	0.350	0.278	0.980	0.417	0.300	0.750	0.500	0.433	0.293

Note: A: Advertisements, B: Algorithm & Models, C: Autos & Vehicles, D: Business & Industrial, E: Computers & Electronics, F: Fairy Tale, G: Films & TV Shows, H: Finance, I: Food & Drink, J: Games, K: Humor, L: Instruction Video (how to), M: Knowledge, N: News, O: Others, P: People, Q: Pets & Animals, R: Science, S: Sports, T: Overall

Table 11: VideoChat2-HD performance across different video types

Noise Type	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT	GPT
	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT	SBERT
	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc	Acc
Clean	0.915	0.423	0.167	0.125	0.343	0.121	0.472	0.846	0.918	0.375	0.389	0.981	0.380	0.500	1.000	0.438	0.084	0.491	0.517	0.438
Gaussian	0.199	0.403	0.671	1.125	0.361	1.214	0.456	0.692	0.713	0.361	0.389	0.904	0.380	0.500	1.000	0.438	0.084	0.491	0.517	0.438
Impulse	0.732	0.382	0.337	0.500	0.332	1.241	0.446	0.923	0.773	0.365	0.389	0.779	0.374	0.500	0.500	0.443	0.877	0.488	0.500	0.517
Speckle	0.894	0.403	0.670	0.750	0.376	1.045	0.456	0.923	0.809	0.365	0.389	0.856	0.374	0.500	0.500	0.443	0.877	0.488	0.500	0.517
Poisson	0.829	0.401	0.671	1.125	0.372	1.250	0.452	0.846	0.809	0.371	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Gaussian Blur	0.724	0.390	0.167	0.875	0.368	1.179	0.449	0.846	0.809	0.358	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Motion Blur	0.950	0.373	0.167	0.625	0.361	1.273	0.461	0.846	0.809	0.374	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Distortion	0.950	0.373	0.167	0.625	0.361	1.273	0.461	0.846	0.809	0.374	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Glass Blur	1.098	0.419	0.167	0.625	0.364	1.304	0.468	0.846	0.809	0.382	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Zoom Blur	0.911	0.398	0.388	0.333	0.368	1.188	0.466	0.769	0.782	0.355	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Frame Drop	0.789	0.388	0.333	0.333	0.368	1.152	0.442	0.769	0.782	0.355	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Frame Replace	1.065	0.427	0.167	0.750	0.361	1.259	0.473	0.846	1.009	0.383	0.444	1.048	0.391	0.500	0.500	0.446	1.128	0.511	0.500	0.517
Temporal Jitter	1.106	0.417	0.000	1.000	0.373	0.830	0.432	0.385	0.736	0.354	0.444	0.740	0.357	0.500	0.500	0.446	1.026	0.497	0.500	0.517
Edge Sawtooth	0.894	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Color Shift	1.187	0.421	0.167	0.750	0.361	1.259	0.473	0.846	1.009	0.383	0.444	1.048	0.391	0.500	0.500	0.446	1.026	0.497	0.500	0.517
Contrast	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Brightness	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Quantized	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Random Block	0.927	0.414	0.167	0.500	0.342	1.170	0.459	0.692	0.903	0.374	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Rolling Shutter	0.967	0.408	0.167	0.875	0.349	1.152	0.452	0.846	0.945	0.381	0.444	0.933	0.381	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Resolution Degradation	0.886	0.400	0.333	0.750	0.342	1.054	0.453	0.846	0.864	0.365	0.500	0.865	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Stretch Squish	0.976	0.409	0.167	0.750	0.337	1.152	0.456	0.692	0.736	0.364	0.278	0.952	0.376	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Edge Sawtooth	0.992	0.399	0.333	0.750	0.344	1.098	0.443	0.692	0.800	0.362	0.389	0.846	0.364	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Color Shift	1.187	0.421	0.167	0.750	0.361	1.259	0.473	0.846	1.009	0.383	0.444	1.048	0.391	0.500	0.500	0.446	1.026	0.497	0.500	0.517
Contrast	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Brightness	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Quantized	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Random Block	0.927	0.414	0.167	0.500	0.342	1.170	0.459	0.692	0.903	0.374	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Rolling Shutter	0.967	0.408	0.167	0.875	0.349	1.152	0.452	0.846	0.945	0.381	0.444	0.933	0.381	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Resolution Degradation	0.886	0.400	0.333	0.750	0.342	1.054	0.453	0.846	0.864	0.365	0.500	0.865	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Stretch Squish	0.976	0.409	0.167	0.750	0.337	1.152	0.456	0.692	0.736	0.364	0.278	0.952	0.376	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Edge Sawtooth	0.992	0.399	0.333	0.750	0.344	1.098	0.443	0.692	0.800	0.362	0.389	0.846	0.364	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Color Shift	1.187	0.421	0.167	0.750	0.361	1.259	0.473	0.846	1.009	0.383	0.444	1.048	0.391	0.500	0.500	0.446	1.026	0.497	0.500	0.517
Contrast	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Brightness	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Quantized	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Random Block	0.927	0.414	0.167	0.500	0.342	1.170	0.459	0.692	0.903	0.374	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Rolling Shutter	0.967	0.408	0.167	0.875	0.349	1.152	0.452	0.846	0.945	0.381	0.444	0.933	0.381	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Resolution Degradation	0.886	0.400	0.333	0.750	0.342	1.054	0.453	0.846	0.864	0.365	0.500	0.865	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Stretch Squish	0.976	0.409	0.167	0.750	0.337	1.152	0.456	0.692	0.736	0.364	0.278	0.952	0.376	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Edge Sawtooth	0.992	0.399	0.333	0.750	0.344	1.098	0.443	0.692	0.800	0.362	0.389	0.846	0.364	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Color Shift	1.187	0.421	0.167	0.750	0.361	1.259	0.473	0.846	1.009	0.383	0.444	1.048	0.391	0.500	0.500	0.446	1.026	0.497	0.500	0.517
Contrast	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Brightness	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Quantized	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Random Block	0.927	0.414	0.167	0.500	0.342	1.170	0.459	0.692	0.903	0.374	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Rolling Shutter	0.967	0.408	0.167	0.875	0.349	1.152	0.452	0.846	0.945	0.381	0.444	0.933	0.381	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Resolution Degradation	0.886	0.400	0.333	0.750	0.342	1.054	0.453	0.846	0.864	0.365	0.500	0.865	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Stretch Squish	0.976	0.409	0.167	0.750	0.337	1.152	0.456	0.692	0.736	0.364	0.278	0.952	0.376	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Edge Sawtooth	0.992	0.399	0.333	0.750	0.344	1.098	0.443	0.692	0.800	0.362	0.389	0.846	0.364	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Color Shift	1.187	0.421	0.167	0.750	0.361	1.259	0.473	0.846	1.009	0.383	0.444	1.048	0.391	0.500	0.500	0.446	1.026	0.497	0.500	0.517
Contrast	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Brightness	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Quantized	0.846	0.409	0.333	0.500	0.370	1.214	0.460	0.692	0.882	0.368	0.389	0.109	0.384	0.500	0.500	0.446	0.945	0.488	0.500	0.517
Random Block	0.927	0.414	0.167	0.500	0.342	1.170	0.459	0.692	0.903	0.374	0.389	0.856	0.374	0.500	0.500	0.446	0.945	0.488	0.500	0.517

Table 12: Chat-UniVi-v1.5 performance across different question types

Noise Type	CP		FP-S		FP-C		HL		LR		AR		RR		CSR		TR		Mean2		Overall															
	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT	GPT	SBERT														
Clean	1.080	0.405	0.462	0.860	0.413	0.543	0.770	0.453	0.200	0.870	0.151	0.000	0.900	0.408	0.500	0.500	0.491	0.222	1.190	0.487	0.429	0.920	0.550	0.333	0.600	0.419	0.000	0.850	0.441	0.480	0.860	0.475	0.386	0.900	0.430	0.461
Bit Error	1.000	0.408	0.231	0.890	0.413	0.529	0.770	0.458	0.200	0.890	0.152	0.000	0.900	0.410	0.457	0.560	0.501	0.222	1.080	0.481	0.429	0.960	0.556	0.367	0.740	0.407	0.000	0.760	0.448	0.480	0.820	0.477	0.386	0.880	0.431	0.434
Bright Transform	0.930	0.397	0.231	0.840	0.412	0.529	0.800	0.460	0.200	0.870	0.155	0.000	0.860	0.407	0.457	0.520	0.495	0.222	1.080	0.481	0.429	0.960	0.556	0.367	0.640	0.417	0.000	0.780	0.445	0.520	0.830	0.477	0.400	0.870	0.429	0.441
Color Quantized	1.070	0.404	0.308	0.830	0.412	0.529	0.770	0.457	0.200	0.760	0.156	0.000	0.880	0.409	0.467	0.360	0.482	0.222	1.050	0.471	0.571	0.910	0.557	0.367	0.800	0.422	0.000	0.800	0.442	0.440	0.820	0.472	0.400	0.870	0.429	0.447
Color Shift	1.030	0.404	0.385	0.880	0.415	0.529	0.750	0.460	0.133	0.770	0.148	0.000	0.890	0.410	0.478	0.510	0.502	0.222	1.070	0.483	0.429	0.950	0.556	0.367	0.770	0.415	0.000	0.810	0.449	0.440	0.850	0.479	0.371	0.890	0.432	0.441
Contrast	1.080	0.413	0.385	0.840	0.409	0.529	0.740	0.457	0.200	0.920	0.144	0.000	0.890	0.407	0.478	0.470	0.488	0.222	1.040	0.485	0.429	0.890	0.555	0.367	0.680	0.405	0.000	0.780	0.456	0.480	0.810	0.478	0.400	0.870	0.429	0.447
Defocus Blur	1.020	0.401	0.385	0.860	0.411	0.529	0.790	0.455	0.200	0.790	0.166	0.000	0.900	0.406	0.478	0.560	0.501	0.222	1.050	0.466	0.429	0.880	0.558	0.367	0.700	0.408	0.000	0.700	0.448	0.440	0.780	0.476	0.371	0.870	0.427	0.441
Edge Sawtooth	0.940	0.394	0.308	0.840	0.411	0.529	0.720	0.457	0.200	0.720	0.163	0.000	0.850	0.406	0.478	0.420	0.493	0.222	1.000	0.480	0.429	0.860	0.558	0.367	0.650	0.420	0.000	0.770	0.442	0.440	0.790	0.476	0.371	0.840	0.429	0.428
Elastic	1.060	0.413	0.538	0.830	0.414	0.543	0.710	0.458	0.200	0.810	0.155	0.000	0.870	0.410	0.500	0.440	0.496	0.333	0.990	0.473	0.429	0.930	0.554	0.367	0.640	0.410	0.000	0.790	0.449	0.480	0.800	0.475	0.414	0.870	0.431	0.474
Flicker	1.150	0.404	0.462	0.860	0.412	0.543	0.750	0.455	0.200	0.760	0.151	0.000	0.910	0.407	0.489	0.580	0.499	0.333	0.990	0.467	0.429	0.860	0.554	0.367	0.860	0.416	0.000	0.810	0.445	0.480	0.830	0.473	0.414	0.890	0.429	0.467
Foggy	1.010	0.404	0.385	0.870	0.415	0.529	0.810	0.463	0.133	0.630	0.158	0.000	0.880	0.411	0.478	0.480	0.498	0.222	1.070	0.480	0.571	0.910	0.556	0.367	0.840	0.429	0.000	0.760	0.449	0.440	0.830	0.479	0.400	0.870	0.433	0.454
Frame Drop	0.800	0.370	0.308	0.720	0.396	0.471	0.680	0.449	0.133	0.600	0.147	0.000	0.740	0.390	0.424	0.410	0.486	0.222	1.060	0.449	0.286	0.870	0.552	0.400	0.410	0.368	0.000	0.800	0.446	0.440	0.730	0.462	0.386	0.750	0.413	0.414
Frame Repeat	0.820	0.371	0.385	0.710	0.392	0.486	0.660	0.450	0.200	0.660	0.147	0.000	0.730	0.388	0.435	0.500	0.488	0.222	1.040	0.461	0.286	0.890	0.542	0.367	0.680	0.389	0.000	0.750	0.434	0.480	0.780	0.462	0.386	0.750	0.412	0.421
Frame Replace	1.040	0.404	0.308	0.920	0.418	0.543	0.810	0.456	0.200	0.850	0.152	0.000	0.930	0.412	0.467	0.460	0.491	0.222	1.100	0.480	0.429	0.970	0.555	0.367	0.740	0.428	0.000	0.830	0.447	0.400	0.860	0.478	0.371	0.920	0.432	0.434
Frost	0.890	0.391	0.308	0.830	0.410	0.529	0.760	0.462	0.200	0.900	0.143	0.000	0.840	0.404	0.467	0.410	0.490	0.222	1.060	0.474	0.429	0.930	0.560	0.367	0.590	0.398	0.000	0.780	0.441	0.400	0.780	0.472	0.371	0.840	0.425	0.434
Gaussian	0.990	0.402	0.308	0.830	0.410	0.514	0.780	0.465	0.200	0.740	0.156	0.000	0.860	0.407	0.446	0.510	0.498	0.222	1.090	0.479	0.286	0.990	0.561	0.333	0.860	0.412	0.000	0.760	0.448	0.400	0.830	0.478	0.343	0.860	0.430	0.408
Glass Blur	0.870	0.387	0.308	0.760	0.402	0.514	0.750	0.448	0.067	0.890	0.151	0.000	0.800	0.398	0.446	0.480	0.500	0.222	1.050	0.479	0.286	0.800	0.544	0.400	0.880	0.424	0.000	0.740	0.442	0.440	0.710	0.464	0.386	0.780	0.419	0.428
H265 Artifacts	1.020	0.406	0.308	0.830	0.412	0.529	0.780	0.462	0.133	0.030	0.155	0.000	0.880	0.408	0.457	0.480	0.497	0.222	1.050	0.479	0.286	0.800	0.544	0.400	0.880	0.424	0.000	0.800	0.448	0.480	0.830	0.477	0.400	0.870	0.430	0.441
Impulse	0.880	0.383	0.154	0.750	0.397	0.514	0.610	0.435	0.067	1.020	0.169	0.000	0.780	0.392	0.424	0.500	0.491	0.222	1.080	0.456	0.429	0.760	0.544	0.367	0.420	0.384	0.000	0.710	0.443	0.440	0.720	0.467	0.371	0.760	0.415	0.408
Jpeg Artifact	0.900	0.390	0.308	0.760	0.403	0.514	0.700	0.445	0.067	0.840	0.155	0.000	0.790	0.398	0.435	0.430	0.491	0.222	0.910	0.462	0.429	0.760	0.544	0.367	0.420	0.384	0.000	0.710	0.443	0.480	0.700	0.466	0.400	0.780	0.420	0.428
Motion Blur	0.940	0.398	0.231	0.820	0.411	0.514	0.770	0.468	0.133	0.810	0.162	0.000	0.850	0.408	0.446	0.500	0.502	0.222	0.980	0.468	0.286	0.870	0.561	0.400	0.600	0.403	0.000	0.820	0.444	0.440	0.800	0.474	0.386	0.840	0.428	0.428
Overexposure	0.810	0.379	0.231	0.770	0.398	0.529	0.760	0.451	0.133	1.060	0.154	0.000	0.800	0.394	0.435	0.440	0.496	0.222	0.830	0.454	0.286	0.770	0.538	0.333	0.530	0.399	0.000	0.720	0.441	0.400	0.700	0.465	0.343	0.780	0.417	0.401
Poisson	1.050	0.399	0.462	0.860	0.417	0.543	0.820	0.456	0.200	0.820	0.157	0.000	0.900	0.410	0.500	0.450	0.499	0.222	0.940	0.472	0.286	0.860	0.548	0.333	0.750	0.411	0.000	0.770	0.441	0.480	0.780	0.472	0.371	0.870	0.429	0.454
Rainy	0.990	0.394	0.385	0.880	0.412	0.529	0.730	0.461	0.200	1.020	0.147	0.000	0.890	0.406	0.467	0.530	0.492	0.222	0.940	0.472	0.286	0.880	0.558	0.367	0.670	0.422	0.000	0.790	0.446	0.480	0.810	0.474	0.386	0.880	0.428	0.441
Random Block	1.010	0.403	0.308	0.830	0.412	0.543	0.750	0.460	0.133	0.870	0.147	0.000	0.870	0.408	0.467	0.470	0.494	0.222	1.060	0.480	0.286	1.010	0.562	0.367	0.670	0.422	0.000	0.790	0.448	0.440	0.810	0.479	0.371	0.860	0.430	0.434
Reflect	1.030	0.401	0.308	0.820	0.407	0.557	0.720	0.466	0.200	0.850	0.151	0.000	0.860	0.406	0.489	0.510	0.498	0.222	1.070	0.469	0.429	0.870	0.560	0.333	0.790	0.418	0.000	0.790	0.445	0.520	0.840	0.476	0.400	0.870	0.428	0.461
Resolution Degradation	1.060	0.408	0.308	0.830	0.413	0.529	0.750	0.455	0.200	0.920	0.152	0.000	0.880	0.409	0.457	0.430	0.494	0.222	1.090	0.481	0.429	0.860	0.559	0.367	0.630	0.408	0.000	0.760	0.448	0.440	0.780	0.477	0.371	0.860	0.431	0.428
Rolling Shutter	0.930	0.398	0.231	0.790	0.403	0.514	0.620	0.448	0.200	0.660	0.147	0.000	0.800	0.399	0.446	0.490	0.482	0.222	1.080	0.476	0.286	0.950	0.559	0.367	0.630	0.411	0.000	0.760	0.439	0.440	0.820	0.472	0.357	0.820	0.422	0.414
Shadow	0.980	0.403	0.385	0.830	0.411	0.514	0.770	0.464	0.200	0.810	0.158	0.000	0.860	0.409	0.467	0.460	0.496	0.222	1.050	0.479	0.286	0.860	0.565	0.367	0.690	0.419	0.000	0.790	0.444	0.440	0.820	0.478	0.371	0.860	0.431	0.434
Snow	1.020	0.408	0.538	0.850	0.413	0.586	0.780	0.462	0.200	0.630	0.144	0.000	0.860	0.409	0.533	0.480	0.498	0.222	1.040	0.476	0.286	0.960	0.557	0.333	0.700	0.406	0.000	0.740	0.449	0.520	0.820	0.477	0.386	0.860	0.430	0.474
Speckle	0.920	0.394	0.308	0.830	0.407	0.557	0.700	0.448	0.200	0.970	0.154	0.000	0.850	0.402	0.478	0.510	0.491	0.222	0.920	0.468	0.429	0.870	0.556	0.333	0.600	0.395	0.000	0.740	0.440	0.480	0.780	0.469	0.386	0.840	0.424	0.447
Stretch Squish	0.900	0.388	0.308	0.750	0.400	0.529	0.660	0.456	0.067	0.940	0.159	0.000	0.800	0.398	0.446	0.460	0.460	0.222	1.050	0.453	0.429	0.770	0.547	0.367	0.580											

Table 13: Chat-UniVi performance across different question types

Noise Type	CP		FP-S		FP-C		HL		Mean		LR		AR		RR		CSR		TR		Mean2		Overall														
	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc	GPT	SBERT	Acc													
Clean	1.110	0.404	0.538	0.870	0.413	0.600	0.710	0.437	0.200	0.270	0.142	0.000	0.870	0.407	0.533	0.520	0.492	0.556	1.040	0.477	0.571	0.890	0.539	0.367	0.640	0.406	0.000	0.830	0.440	0.560	0.830	0.473	0.486	0.870	0.429	0.507	
Bit Error	0.960	0.394	0.385	0.880	0.417	0.586	0.620	0.450	0.267	0.180	0.134	0.000	0.840	0.407	0.522	0.600	0.491	0.556	1.060	0.477	0.571	1.010	0.567	0.367	0.650	0.413	0.000	0.850	0.444	0.400	0.880	0.476	0.429	0.850	0.429	0.487	
Bright Transform	1.000	0.394	0.385	0.860	0.414	0.614	0.720	0.455	0.133	0.290	0.144	0.333	0.850	0.406	0.533	0.510	0.479	0.778	1.010	0.477	0.571	0.970	0.564	0.367	0.790	0.421	0.000	0.800	0.440	0.560	0.840	0.474	0.514	0.850	0.428	0.513	
Color Quantized	0.970	0.400	0.308	0.850	0.414	0.600	0.660	0.455	0.200	0.190	0.140	0.333	0.820	0.407	0.523	0.560	0.490	0.556	1.160	0.489	0.286	1.010	0.554	0.300	0.810	0.413	0.000	0.830	0.443	0.480	0.900	0.476	0.400	0.860	0.429	0.461	
Color Shift	1.030	0.403	0.385	0.860	0.414	0.614	0.710	0.459	0.200	0.220	0.143	0.333	0.850	0.408	0.543	0.560	0.486	0.778	1.000	0.477	0.429	1.000	0.562	0.367	0.630	0.407	0.000	0.900	0.441	0.440	0.870	0.474	0.457	0.870	0.430	0.493	
Contrast	1.040	0.404	0.385	0.880	0.416	0.643	0.700	0.460	0.200	0.160	0.130	0.333	0.870	0.410	0.565	0.620	0.490	0.556	1.110	0.487	0.429	1.100	0.565	0.400	0.740	0.424	0.000	0.810	0.441	0.360	0.900	0.478	0.414	0.890	0.432	0.493	
Defocus Blur	1.050	0.406	0.308	0.860	0.412	0.571	0.770	0.463	0.133	0.270	0.138	0.000	0.860	0.408	0.478	0.530	0.481	0.556	1.120	0.483	0.429	1.050	0.567	0.400	0.840	0.424	0.000	0.790	0.440	0.440	0.900	0.477	0.443	0.880	0.430	0.461	
Edge Sawtooth	0.990	0.392	0.385	0.850	0.413	0.629	0.720	0.460	0.133	0.290	0.142	0.000	0.840	0.406	0.533	0.540	0.486	0.667	1.090	0.491	0.143	1.060	0.562	0.400	0.790	0.425	0.000	0.890	0.441	0.440	0.890	0.476	0.429	0.880	0.432	0.480	
Elastic	1.030	0.408	0.385	0.890	0.416	0.629	0.730	0.459	0.067	0.110	0.135	0.333	0.870	0.411	0.533	0.420	0.477	0.667	1.070	0.483	0.571	1.100	0.563	0.333	0.680	0.423	0.000	0.840	0.441	0.440	0.870	0.477	0.443	0.860	0.430	0.461	
Flicker	1.020	0.405	0.462	0.840	0.415	0.514	0.650	0.455	0.200	0.290	0.140	0.333	0.840	0.408	0.478	0.510	0.491	0.667	1.070	0.483	0.571	1.100	0.563	0.333	0.680	0.423	0.000	0.840	0.441	0.440	0.870	0.477	0.443	0.860	0.430	0.461	
Foggy	0.950	0.396	0.385	0.880	0.416	0.614	0.700	0.452	0.200	0.340	0.150	0.333	0.860	0.408	0.486	0.430	0.520	0.488	0.778	0.990	0.463	0.286	0.910	0.554	0.400	0.570	0.396	0.000	0.790	0.437	0.480	0.790	0.466	0.471	0.770	0.418	0.487
Frame Drop	0.850	0.374	0.385	0.750	0.404	0.571	0.590	0.444	0.133	0.230	0.136	0.667	0.740	0.395	0.522	0.500	0.488	0.778	0.990	0.463	0.286	0.910	0.554	0.400	0.570	0.396	0.000	0.790	0.437	0.480	0.790	0.466	0.471	0.770	0.418	0.487	
Frame Repeat	0.770	0.362	0.308	0.690	0.396	0.557	0.530	0.431	0.267	0.350	0.142	0.333	0.680	0.386	0.500	0.560	0.483	0.667	0.930	0.454	0.286	0.950	0.550	0.400	0.470	0.387	0.000	0.750	0.435	0.360	0.790	0.462	0.414	0.720	0.410	0.454	
Frame Replace	1.020	0.399	0.385	0.890	0.416	0.629	0.720	0.458	0.200	0.210	0.145	0.000	0.870	0.410	0.543	0.510	0.487	0.556	1.090	0.480	0.429	1.020	0.565	0.333	0.730	0.420	0.000	0.840	0.445	0.480	0.880	0.477	0.429	0.880	0.431	0.487	
Frost	0.980	0.403	0.462	0.820	0.412	0.600	0.630	0.444	0.200	0.180	0.133	0.333	0.800	0.405	0.543	0.480	0.480	0.667	1.101	0.479	0.571	1.050	0.562	0.433	0.650	0.408	0.000	0.860	0.444	0.440	0.850	0.474	0.486	0.830	0.427	0.513	
Gaussian	0.950	0.402	0.308	0.830	0.412	0.600	0.710	0.459	0.200	0.260	0.140	0.333	0.820	0.406	0.522	0.600	0.493	0.556	1.020	0.476	0.429	0.920	0.557	0.367	0.630	0.410	0.000	0.820	0.442	0.480	0.890	0.477	0.486	0.850	0.429	0.500	
Gaussian Blur	0.870	0.388	0.385	0.790	0.405	0.571	0.640	0.447	0.133	0.180	0.137	0.333	0.770	0.398	0.500	0.470	0.490	0.556	1.020	0.472	0.429	0.920	0.557	0.367	0.630	0.403	0.000	0.810	0.440	0.360	0.800	0.470	0.400	0.790	0.421	0.461	
Glass Blur	1.040	0.404	0.385	0.820	0.409	0.643	0.770	0.463	0.200	0.240	0.137	0.000	0.840	0.406	0.554	0.580	0.491	0.667	1.110	0.476	0.429	0.920	0.551	0.333	0.720	0.404	0.000	0.900	0.441	0.400	0.890	0.471	0.414	0.860	0.427	0.500	
H265 Artifacts	0.870	0.388	0.385	0.780	0.403	0.557	0.630	0.444	0.133	0.230	0.141	0.333	0.760	0.397	0.489	0.580	0.488	0.667	1.060	0.476	0.429	0.860	0.553	0.367	0.780	0.408	0.000	0.770	0.434	0.400	0.770	0.463	0.414	0.770	0.418	0.454	
Impulse	0.810	0.382	0.385	0.790	0.404	0.586	0.580	0.446	0.133	0.260	0.134	0.000	0.740	0.395	0.500	0.470	0.484	0.556	1.020	0.471	0.286	1.010	0.552	0.433	0.650	0.402	0.000	0.740	0.438	0.360	0.810	0.468	0.414	0.770	0.418	0.461	
Jpeg Artifact	0.880	0.395	0.308	0.820	0.413	0.614	0.690	0.458	0.200	0.230	0.141	0.333	0.800	0.406	0.533	0.540	0.490	0.667	1.030	0.486	0.429	1.020	0.560	0.367	0.690	0.406	0.000	0.750	0.433	0.440	0.840	0.473	0.443	0.820	0.427	0.487	
Motion Blur	0.820	0.378	0.385	0.790	0.401	0.600	0.670	0.443	0.133	0.260	0.144	0.000	0.760	0.393	0.500	0.580	0.488	0.556	0.910	0.467	0.429	0.900	0.551	0.400	0.600	0.393	0.000	0.860	0.440	0.440	0.810	0.466	0.443	0.780	0.416	0.474	
Overexposure	1.010	0.400	0.462	0.860	0.411	0.629	0.700	0.454	0.200	0.290	0.148	0.333	0.850	0.406	0.565	0.400	0.480	0.667	0.940	0.477	0.286	1.080	0.564	0.367	0.530	0.406	0.000	0.810	0.445	0.440	0.810	0.473	0.429	0.850	0.428	0.493	
Poisson	0.980	0.393	0.462	0.850	0.410	0.614	0.620	0.450	0.200	0.290	0.144	0.333	0.830	0.403	0.554	0.460	0.489	0.667	1.010	0.476	0.286	1.080	0.568	0.367	0.620	0.417	0.000	0.780	0.440	0.360	0.830	0.475	0.400	0.840	0.426	0.480	
Rainy	1.030	0.402	0.385	0.860	0.414	0.600	0.640	0.458	0.200	0.260	0.143	0.333	0.850	0.408	0.533	0.470	0.481	0.667	1.040	0.477	0.429	0.940	0.553	0.367	0.630	0.420	0.000	0.800	0.441	0.440	0.820	0.472	0.443	0.850	0.426	0.474	
Random Block	1.070	0.398	0.462	0.840	0.410	0.571	0.650	0.458	0.200	0.160	0.138	0.000	0.840	0.404	0.511	0.570	0.487	0.667	1.060	0.481	0.286	1.020	0.560	0.333	0.750	0.422	0.000	0.840	0.440	0.480	0.880	0.474	0.429	0.850	0.426	0.474	
Reflect	0.920	0.396	0.385	0.830	0.409	0.586	0.670	0.455	0.200	0.210	0.131	0.000	0.810	0.404	0.511	0.570	0.484	0.556	1.110	0.484	0.429	0.980	0.566	0.300	0.640	0.408	0.000	0.820	0.440	0.400	0.860	0.474	0.386	0.830	0.426	0.454	
Resolution Degrade	0.970	0.398	0.308	0.830	0.410	0.600	0.650	0.451	0.133	0.210	0.141	0.333	0.820	0.404	0.511	0.590	0.488	0.667	1.010	0.475	0.286	1.120	0.566	0.367	0.670	0.403	0.000	0.830	0.443	0.440	0.880	0.474	0.429	0.850	0.426	0.474	
Rolling Shutter	0.930	0.389	0.462	0.830	0.410	0.600	0.700	0.451	0.133	0.230	0.145	0.000	0.810	0.402	0.511	0.450	0.493	0.556	0.940	0.474	0.143	1.110	0.572	0.267	0.690	0.421	0.000	0.780	0.432	0.400	0.830	0.473	0.343	0.820	0.425	0.434	
Shadow	0.990	0.403	0.380	0.870	0.415	0.614	0.650	0.455	0.133	0.210	0.139	0.000	0.840	0.408	0.511	0.600	0.492	0.66																			

Table 14: GPT-4o performance across different question types

Noise Type	CP	FP-S	FP-C	HL	Mean	LR	AR	RR	CSR	TR	Mean2	Overall																								
Clean	1.980	0.529	0.538	1.670	0.531	0.614	1.570	0.515	0.600	2.400	0.188	0.667	1.740	0.516	0.609	1.360	0.653	1.000	1.940	0.527	0.571	1.750	0.599	0.467	1.910	0.524	0.000	1.630	0.493	0.360	1.720	0.550	0.500	1.740	0.527	0.546
Bit Error	1.770	0.517	0.615	1.740	0.553	0.600	1.380	0.503	0.667	0.970	0.144	1.000	1.670	0.523	0.620	1.420	0.681	0.667	2.100	0.540	0.857	0.050	0.620	0.700	1.900	0.544	0.500	1.710	0.513	0.520	1.830	0.570	0.643	1.720	0.536	0.638
Bright Transform	1.820	0.520	0.923	1.690	0.550	0.643	1.540	0.521	0.600	1.180	0.133	1.000	1.680	0.524	0.685	1.420	0.681	0.889	2.170	0.546	0.714	1.860	0.634	0.600	1.880	0.556	0.500	1.670	0.514	0.520	1.800	0.574	0.643	1.730	0.540	0.664
Color Quantized	1.660	0.466	0.462	1.490	0.487	0.557	0.920	0.335	0.333	0.980	0.144	0.667	1.410	0.445	0.500	0.020	0.449	1.000	0.850	0.265	0.429	0.480	0.190	0.033	1.600	0.491	0.500	0.840	0.278	0.240	0.870	0.304	0.286	1.220	0.396	0.388
Color Shift	1.840	0.528	0.692	1.730	0.551	0.643	1.530	0.522	0.333	1.000	0.129	1.000	1.700	0.526	0.674	1.540	0.700	1.000	2.020	0.543	0.857	1.960	0.629	0.667	2.070	0.569	0.000	1.630	0.513	0.520	1.820	0.576	0.671	1.740	0.541	0.671
Contrast	1.890	0.520	0.692	1.760	0.548	0.729	1.460	0.512	0.733	1.310	0.146	0.333	1.720	0.521	0.717	1.590	0.682	1.000	2.030	0.546	1.000	1.960	0.622	0.533	1.880	0.534	1.000	1.710	0.512	0.480	1.820	0.569	0.643	1.750	0.536	0.678
Defocus Blur	1.740	0.518	0.846	1.670	0.548	0.645	1.460	0.507	0.600	0.920	0.144	1.000	1.620	0.521	0.674	1.220	0.659	0.889	1.990	0.538	0.571	1.950	0.624	0.633	1.680	0.544	0.500	1.600	0.502	0.560	1.700	0.557	0.657	1.660	0.532	0.658
Edge Sawtooth	1.810	0.530	0.846	1.700	0.549	0.657	1.580	0.514	0.600	0.890	0.134	1.000	1.660	0.524	0.685	1.420	0.680	0.778	2.080	0.540	1.000	1.950	0.628	0.367	1.950	0.557	0.500	1.680	0.516	0.520	1.830	0.573	0.629	1.720	0.539	0.658
Elastic	1.910	0.530	0.692	1.740	0.552	0.714	1.500	0.511	0.733	0.900	0.136	1.000	1.700	0.525	0.717	1.420	0.682	0.667	2.140	0.555	0.857	2.080	0.635	0.633	1.790	0.550	1.000	1.690	0.520	0.560	1.840	0.578	0.686	1.750	0.541	0.704
Flicker	1.820	0.516	0.846	1.680	0.548	0.671	1.480	0.515	0.667	0.970	0.145	0.667	1.650	0.521	0.685	1.350	0.669	0.778	2.040	0.557	0.714	2.020	0.633	0.633	1.790	0.550	1.000	1.620	0.512	0.400	1.780	0.574	0.586	1.690	0.537	0.651
Foggy	1.860	0.522	0.692	1.690	0.546	0.643	1.510	0.501	0.733	1.060	0.137	1.000	1.670	0.519	0.663	1.440	0.678	0.667	2.070	0.549	0.857	1.910	0.629	0.567	1.830	0.540	0.500	1.680	0.520	0.480	1.780	0.574	0.586	1.710	0.536	0.632
Frame Drop	1.340	0.457	0.692	1.190	0.497	0.543	0.970	0.460	0.667	0.810	0.143	0.667	1.170	0.471	0.563	0.920	0.571	0.667	1.470	0.486	0.571	1.470	0.598	0.533	1.540	0.543	0.500	1.430	0.472	0.760	1.400	0.522	0.643	1.250	0.486	0.599
Frame Repeat	1.300	0.457	0.538	1.090	0.469	0.386	1.090	0.455	0.467	0.100	0.145	0.667	1.150	0.454	0.535	0.850	0.592	0.556	1.630	0.505	0.429	1.540	0.569	0.567	1.440	0.514	0.500	1.370	0.487	0.560	1.390	0.525	0.557	1.220	0.475	0.487
Frame Replace	1.810	0.527	0.692	1.730	0.546	0.671	1.460	0.506	0.733	1.160	0.135	1.000	1.690	0.521	0.685	1.470	0.667	0.778	2.100	0.561	0.857	1.950	0.618	0.633	1.930	0.544	1.000	1.690	0.518	0.400	1.820	0.573	0.629	1.740	0.536	0.664
Frost	1.790	0.522	0.769	1.680	0.554	0.671	1.400	0.496	0.733	0.920	0.134	0.667	1.620	0.523	0.685	1.460	0.656	0.667	1.890	0.537	0.857	1.890	0.621	0.633	1.670	0.536	0.500	1.600	0.511	0.360	1.720	0.563	0.571	1.650	0.534	0.638
Gaussian	1.760	0.514	0.769	1.610	0.540	0.657	1.400	0.501	0.733	1.160	0.140	1.000	1.590	0.513	0.685	1.440	0.659	0.889	2.020	0.532	0.714	1.890	0.622	0.633	1.730	0.537	1.000	1.690	0.509	0.440	1.760	0.559	0.643	1.650	0.528	0.658
Glass Blur	1.570	0.492	0.769	1.450	0.517	0.629	1.250	0.492	0.600	0.980	0.136	1.000	1.430	0.495	0.641	1.120	0.647	0.667	1.880	0.513	0.857	1.930	0.603	0.733	1.620	0.518	0.500	1.520	0.492	0.640	1.620	0.544	0.686	1.490	0.510	0.664
H265 Artifacts	1.740	0.509	0.769	1.620	0.537	0.600	1.310	0.497	0.467	1.060	0.138	0.667	1.380	0.512	0.598	1.210	0.673	0.444	1.930	0.540	0.714	1.970	0.611	0.667	1.800	0.536	1.000	1.690	0.517	0.600	1.740	0.568	0.643	1.630	0.528	0.625
Impulse	1.530	0.482	0.769	1.380	0.515	0.543	1.190	0.480	0.533	0.740	0.127	0.667	1.360	0.490	0.565	1.220	0.640	0.667	1.810	0.520	1.000	1.760	0.602	0.700	1.530	0.510	1.000	1.520	0.495	0.640	1.590	0.547	0.686	1.430	0.507	0.605
Jpeg Artifact	1.450	0.477	0.692	1.410	0.515	0.643	1.210	0.473	0.667	0.840	0.169	1.000	1.370	0.489	0.641	1.060	0.619	0.778	1.750	0.516	0.429	1.830	0.589	0.600	1.490	0.512	0.000	1.560	0.502	0.480	1.570	0.541	0.557	1.430	0.505	0.612
Motion Blur	1.790	0.518	0.769	1.650	0.535	0.686	1.360	0.497	0.733	0.980	0.132	1.000	1.610	0.512	0.707	1.430	0.662	0.778	2.020	0.540	0.714	2.020	0.623	0.700	1.750	0.531	1.000	1.690	0.512	0.640	1.800	0.565	0.714	1.680	0.528	0.724
Overexposure	1.480	0.477	0.846	1.350	0.504	0.557	1.320	0.492	0.600	1.030	0.147	0.333	1.360	0.484	0.587	1.080	0.654	0.667	1.670	0.514	0.714	1.750	0.606	0.567	1.590	0.522	0.500	1.460	0.488	0.520	1.500	0.544	0.571	1.410	0.503	0.579
Poisson	1.770	0.518	0.769	1.700	0.552	0.700	1.410	0.508	0.667	0.890	0.125	0.667	1.640	0.523	0.696	1.420	0.671	0.778	1.990	0.545	1.000	1.960	0.620	0.633	1.790	0.530	0.500	1.550	0.511	0.520	1.730	0.567	0.657	1.680	0.537	0.684
Rainy	1.820	0.526	0.769	1.590	0.546	0.586	1.360	0.492	0.667	1.020	0.136	0.333	1.550	0.517	0.598	1.260	0.660	0.778	1.980	0.544	0.714	1.890	0.625	0.600	1.810	0.530	1.000	1.560	0.499	0.520	1.690	0.562	0.629	1.600	0.531	0.599
Random Block	1.770	0.514	0.769	1.680	0.538	0.743	1.420	0.502	0.800	1.050	0.139	1.000	1.640	0.513	0.761	1.320	0.670	1.000	2.100	0.554	0.857	2.010	0.603	0.733	1.830	0.547	0.500	1.670	0.513	0.560	1.790	0.568	0.700	1.690	0.529	0.730
Reflect	1.830	0.522	0.769	1.710	0.545	0.614	1.450	0.517	0.600	1.060	0.129	0.667	1.670	0.520	0.620	1.420	0.676	0.667	2.020	0.536	0.714	1.970	0.614	0.667	1.850	0.538	0.500	1.690	0.508	0.480	1.800	0.565	0.614	1.720	0.533	0.618
Resolution Degrade	1.740	0.513	0.769	1.570	0.533	0.629	1.270	0.453	0.600	0.820	0.151	1.000	1.540	0.503	0.652	1.200	0.628	0.667	1.430	0.423	0.857	1.410	0.426	0.200	1.810	0.564	0.500	1.420	0.478	0.480	1.440	0.492	0.429	1.500	0.496	0.546
Rolling Shutter	1.710	0.510	0.615	1.550	0.533	0.557	1.270	0.499	0.667	1.020	0.139	1.000	1.520	0.509	0.576	1.120	0.674	0.556	1.920	0.529	0.857	2.020	0.628	0.733	1.720	0.536	0.500	1.600	0.514	0.480	1.690	0.566	0.629	1.580	0.526	0.612
Shadow	1.820	0.510	0.769	1.750	0.554	0.714	1.520	0.511	0.733	1.020	0.140	1.000	1.700	0.525	0.728	1.490	0.697	0.889	1.970	0.542	0.714	1.950	0.612	0.533	1.950	0.555	0.500	1.730	0.518	0.520	1.800	0.573	0.600	1.730	0.539	0.678
Snow	1.880	0.517	0.769	1.710	0.547	0.651	1.490	0.518	0.733	1.100	0.138	1.000	1.690	0.522	0.685	1.380	0.664	0.778	2.140	0.528	0.857	1.910	0.628	0.600	1.910	0.550	1.000	1.730	0.518	0.480	1.830	0.568	0.643	1.740	0.535	0.678
Speckle	1.850	0.527	0.615	1.610	0.546	0.651	1.380	0.497	0.600	0.920	0.144	0.667	1.600	0.519	0.609	1.270	0.663	0.667	2.010	0.534	1.000	2.000	0.622	0.633	1.900	0.565	0.500	1.590	0.502	0.520	1.750	0.564	0.643	1.640	0.533	0.612
Stretch Squish	1.570	0.497	0.846	1.470	0.513	0.571	1.190	0.496	0.600	0.790	0.146	1.000	1.430	0.495	0.630	1.010	0.644	0.333	1.910	0.530	0.714	1.800	0.613	0.600	1.700	0.518	1.000	1.510	0.508	0.560	1.600	0.555	0.586	1.490	0.514	0.632
Target Block	1.780	0.511	0.769	1.590	0.542	0.614	1.330	0.498	0.667	1.160	0.145	1.000	1.580	0.515	0.641	1.400	0.673	0.667	2.060	0.555	0.857	1.750	0.													

Table 15: LLaMA-VID performance across different question types

Noise Type	CP	FP-S	FP-C	HL	Mean	LR	AR	RR	CSR	TR	Mean2	Overall																								
Clean	1.020	0.411	0.385	0.890	0.417	0.514	0.800	0.460	0.133	0.600	0.168	0.000	0.910	0.414	0.467	0.530	0.496	0.222	1.120	0.473	0.429	0.900	0.562	0.367	0.730	0.425	0.000	0.900	0.455	0.480	0.890	0.481	0.400	0.910	0.433	0.447
Bit Error	1.010	0.409	0.385	0.890	0.417	0.514	0.710	0.460	0.133	0.790	0.163	0.000	0.910	0.413	0.467	0.580	0.507	0.222	1.040	0.471	0.429	0.890	0.557	0.367	0.720	0.417	0.000	0.870	0.452	0.440	0.860	0.480	0.411	0.910	0.436	0.447
Bright Transform	1.040	0.413	0.385	0.870	0.419	0.500	0.800	0.458	0.133	0.610	0.167	0.000	0.900	0.416	0.457	0.380	0.507	0.222	1.040	0.467	0.429	0.920	0.559	0.367	0.770	0.416	0.000	0.900	0.456	0.520	0.880	0.480	0.414	0.910	0.436	0.447
Color Quantized	1.030	0.407	0.385	0.850	0.416	0.514	0.770	0.460	0.133	0.610	0.166	0.000	0.890	0.414	0.467	0.570	0.508	0.222	1.090	0.478	0.429	0.790	0.556	0.367	0.750	0.423	0.000	0.890	0.454	0.440	0.860	0.481	0.386	0.890	0.435	0.441
Color Shift	0.990	0.411	0.385	0.840	0.414	0.500	0.740	0.460	0.133	0.630	0.165	0.000	0.870	0.412	0.467	0.570	0.501	0.222	1.030	0.473	0.429	0.890	0.553	0.367	0.740	0.419	0.000	0.900	0.455	0.440	0.860	0.479	0.371	0.870	0.434	0.428
Contrast	0.980	0.406	0.385	0.870	0.417	0.514	0.790	0.460	0.133	0.650	0.170	0.000	0.890	0.414	0.467	0.530	0.499	0.222	1.070	0.472	0.429	1.020	0.562	0.367	0.680	0.412	0.000	0.890	0.454	0.440	0.890	0.480	0.386	0.900	0.435	0.441
Defocus Blur	1.010	0.409	0.385	0.840	0.413	0.514	0.780	0.458	0.133	0.660	0.165	0.000	0.880	0.411	0.467	0.510	0.499	0.222	1.090	0.475	0.428	0.900	0.557	0.367	0.810	0.416	0.000	0.900	0.452	0.440	0.860	0.479	0.371	0.880	0.433	0.434
Edge Sawtooth	1.020	0.406	0.385	0.860	0.418	0.500	0.780	0.452	0.133	0.580	0.164	0.000	0.890	0.414	0.467	0.560	0.492	0.222	1.030	0.471	0.429	0.860	0.553	0.367	0.750	0.427	0.000	0.880	0.456	0.440	0.860	0.480	0.386	0.890	0.434	0.434
Elastic	1.020	0.413	0.385	0.870	0.416	0.500	0.800	0.460	0.133	0.550	0.170	0.000	0.890	0.414	0.467	0.550	0.499	0.222	1.060	0.473	0.429	0.890	0.552	0.367	0.690	0.410	0.000	0.900	0.461	0.440	0.860	0.481	0.386	0.890	0.435	0.434
Flicker	1.100	0.415	0.385	0.860	0.415	0.514	0.810	0.460	0.133	0.630	0.164	0.000	0.910	0.414	0.467	0.540	0.502	0.222	1.030	0.469	0.429	0.890	0.559	0.367	0.690	0.410	0.000	0.900	0.459	0.440	0.860	0.480	0.386	0.910	0.435	0.441
Foggy	0.940	0.402	0.385	0.840	0.416	0.500	0.750	0.457	0.133	0.630	0.169	0.000	0.860	0.412	0.467	0.580	0.501	0.222	1.060	0.470	0.429	0.800	0.557	0.367	0.830	0.416	0.000	0.750	0.454	0.440	0.820	0.480	0.386	0.860	0.432	0.434
Frame Drop	0.810	0.382	0.231	0.730	0.402	0.500	0.590	0.442	0.133	0.500	0.162	0.000	0.730	0.395	0.435	0.570	0.503	0.222	1.010	0.463	0.428	0.870	0.557	0.367	0.590	0.399	0.000	0.840	0.449	0.440	0.820	0.474	0.371	0.760	0.420	0.410
Frame Repeat	0.810	0.370	0.385	0.680	0.389	0.500	0.510	0.435	0.133	0.520	0.154	0.000	0.690	0.384	0.457	0.380	0.497	0.222	0.840	0.456	0.429	0.860	0.545	0.367	0.520	0.383	0.000	0.780	0.448	0.440	0.750	0.466	0.386	0.720	0.410	0.434
Frame Replace	1.060	0.412	0.385	0.860	0.417	0.514	0.800	0.461	0.133	0.660	0.166	0.000	0.900	0.414	0.467	0.480	0.500	0.222	1.040	0.473	0.429	0.970	0.564	0.367	0.730	0.415	0.000	0.880	0.455	0.440	0.880	0.480	0.386	0.900	0.435	0.441
Frost	0.870	0.397	0.308	0.800	0.412	0.500	0.680	0.454	0.133	0.680	0.168	0.000	0.810	0.407	0.446	0.500	0.501	0.222	0.990	0.472	0.429	0.820	0.557	0.367	0.670	0.394	0.000	0.820	0.454	0.440	0.800	0.477	0.386	0.820	0.430	0.428
Gaussian	0.980	0.410	0.385	0.840	0.412	0.500	0.750	0.455	0.133	0.650	0.166	0.000	0.860	0.410	0.457	0.500	0.503	0.222	1.060	0.471	0.428	0.930	0.558	0.367	0.690	0.421	0.000	0.860	0.458	0.440	0.860	0.481	0.357	0.870	0.433	0.421
Gaussian Blur	0.920	0.395	0.385	0.790	0.406	0.500	0.700	0.455	0.133	0.710	0.164	0.000	0.810	0.403	0.457	0.630	0.490	0.222	0.960	0.459	0.429	0.820	0.555	0.367	0.670	0.407	0.000	0.900	0.455	0.480	0.840	0.472	0.440	0.830	0.425	0.441
Glass Blur	1.010	0.412	0.385	0.840	0.413	0.500	0.700	0.453	0.133	0.710	0.168	0.000	0.870	0.411	0.457	0.530	0.502	0.222	1.050	0.472	0.429	0.920	0.558	0.367	0.790	0.418	0.000	0.820	0.455	0.440	0.850	0.480	0.371	0.880	0.433	0.441
H265 Artifacts	0.930	0.390	0.308	0.780	0.404	0.500	0.620	0.450	0.133	0.740	0.161	0.000	0.800	0.399	0.446	0.600	0.504	0.222	0.980	0.470	0.429	0.830	0.548	0.367	0.850	0.418	0.000	0.810	0.451	0.520	0.790	0.473	0.411	0.810	0.422	0.441
Impulse	0.890	0.398	0.231	0.770	0.405	0.500	0.650	0.452	0.133	0.660	0.165	0.000	0.790	0.403	0.435	0.630	0.502	0.222	0.930	0.454	0.429	0.860	0.552	0.367	0.780	0.413	0.000	0.830	0.451	0.440	0.840	0.473	0.371	0.810	0.425	0.441
Jpeg Artifact	0.970	0.411	0.308	0.840	0.413	0.514	0.760	0.457	0.133	0.680	0.165	0.000	0.870	0.411	0.457	0.510	0.504	0.222	0.960	0.461	0.429	0.890	0.556	0.367	0.620	0.410	0.000	0.890	0.454	0.440	0.830	0.476	0.371	0.860	0.432	0.428
Motion Blur	0.830	0.387	0.308	0.750	0.400	0.500	0.620	0.444	0.133	0.650	0.170	0.000	0.760	0.397	0.446	0.570	0.494	0.222	0.890	0.448	0.429	0.730	0.547	0.367	0.690	0.395	0.000	0.820	0.449	0.520	0.770	0.467	0.410	0.770	0.419	0.419
Overexposure	0.990	0.408	0.385	0.850	0.416	0.500	0.800	0.457	0.133	0.660	0.164	0.000	0.890	0.413	0.457	0.550	0.499	0.222	1.060	0.472	0.429	0.800	0.558	0.367	0.800	0.431	0.000	0.840	0.460	0.440	0.840	0.483	0.386	0.880	0.435	0.434
Poisson	0.990	0.409	0.380	0.840	0.412	0.500	0.760	0.459	0.133	0.680	0.167	0.000	0.850	0.410	0.446	0.540	0.504	0.222	0.990	0.465	0.429	0.860	0.561	0.367	0.720	0.419	0.000	0.890	0.454	0.440	0.850	0.478	0.371	0.860	0.432	0.421
Rainy	0.980	0.407	0.385	0.840	0.414	0.500	0.680	0.456	0.133	0.650	0.164	0.000	0.860	0.410	0.457	0.550	0.498	0.222	1.010	0.474	0.429	0.890	0.558	0.367	0.800	0.420	0.000	0.870	0.455	0.480	0.870	0.480	0.400	0.870	0.433	0.441
Random Block	0.970	0.407	0.385	0.810	0.411	0.500	0.760	0.465	0.133	0.550	0.163	0.000	0.840	0.410	0.457	0.590	0.502	0.222	1.030	0.472	0.429	0.910	0.555	0.367	0.700	0.408	0.000	0.860	0.455	0.440	0.850	0.478	0.371	0.860	0.432	0.428
Reflect	1.020	0.414	0.385	0.840	0.415	0.514	0.730	0.457	0.133	0.610	0.169	0.000	0.880	0.413	0.467	0.570	0.504	0.222	1.060	0.474	0.429	0.900	0.564	0.367	0.830	0.421	0.000	0.880	0.455	0.440	0.870	0.482	0.386	0.880	0.435	0.441
Resolution Degradation	0.960	0.406	0.385	0.810	0.411	0.500	0.740	0.459	0.133	0.580	0.167	0.000	0.850	0.409	0.457	0.530	0.496	0.222	1.020	0.475	0.429	0.910	0.554	0.367	0.680	0.420	0.000	0.870	0.458	0.440	0.850	0.480	0.386	0.860	0.432	0.434
Rolling Shutter	0.910	0.397	0.385	0.830	0.410	0.500	0.730	0.452	0.133	0.610	0.168	0.000	0.830	0.406	0.457	0.510	0.509	0.222	0.940	0.464	0.429	0.880	0.552	0.367	0.730	0.418	0.000	0.920	0.458	0.440	0.860	0.478	0.371	0.850	0.429	0.428
Shadow	0.960	0.407	0.385	0.860	0.417	0.500	0.760	0.462	0.133	0.560	0.162	0.000	0.870	0.413	0.457	0.380	0.508	0.222	1.070	0.474	0.429	0.800	0.558	0.367	0.680	0.419	0.000	0.900	0.457	0.480	0.870	0.480	0.400	0.880	0.436	0.441
Snow	1.070	0.415	0.385	0.870	0.417	0.500	0.780	0.463	0.133	0.560	0.165	0.000	0.900	0.415	0.457	0.500	0.503	0.222	1.070	0.470	0.429	0.890	0.562	0.367	0.690	0.421	0.000	0.930	0.458	0.440	0.890	0.482	0.386	0.900	0.436	0.434
Speckle	0.950	0.412	0.388	0.830	0.415	0.500	0.730	0.455	0.133	0.530	0.163	0.000	0.840	0.412	0.446	0.560	0.508	0.222	1.020	0.469	0.429	0.890	0.558	0.367	0.690	0.421	0.000	0.870	0.454	0.440	0.870	0.483	0.371	0.850	0.434	0.421
Stretch Squish	0.910	0.400	0.385	0.780	0.407	0.500	0.680	0.457	0.133	0.770	0.170	0.000	0.810	0.405	0.447	0.510	0.493	0.222	1.020	0.470	0.429	0.820	0.556	0.367	0.620	0.408	0.000	0.810	0.456	0.440	0.800	0.476	0.386	0.820	0.428	0.434
Target Block	0.950	0.407	0.388	0.760																																

Table 16: PLLaVA 13B performance across different question types

Noise Type	CP	FP-S	FP-C	HL	Mean	LR	AR	RR	CSR	TR	Mean2	Overall
	GPT SBERT Acc	GPT SBERT Acc	GPT SBERT Acc	GPT SBERT Acc	GPT SBERT Acc	GPT SBERT Acc	GPT SBERT Acc	GPT SBERT Acc	GPT SBERT Acc	GPT SBERT Acc	GPT SBERT Acc	GPT SBERT Acc
Clean	1.140	0.421	0.308	0.940	0.412	0.214	0.740	0.451	0.000	0.970	0.410	0.207
Bit Error	1.060	0.408	0.154	0.810	0.398	0.214	0.700	0.441	0.000	0.890	0.414	0.207
Bright Transform	1.040	0.406	0.231	0.850	0.405	0.200	0.660	0.451	0.000	0.840	0.413	0.200
Color Quantized	1.080	0.408	0.231	0.870	0.405	0.200	0.690	0.445	0.000	0.880	0.413	0.200
Color Shift	1.070	0.413	0.308	0.880	0.406	0.186	0.780	0.447	0.000	0.810	0.413	0.200
Contrast	1.130	0.417	0.308	0.910	0.410	0.214	0.700	0.454	0.000	0.820	0.413	0.200
Defocus Blur	1.020	0.406	0.308	0.810	0.397	0.214	0.720	0.451	0.000	0.890	0.413	0.200
Edge Sawtooth	1.020	0.405	0.154	0.850	0.399	0.186	0.750	0.445	0.000	0.740	0.413	0.200
Elastic	1.070	0.412	0.308	0.880	0.407	0.186	0.680	0.447	0.000	0.890	0.413	0.200
Flicker	1.080	0.411	0.308	0.870	0.405	0.200	0.740	0.450	0.000	0.810	0.413	0.200
Foggy	1.090	0.412	0.231	0.890	0.408	0.200	0.700	0.443	0.000	0.880	0.413	0.200
Frame Drop	0.610	0.360	0.154	0.530	0.368	0.114	0.530	0.415	0.000	0.980	0.413	0.200
Frame Repeat	0.560	0.344	0.231	0.480	0.352	0.086	0.480	0.399	0.067	1.210	0.156	0.000
Frame Replace	1.200	0.420	0.308	0.890	0.409	0.200	0.780	0.449	0.000	0.880	0.413	0.200
Frost	0.970	0.389	0.154	0.800	0.397	0.200	0.690	0.430	0.000	0.710	0.413	0.200
Gaussian	0.980	0.400	0.154	0.790	0.401	0.200	0.660	0.446	0.000	0.680	0.413	0.200
Glass Blur	0.910	0.400	0.154	0.600	0.372	0.143	0.540	0.416	0.000	0.850	0.413	0.200
H265 Artifacts	0.580	0.354	0.154	0.530	0.366	0.143	0.440	0.410	0.000	0.950	0.413	0.200
Impulse	0.750	0.381	0.077	0.670	0.379	0.171	0.580	0.419	0.000	1.080	0.413	0.200
Jpeg Artifact	0.880	0.398	0.154	0.780	0.400	0.200	0.650	0.442	0.000	0.810	0.413	0.200
Motion Blur	0.570	0.357	0.154	0.580	0.365	0.114	0.520	0.406	0.000	0.760	0.413	0.200
Overexposure	1.030	0.408	0.308	0.860	0.405	0.214	0.710	0.445	0.000	0.870	0.413	0.200
Poisson	0.890	0.391	0.154	0.750	0.392	0.186	0.550	0.432	0.000	0.820	0.413	0.200
Rainy	1.000	0.408	0.231	0.870	0.406	0.229	0.690	0.449	0.000	0.850	0.413	0.200
Random Block	0.960	0.401	0.385	0.770	0.396	0.200	0.640	0.431	0.000	0.790	0.413	0.200
Reflect	1.030	0.407	0.231	0.890	0.406	0.214	0.690	0.442	0.000	0.770	0.413	0.200
Resolution Degradation	1.010	0.402	0.308	0.780	0.393	0.200	0.710	0.440	0.000	0.850	0.413	0.200
Rolling Shutter	0.960	0.388	0.308	0.740	0.387	0.157	0.670	0.433	0.000	0.870	0.413	0.200
Shadow	1.150	0.416	0.231	0.890	0.408	0.200	0.800	0.445	0.000	0.730	0.413	0.200
Snow	1.130	0.423	0.308	0.870	0.408	0.229	0.710	0.453	0.000	0.850	0.413	0.200
Snow	0.890	0.393	0.154	0.760	0.396	0.186	0.670	0.435	0.000	0.880	0.413	0.200
Speckle	0.770	0.379	0.154	0.670	0.380	0.186	0.500	0.422	0.000	0.760	0.413	0.200
Stretch Squish	1.050	0.410	0.308	0.780	0.397	0.200	0.670	0.434	0.000	1.000	0.413	0.200
Target Block	0.860	0.399	0.231	0.700	0.392	0.143	0.680	0.447	0.000	0.940	0.413	0.200
Temporal Jitter	1.050	0.416	0.308	0.880	0.406	0.214	0.750	0.445	0.000	0.770	0.413	0.200
Underexposure	0.780	0.371	0.154	0.630	0.374	0.143	0.520	0.420	0.000	0.840	0.413	0.200
Zoom Blur	0.780	0.371	0.154	0.630	0.374	0.143	0.520	0.420	0.000	0.840	0.413	0.200

Table 17: PLLaVA 7B performance across different question types

Noise Type	CP	FP-S	FP-C	HL	Mean	LR	AR	GPT SBERT Acc	RR	CSR	TR	Mean2	Overall
Clean	0.890 0.391 0.0000.650 0.383 0.1140.570 0.413 0.0000.900 0.144 0.0000.710 0.380 0.0870.340 0.460 0.0000.950 0.453 0.0000.660 0.502 0.0330.460 0.384 0.0000.430 0.384 0.0000.580 0.431 0.0140.690 0.397 0.059												
Bit Error	0.700 0.369 0.0000.530 0.372 0.1290.420 0.399 0.0000.890 0.137 0.0000.570 0.367 0.0980.340 0.462 0.0000.840 0.440 0.0000.590 0.486 0.0000.420 0.382 0.0000.350 0.379 0.0000.520 0.423 0.0000.560 0.385 0.059												
Bright Transform	0.730 0.379 0.0000.580 0.380 0.1000.540 0.408 0.0000.840 0.137 0.0000.610 0.375 0.0760.350 0.470 0.0000.810 0.444 0.0000.620 0.487 0.0000.510 0.384 0.0000.330 0.378 0.0000.520 0.426 0.0000.590 0.391 0.046												
Color Quantized	0.760 0.381 0.0000.620 0.377 0.1140.510 0.398 0.0000.900 0.147 0.0000.660 0.373 0.0870.470 0.459 0.0001.010 0.449 0.0000.760 0.488 0.0000.600 0.389 0.0000.410 0.394 0.0000.640 0.431 0.0000.660 0.392 0.053												
Color Shift	0.650 0.377 0.0000.590 0.380 0.1000.440 0.405 0.0000.820 0.131 0.0000.690 0.374 0.0760.410 0.463 0.0000.810 0.447 0.0000.600 0.494 0.0330.430 0.379 0.0000.390 0.376 0.0000.530 0.425 0.0140.580 0.390 0.053												
Contrast	0.780 0.390 0.0000.610 0.384 0.1140.510 0.410 0.0000.870 0.138 0.0000.640 0.380 0.0870.300 0.452 0.0000.890 0.449 0.0000.650 0.494 0.0000.590 0.396 0.0000.310 0.376 0.0000.530 0.426 0.0000.620 0.394 0.053												
Defocus Blur	0.630 0.369 0.0000.550 0.374 0.1000.460 0.401 0.0000.690 0.133 0.0000.580 0.368 0.0760.290 0.457 0.0000.820 0.447 0.0000.560 0.499 0.0330.600 0.388 0.0000.370 0.388 0.0000.530 0.431 0.0140.560 0.388 0.053												
Edge Sawtooth	0.680 0.372 0.0000.540 0.377 0.1000.520 0.403 0.0000.850 0.132 0.0000.580 0.371 0.0760.420 0.463 0.0000.880 0.439 0.0000.570 0.389 0.0000.420 0.381 0.0000.540 0.426 0.0000.580 0.388 0.046												
Elastic	0.730 0.379 0.0000.590 0.380 0.1000.530 0.400 0.0000.850 0.139 0.0000.630 0.375 0.0760.410 0.465 0.0000.800 0.443 0.0000.710 0.488 0.0000.540 0.390 0.0000.340 0.375 0.0000.540 0.424 0.0000.610 0.391 0.046												
Flicker	0.720 0.386 0.0000.570 0.376 0.0860.490 0.404 0.0000.770 0.137 0.0000.610 0.374 0.0650.350 0.463 0.0000.800 0.444 0.0000.670 0.494 0.0000.480 0.379 0.0000.330 0.387 0.0000.540 0.432 0.0000.600 0.392 0.039												
Foggy	0.710 0.376 0.0000.550 0.377 0.1140.450 0.397 0.0000.790 0.139 0.0000.580 0.371 0.0870.400 0.415 0.0000.800 0.440 0.0000.670 0.494 0.0000.480 0.379 0.0000.330 0.380 0.0000.530 0.423 0.0000.570 0.387 0.053												
Frame Drop	0.260 0.283 0.0000.220 0.292 0.0860.220 0.313 0.0000.850 0.117 0.0000.250 0.290 0.0650.290 0.451 0.0000.500 0.387 0.0000.380 0.460 0.0000.210 0.327 0.0000.330 0.339 0.0000.360 0.383 0.0000.300 0.319 0.039												
Frame Repeat	0.310 0.319 0.0000.350 0.332 0.0860.340 0.343 0.0000.820 0.137 0.0000.360 0.325 0.0650.370 0.421 0.0000.490 0.387 0.0000.520 0.460 0.0330.320 0.335 0.0000.420 0.389 0.0140.390 0.345 0.046												
Frame Replace	0.820 0.390 0.0770.630 0.381 0.1140.560 0.413 0.0000.790 0.127 0.0000.670 0.378 0.0980.350 0.454 0.0000.900 0.455 0.0000.610 0.498 0.0000.510 0.380 0.0000.450 0.403 0.0000.580 0.436 0.0000.650 0.397 0.059												
Frost	0.720 0.371 0.0770.530 0.372 0.1000.420 0.388 0.0000.730 0.135 0.0000.570 0.366 0.0870.370 0.459 0.0000.730 0.441 0.0000.610 0.485 0.0330.460 0.376 0.0000.360 0.375 0.0000.510 0.421 0.0140.560 0.384 0.059												
Gaussian	0.530 0.352 0.0000.490 0.367 0.1140.440 0.401 0.0000.890 0.141 0.0000.510 0.361 0.0870.300 0.464 0.0000.710 0.435 0.0000.580 0.487 0.0000.470 0.387 0.0000.300 0.382 0.0000.480 0.425 0.0000.510 0.381 0.053												
Glass Blur	0.420 0.342 0.0000.360 0.346 0.0570.190 0.374 0.0000.840 0.132 0.0000.370 0.342 0.0430.300 0.446 0.0000.660 0.418 0.0000.520 0.508 0.0000.310 0.373 0.0000.380 0.374 0.0000.450 0.417 0.0000.400 0.365 0.026												
H265 Artifacts	0.660 0.367 0.0000.570 0.372 0.1140.490 0.410 0.0001.060 0.140 0.0000.600 0.368 0.0870.370 0.461 0.0000.840 0.445 0.0000.520 0.496 0.0000.230 0.354 0.0000.400 0.382 0.0000.540 0.430 0.0140.580 0.388 0.059												
Impulse	0.350 0.324 0.0000.380 0.347 0.0570.220 0.361 0.0001.100 0.132 0.0000.380 0.337 0.0430.300 0.439 0.0000.470 0.409 0.0000.520 0.496 0.0000.260 0.361 0.0000.310 0.363 0.0000.390 0.407 0.0000.390 0.359 0.026												
Jpeg Artifact	0.320 0.327 0.0000.330 0.336 0.0710.270 0.363 0.0000.900 0.135 0.0000.340 0.330 0.0540.210 0.442 0.0000.560 0.423 0.0000.560 0.493 0.0000.420 0.379 0.0000.260 0.340 0.0000.370 0.402 0.0000.360 0.351 0.033												
Motion Blur	0.600 0.362 0.0000.460 0.369 0.1000.380 0.397 0.0000.900 0.139 0.0000.500 0.364 0.0760.350 0.448 0.0000.660 0.423 0.0000.660 0.480 0.0000.480 0.389 0.01430.450 0.503 0.0000.320 0.369 0.0000.370 0.402 0.0140.360 0.355 0.039												
Overexposure	0.320 0.328 0.0000.320 0.338 0.0710.240 0.368 0.0000.920 0.121 0.0000.340 0.334 0.0540.220 0.426 0.0000.480 0.389 0.01430.450 0.503 0.0000.460 0.495 0.0000.460 0.384 0.0000.340 0.390 0.0000.540 0.430 0.0000.570 0.389 0.046												
Poisson	0.690 0.370 0.0000.560 0.378 0.1000.390 0.391 0.0000.950 0.139 0.0000.580 0.371 0.0760.380 0.461 0.0000.860 0.445 0.0000.640 0.495 0.0000.460 0.374 0.0000.340 0.368 0.0000.500 0.415 0.0000.480 0.374 0.046												
Rainy	0.520 0.349 0.0000.450 0.363 0.1000.350 0.389 0.0001.030 0.133 0.0000.480 0.357 0.0760.330 0.443 0.0000.770 0.428 0.0000.560 0.498 0.0000.520 0.405 0.0000.360 0.380 0.0000.530 0.428 0.0000.600 0.391 0.053												
Random Block	0.650 0.365 0.0000.510 0.381 0.1000.340 0.390 0.0000.940 0.139 0.0000.540 0.371 0.0760.330 0.464 0.0000.750 0.441 0.0000.410 0.487 0.0000.520 0.386 0.0000.350 0.394 0.0400.470 0.430 0.0140.520 0.389 0.053												
Reflect	0.680 0.373 0.0000.580 0.374 0.1140.460 0.399 0.0000.870 0.136 0.0000.600 0.369 0.0870.500 0.465 0.0000.820 0.439 0.0000.610 0.498 0.0000.510 0.392 0.0000.400 0.384 0.0000.570 0.429 0.0000.600 0.388 0.053												
Resolution Degrade	0.620 0.370 0.0000.540 0.379 0.1000.530 0.410 0.0001.160 0.140 0.0000.590 0.373 0.0760.290 0.437 0.0000.700 0.431 0.0000.540 0.502 0.0000.520 0.398 0.0000.400 0.387 0.0000.500 0.425 0.0000.560 0.390 0.046												
Rolling Shutter	0.630 0.367 0.0000.520 0.368 0.1000.390 0.388 0.0000.870 0.134 0.0000.540 0.362 0.0760.340 0.457 0.0000.810 0.436 0.0000.890 0.448 0.0000.670 0.495 0.0000.620 0.388 0.0000.440 0.385 0.0000.540 0.426 0.0000.550 0.382 0.046												
Shadow	0.830 0.388 0.0000.620 0.378 0.1140.500 0.402 0.0000.870 0.139 0.0000.870 0.375 0.0870.420 0.452 0.0000.890 0.448 0.0000.890 0.448 0.0000.830 0.448 0.0000.550 0.488 0.0000.390 0.384 0.0000.530 0.429 0.0000.600 0.392 0.053												
Snow	0.810 0.381 0.0770.580 0.379 0.1000.470 0.410 0.0000.810 0.137 0.0000.630 0.375 0.0870.290 0.463 0.0000.830 0.448 0.0000.750 0.429 0.0000.470 0.496 0.0000.470 0.383 0.0000.260 0.369 0.0000.430 0.418 0.0000.470 0.376 0.053												
Speckle	0.530 0.355 0.0000.450 0.359 0.1140.360 0.390 0.0000.780 0.144 0.0000.480 0.356 0.0870.330 0.453 0.0000.680 0.433 0.0000.560 0.508 0.0000.300 0.358 0.0000.360 0.390 0.0000.460 0.425 0.0000.440 0.373 0.046												
Stretch Squish	0.460 0.341 0.0000.480 0.356 0.1140.390 0.393 0.0000.850 0.143 0.0000.420 0.349 0.0760.270 0.455 0.0000.810 0.438 0.0000.510 0.482 0.0330.470 0.378 0.0000.360 0.384 0.0000.490 0.422 0.0140.520 0.376 0.059												
Target Block	0.640 0.369 0.0000.490 0.354 0.1000.310 0.390 0.0000.820 0.137 0.0000.520 0.356 0.0870.300 0.436 0.0000.790 0.425 0.0000.520 0.490 0.0330.460 0.378 0.0000.400 0.372 0.0000.500 0.415 0.0140.550 0.372 0.046												
Temporal Jitter	0.650 0.355 0.0000.520 0.355 0.0860.440 0.383 0.0000.980 0.134 0.0000.560 0.353 0.0650.300 0.436 0.0000.900 0.451 0.0000.540 0.490 0.0000.540 0.394 0.0000.510 0.428 0.0000.580 0.389 0.059												
Underexposure	0.730 0.382 0.0000.570 0.376 0.1290.480 0.399 0.0000.900 0.141 0.0000.610 0.372 0.0980.270 0.463 0.0000.900 0.451 0.0000.540 0.490 0.0000.590 0.390 0.0000.300 0.369 0.0000.510 0.428 0.0000.580 0.389 0.059												
Zoom Blur	0.450 0.338 0.0000.390 0.350 0.1000.290 0.374 0.0000.790 0.136 0.0000.410 0.342 0.0760.240 0.416 0.0000.630 0.408 0.0000.500 0.484 0.0000.260 0.344 0.0000.340 0.369 0.0000.410 0.402 0.0000.420 0.362 0.046												

Table 18: Qwen2.5-VL 3B performance across different question types

[illegible]

Table 19: Video-ChatGPT performance across different question types

Noise Type	CP	FP-S	FP-C	HL	Mean	LR	AR	RR	CSR	TR	Mean2	Overall
Clean	0.790 0.396 0.462	0.760 0.411 0.500	0.690 0.445 0.200	0.340 0.154 0.333	0.750 0.404 0.467	0.590 0.496 0.333	0.930 0.477 0.571	1.060 0.572 0.400	0.640 0.414 0.000	0.780 0.441 0.560	0.840 0.477 0.471	0.780 0.427 0.467
Bit Error	0.850 0.399 0.385	0.780 0.411 0.500	0.590 0.442 0.133	0.290 0.148 0.333	0.760 0.404 0.457	0.540 0.498 0.222	0.860 0.461 0.571	1.110 0.583 0.400	0.840 0.418 0.000	0.800 0.444 0.640	0.860 0.478 0.486	0.800 0.427 0.474
Bright Transform	0.810 0.400 0.385	0.740 0.409 0.471	0.650 0.451 0.133	0.310 0.154 0.000	0.740 0.404 0.435	0.520 0.494 0.111	0.890 0.465 0.571	1.140 0.584 0.400	0.850 0.408 0.000	0.740 0.444 0.600	0.840 0.477 0.457	0.770 0.427 0.454
Color Quantized	0.830 0.391 0.308	0.770 0.410 0.457	0.680 0.452 0.067	0.320 0.151 0.333	0.770 0.403 0.413	0.530 0.497 0.222	0.810 0.463 0.571	1.050 0.584 0.400	0.700 0.418 0.000	0.810 0.441 0.600	0.810 0.476 0.471	0.790 0.426 0.441
Contrast	0.760 0.388 0.308	0.790 0.408 0.514	0.650 0.450 0.200	0.350 0.159 0.000	0.760 0.402 0.446	0.440 0.496 0.222	0.930 0.467 0.571	1.140 0.590 0.467	0.750 0.415 0.000	0.790 0.441 0.600	0.830 0.478 0.500	0.790 0.426 0.474
Defocus Blur	0.790 0.389 0.231	0.770 0.409 0.529	0.630 0.446 0.200	0.270 0.158 0.000	0.750 0.402 0.457	0.470 0.498 0.222	0.900 0.460 0.571	1.200 0.586 0.500	0.700 0.406 0.000	0.810 0.442 0.520	0.840 0.475 0.486	0.780 0.425 0.474
Edge Sawtooth	0.780 0.384 0.308	0.760 0.406 0.486	0.670 0.442 0.230	0.350 0.155 0.333	0.730 0.400 0.457	0.500 0.486 0.222	0.900 0.461 0.571	1.080 0.579 0.467	0.770 0.411 0.000	0.740 0.441 0.560	0.820 0.474 0.486	0.770 0.423 0.441
Elastic	0.770 0.394 0.462	0.790 0.407 0.690	0.441 0.200 0.210	0.151 0.333 0.730	0.400 0.457	0.500 0.495 0.222	0.800 0.454 0.571	0.970 0.576 0.400	0.630 0.388 0.000	0.770 0.438 0.600	0.770 0.470 0.471	0.750 0.423 0.447
Flicker	0.790 0.388 0.231	0.760 0.405 0.471	0.640 0.443 0.133	0.400 0.154 0.000	0.770 0.403 0.435	0.470 0.500 0.000	0.880 0.470 0.571	1.130 0.585 0.433	0.640 0.396 0.000	0.770 0.440 0.520	0.800 0.477 0.429	0.760 0.422 0.441
Foggy	0.800 0.392 0.385	0.710 0.403 0.471	0.620 0.444 0.133	0.230 0.154 0.000	0.710 0.399 0.413	0.490 0.483 0.000	0.890 0.463 0.571	1.120 0.585 0.400	0.720 0.408 0.000	0.740 0.435 0.480	0.810 0.473 0.443	0.750 0.422 0.441
Frame Drop	0.540 0.335 0.231	0.630 0.382 0.486	0.550 0.419 0.267	0.610 0.152 0.333	0.610 0.371 0.435	0.470 0.480 0.111	0.600 0.418 0.571	0.810 0.558 0.400	0.470 0.348 0.000	0.690 0.435 0.480	0.650 0.450 0.414	0.630 0.397 0.441
Frame Repeat	0.760 0.376 0.231	0.830 0.406 0.500	0.660 0.433 0.200	0.160 0.148 0.000	0.770 0.395 0.435	0.480 0.490 0.222	0.880 0.470 0.571	0.980 0.572 0.433	0.570 0.385 0.000	0.680 0.431 0.520	0.750 0.469 0.457	0.770 0.418 0.454
Frame Replace	0.830 0.390 0.385	0.790 0.412 0.500	0.630 0.444 0.267	0.420 0.150 0.333	0.770 0.403 0.457	0.590 0.498 0.222	0.890 0.465 0.571	1.110 0.581 0.433	0.890 0.416 0.000	0.790 0.443 0.560	0.880 0.478 0.486	0.800 0.426 0.467
Gaussian	0.700 0.380 0.231	0.720 0.402 0.457	0.630 0.444 0.133	0.260 0.153 0.000	0.710 0.396 0.402	0.530 0.497 0.000	0.740 0.447 0.571	0.990 0.577 0.400	0.790 0.397 0.000	0.740 0.440 0.560	0.760 0.469 0.429	0.730 0.419 0.434
Gaussian Blur	0.680 0.377 0.308	0.700 0.394 0.486	0.620 0.442 0.133	0.290 0.158 0.333	0.690 0.388 0.467	0.450 0.491 0.222	0.710 0.445 0.571	0.970 0.568 0.400	0.720 0.397 0.500	0.740 0.429 0.520	0.760 0.463 0.457	0.710 0.412 0.447
Grayscale	0.790 0.388 0.538	0.760 0.409 0.486	0.670 0.444 0.200	0.350 0.156 0.333	0.690 0.392 0.467	0.550 0.496 0.111	0.960 0.466 0.571	1.170 0.580 0.400	0.740 0.411 0.000	0.810 0.440 0.600	0.870 0.476 0.457	0.800 0.425 0.474
H265 Artifacts	0.720 0.379 0.385	0.720 0.401 0.443	0.540 0.438 0.067	0.320 0.150 0.333	0.690 0.394 0.402	0.470 0.488 0.111	0.790 0.450 0.571	1.060 0.576 0.433	0.750 0.395 0.000	0.740 0.435 0.560	0.780 0.468 0.457	0.730 0.417 0.434
Impulse	0.560 0.356 0.231	0.630 0.385 0.443	0.560 0.428 0.133	0.290 0.149 0.000	0.610 0.378 0.391	0.330 0.484 0.000	0.590 0.424 0.571	0.830 0.555 0.433	0.530 0.369 0.000	0.780 0.434 0.560	0.660 0.454 0.443	0.630 0.402 0.434
Jpeg Artifact	0.770 0.387 0.462	0.750 0.405 0.543	0.550 0.435 0.133	0.310 0.147 0.000	0.720 0.398 0.467	0.460 0.497 0.222	0.860 0.468 0.571	1.070 0.580 0.433	0.800 0.412 0.000	0.740 0.439 0.560	0.800 0.476 0.471	0.750 0.422 0.474
Moire Blur	0.700 0.375 0.308	0.680 0.397 0.471	0.620 0.431 0.133	0.340 0.151 0.000	0.670 0.398 0.413	0.470 0.491 0.111	0.790 0.439 0.571	0.910 0.577 0.400	0.580 0.386 0.000	0.740 0.437 0.560	0.730 0.464 0.443	0.740 0.434 0.434
Overexposure	0.620 0.353 0.308	0.660 0.390 0.443	0.540 0.428 0.133	0.320 0.158 0.000	0.630 0.381 0.402	0.360 0.487 0.000	0.640 0.432 0.222	0.860 0.458 0.571	0.880 0.458 0.571	0.880 0.458 0.571	0.880 0.458 0.571	0.880 0.458 0.571
Poisson	0.770 0.388 0.154	0.730 0.405 0.443	0.650 0.444 0.200	0.310 0.158 0.000	0.720 0.399 0.380	0.460 0.493 0.000	0.890 0.460 0.571	0.950 0.568 0.433	0.770 0.416 0.000	0.790 0.447 0.520	0.800 0.474 0.429	0.750 0.423 0.414
Rainy	0.770 0.383 0.385	0.710 0.401 0.486	0.630 0.441 0.133	0.350 0.148 0.000	0.710 0.395 0.435	0.470 0.497 0.111	0.900 0.462 0.571	1.090 0.585 0.467	0.640 0.401 0.000	0.790 0.444 0.640	0.810 0.477 0.500	0.750 0.421 0.474
Random Block	0.750 0.385 0.385	0.760 0.408 0.514	0.580 0.443 0.133	0.310 0.153 0.333	0.730 0.400 0.467	0.480 0.494 0.444	0.860 0.463 0.571	1.070 0.582 0.400	0.900 0.420 0.000	0.820 0.437 0.560	0.830 0.475 0.486	0.760 0.423 0.474
Reflect	0.810 0.389 0.385	0.750 0.406 0.500	0.610 0.445 0.133	0.420 0.149 0.333	0.750 0.400 0.446	0.520 0.500 0.333	0.990 0.469 0.714	1.110 0.583 0.433	0.720 0.412 0.000	0.810 0.447 0.640	0.860 0.480 0.514	0.790 0.425 0.474
Resolution Degraded	0.840 0.388 0.385	0.790 0.409 0.429	0.660 0.445 0.267	0.230 0.141 0.000	0.760 0.401 0.413	0.460 0.498 0.000	0.880 0.464 0.571	1.000 0.577 0.367	0.700 0.404 0.000	0.740 0.433 0.520	0.790 0.471 0.400	0.780 0.423 0.428
Rolling Shutter	0.790 0.390 0.462	0.780 0.408 0.529	0.650 0.446 0.267	0.420 0.152 0.000	0.760 0.401 0.478	0.500 0.498 0.222	0.930 0.475 0.571	1.110 0.582 0.467	0.730 0.403 0.000	0.830 0.444 0.600	0.860 0.479 0.500	0.800 0.426 0.493
Shadow	0.810 0.396 0.385	0.780 0.409 0.471	0.670 0.437 0.133	0.320 0.151 0.333	0.770 0.403 0.435	0.470 0.488 0.333	1.000 0.479 0.571	1.110 0.584 0.433	0.810 0.409 0.000	0.730 0.442 0.560	0.840 0.478 0.486	0.790 0.426 0.461
Snow	0.730 0.377 0.385	0.720 0.401 0.457	0.620 0.441 0.133	0.260 0.153 0.333	0.700 0.393 0.424	0.370 0.493 0.222	0.840 0.459 0.571	1.040 0.576 0.467	0.650 0.397 0.000	0.710 0.436 0.560	0.750 0.470 0.486	0.720 0.418 0.461
Speckle	0.770 0.383 0.308	0.760 0.401 0.500	0.640 0.440 0.200	0.320 0.151 0.333	0.740 0.395 0.457	0.600 0.491 0.222	0.860 0.452 0.571	1.010 0.570 0.433	0.700 0.398 0.000	0.760 0.440 0.520	0.800 0.470 0.471	0.760 0.419 0.461
Stretch Squish	0.730 0.375 0.385	0.720 0.402 0.514	0.600 0.431 0.200	0.350 0.145 0.000	0.700 0.393 0.457	0.500 0.494 0.222	0.860 0.452 0.571	1.050 0.576 0.400	0.620 0.393 0.000	0.830 0.444 0.600	0.810 0.471 0.740 0.418 0.467	
Target Block	0.690 0.365 0.385	0.670 0.392 0.514	0.640 0.434 0.200	0.440 0.145 0.000	0.670 0.384 0.457	0.450 0.478 0.111	0.741 0.441 0.571	0.960 0.564 0.400	0.620 0.379 0.000	0.720 0.431 0.600	0.730 0.458 0.457	0.700 0.408 0.467
Temporal Jitter	0.810 0.391 0.154	0.760 0.409 0.457	0.640 0.444 0.200	0.340 0.151 0.333	0.750 0.402 0.402	0.380 0.491 0.222	0.940 0.465 0.571	1.050 0.591 0.367	0.700 0.416 0.000	0.800 0.444 0.560	0.810 0.479 0.443	0.770 0.426 0.428
Underexposure	0.670 0.371 0.385	0.700 0.398 0.500	0.580 0.442 0.200	0.370 0.155 0.333	0.680 0.391 0.457	0.440 0.496 0.222	0.710 0.440 0.571	0.880 0.569 0.367	0.530 0.368 0.000	0.730 0.434 0.560	0.700 0.463 0.443	0.700 0.415 0.454
Zoom Blur	0.750 0.387 0.385	0.760 0.408 0.514	0.580 0.443 0.133	0.310 0.153 0.333	0.730 0.400 0.467	0.480 0.494 0.444	0.860 0.463 0.571	1.070 0.582 0.400	0.900 0.420 0.000	0.820 0.437 0.560	0.830 0.475 0.486	0.760 0.423 0.474

Table 20: VideoChat2-HD performance across different question types

Noise Type	CP	FP-S	FP-C	HL	Mean	LR	AR	RR	CSR	TR	Mean2	Overall																								
Clean	1.310 0.445 0.846 1.070 0.441 0.600 0.990 0.483 0.400 0.440 0.141 0.333 1.090 0.437 0.598 0.610 0.519 0.333 1.410 0.511 0.714 1.170 0.593 0.333 1.050 0.419 0.000 1.210 0.470 0.680 1.140 0.503 0.471 1.110 0.459 0.553	1.250 0.440 0.846 1.040 0.440 0.643 0.930 0.470 0.400 0.560 0.149 0.333 1.050 0.434 0.620 0.640 0.524 0.444 1.370 0.503 0.714 1.160 0.586 0.400 1.000 0.420 1.000 1.200 0.473 0.760 1.130 0.502 0.571 1.080 0.456 0.605	1.320 0.441 0.846 1.020 0.439 0.629 0.940 0.478 0.400 0.420 0.142 0.333 1.050 0.434 0.620 0.610 0.524 0.333 1.340 0.501 0.857 1.140 0.593 0.367 1.010 0.429 0.500 1.170 0.470 0.680 1.110 0.503 0.529 1.070 0.456 0.586	1.340 0.440 0.846 1.030 0.440 0.614 0.980 0.478 0.400 0.530 0.146 0.333 1.070 0.435 0.609 0.650 0.519 0.444 1.340 0.507 0.857 1.270 0.593 0.367 1.070 0.424 0.000 1.200 0.471 0.720 1.150 0.503 0.529 1.100 0.458 0.572	1.270 0.440 0.769 1.010 0.438 0.629 0.930 0.478 0.400 0.420 0.145 0.000 1.070 0.435 0.609 0.610 0.521 0.444 1.320 0.497 0.857 1.190 0.590 0.367 1.020 0.429 0.000 1.170 0.469 0.720 1.110 0.500 0.529 1.050 0.455 0.579	1.300 0.444 0.846 1.040 0.439 0.657 0.990 0.476 0.400 0.440 0.142 0.000 1.070 0.435 0.641 0.690 0.516 0.444 1.250 0.498 0.857 1.200 0.594 0.433 1.060 0.425 0.000 1.120 0.470 0.720 1.090 0.500 0.557 1.080 0.456 0.605	1.250 0.438 0.769 1.000 0.435 0.586 0.980 0.481 0.400 0.650 0.153 0.333 1.040 0.432 0.609 0.570 0.515 0.222 1.340 0.497 0.714 1.200 0.589 0.367 0.990 0.428 0.500 1.070 0.468 0.600 1.060 0.499 0.471 1.060 0.453 0.553	1.230 0.444 0.846 1.020 0.440 0.586 1.000 0.479 0.400 0.480 0.141 0.333 1.060 0.436 0.587 0.590 0.518 0.222 1.290 0.491 0.857 1.200 0.591 0.400 1.010 0.420 1.000 1.140 0.469 0.720 1.080 0.498 0.543 1.060 0.453 0.592	1.330 0.444 0.846 1.020 0.440 0.629 0.970 0.482 0.333 0.400 0.151 0.333 1.060 0.436 0.587 0.600 0.524 0.444 1.240 0.497 0.714 1.170 0.595 0.333 0.960 0.417 0.500 1.150 0.467 0.680 1.100 0.500 0.486 1.080 0.457 0.553	1.290 0.441 0.923 1.010 0.437 0.629 0.970 0.482 0.333 0.400 0.150 0.333 1.040 0.434 0.630 0.660 0.524 0.444 1.240 0.497 0.714 1.170 0.595 0.333 0.930 0.412 0.000 1.170 0.473 0.720 1.090 0.501 0.500 1.060 0.456 0.572	1.240 0.439 0.692 1.010 0.439 0.629 0.950 0.469 0.400 0.450 0.142 0.000 1.030 0.433 0.609 0.670 0.522 0.556 1.290 0.497 0.857 1.180 0.596 0.400 0.860 0.430 0.000 1.190 0.469 0.720 1.100 0.502 0.557 1.060 0.456 0.586	1.020 0.404 0.846 0.830 0.416 0.571 0.740 0.451 0.333 0.690 0.156 0.333 0.860 0.411 0.587 0.540 0.514 0.444 1.070 0.464 0.714 1.080 0.574 0.400 0.830 0.406 0.500 1.060 0.454 0.640 0.960 0.481 0.529 0.900 0.433 0.566	0.970 0.392 0.692 0.790 0.413 0.486 0.860 0.467 0.200 0.560 0.156 0.333 0.830 0.409 0.467 0.560 0.516 0.444 1.090 0.465 0.714 1.050 0.566 0.433 0.830 0.403 0.000 0.870 0.448 0.560 0.900 0.477 0.486 0.860 0.431 0.480	1.270 0.437 0.923 1.070 0.443 0.657 1.060 0.478 0.400 0.450 0.140 0.333 1.080 0.436 0.652 0.630 0.521 0.444 1.430 0.508 0.714 1.220 0.598 0.333 0.890 0.406 0.000 1.190 0.475 0.720 1.130 0.504 0.500 1.110 0.459 0.592	1.180 0.430 0.769 1.010 0.436 0.657 0.910 0.469 0.467 0.310 0.140 0.000 1.010 0.430 0.641 0.650 0.527 0.556 1.270 0.491 0.857 1.140 0.577 0.400 0.790 0.413 0.000 1.170 0.465 0.680 1.070 0.495 0.543 1.040 0.451 0.599	1.060 0.416 0.692 0.890 0.428 0.571 0.750 0.459 0.133 0.770 0.147 0.333 0.910 0.421 0.533 0.550 0.515 0.333 1.150 0.479 0.857 1.200 0.588 0.467 0.930 0.426 0.000 1.010 0.464 0.720 1.000 0.493 0.557 0.940 0.444 0.546	1.130 0.413 0.769 0.900 0.422 0.571 0.870 0.454 0.333 0.610 0.157 0.333 0.940 0.418 0.576 0.520 0.512 0.556 1.140 0.475 0.714 1.090 0.576 0.367 0.830 0.404 0.000 1.040 0.461 0.640 0.970 0.485 0.514 0.960 0.440 0.546	1.310 0.440 0.846 1.040 0.440 0.643 0.930 0.472 0.400 0.550 0.151 0.333 1.070 0.431 0.598 0.600 0.524 0.111 1.140 0.501 0.714 1.230 0.588 0.367 1.020 0.427 0.500 1.160 0.470 0.640 1.100 0.501 0.471 1.090 0.454 0.559	1.090 0.418 0.692 0.940 0.421 0.614 0.880 0.464 0.200 0.680 0.156 0.000 0.950 0.417 0.576 0.680 0.523 0.444 1.080 0.473 0.714 1.140 0.585 0.433 0.960 0.411 0.500 0.990 0.454 0.560 0.990 0.487 0.514 0.970 0.440 0.546	1.000 0.412 0.692 0.850 0.416 0.600 0.780 0.451 0.333 0.880 0.157 0.333 0.880 0.412 0.576 0.630 0.519 0.333 1.040 0.470 0.714 1.130 0.574 0.433 0.850 0.404 0.500 0.980 0.463 0.600 0.930 0.485 0.514 0.900 0.435 0.553	1.150 0.427 0.615 0.970 0.433 0.600 0.890 0.467 0.400 0.550 0.147 0.333 0.990 0.428 0.576 0.630 0.519 0.333 1.230 0.489 0.857 1.140 0.596 0.400 1.010 0.407 0.500 1.130 0.461 0.720 1.070 0.494 0.557 1.020 0.449 0.566	1.120 0.415 0.769 0.920 0.425 0.543 0.810 0.454 0.267 0.530 0.153 0.333 0.930 0.419 0.543 0.490 0.508 0.556 1.110 0.472 0.714 1.110 0.576 0.433 0.850 0.398 0.000 1.070 0.463 0.680 0.970 0.484 0.543 0.950 0.440 0.539	1.230 0.436 0.923 1.010 0.436 0.643 0.880 0.469 0.400 0.420 0.145 0.333 1.010 0.430 0.641 0.600 0.517 0.444 1.320 0.494 0.857 1.140 0.588 0.333 1.020 0.418 0.000 1.160 0.469 0.760 1.090 0.497 0.529 1.040 0.451 0.592	1.030 0.414 0.615 0.870 0.423 0.571 0.740 0.450 0.267 0.760 0.150 0.333 0.880 0.416 0.543 0.580 0.521 0.444 1.300 0.501 0.857 1.190 0.595 0.333 0.940 0.421 0.000 1.100 0.468 0.720 1.080 0.500 0.514 1.070 0.456 0.553	1.220 0.440 0.921 1.030 0.439 0.657 1.000 0.476 0.467 0.630 0.145 0.333 1.050 0.434 0.652 0.610 0.522 0.333 1.300 0.500 0.714 1.260 0.591 0.400 0.940 0.427 0.500 1.120 0.468 0.760 1.100 0.500 0.543 1.080 0.456 0.612	1.210 0.433 0.846 1.000 0.438 0.643 0.950 0.471 0.400 0.440 0.143 0.333 1.020 0.432 0.650 0.640 0.519 0.556 1.340 0.495 0.857 1.210 0.590 0.333 0.910 0.414 0.500 1.130 0.463 0.760 1.090 0.496 0.557 1.050 0.453 0.599	1.210 0.429 0.769 0.970 0.430 0.600 0.970 0.474 0.400 0.660 0.152 0.333 1.010 0.427 0.598 0.640 0.510 0.111 1.320 0.494 0.714 1.230 0.593 0.367 0.910 0.418 0.500 1.040 0.461 0.560 1.040 0.494 0.443 1.030 0.449 0.536	1.210 0.430 0.769 0.950 0.429 0.614 0.970 0.474 0.400 0.400 0.149 0.667 0.980 0.437 0.669 0.650 0.519 0.556 1.250 0.495 0.571 1.140 0.582 0.400 0.950 0.418 0.000 1.040 0.462 0.760 1.030 0.495 0.543 1.010 0.449 0.579	1.370 0.452 0.923 1.060 0.440 0.657 1.010 0.478 0.400 0.480 0.141 0.333 1.090 0.437 0.667 0.950 0.520 0.444 1.300 0.497 0.714 1.250 0.596 0.367 0.930 0.417 0.000 1.210 0.473 0.720 1.150 0.514 1.010 0.458 0.592	1.270 0.437 0.923 1.020 0.439 0.643 0.940 0.472 0.400 0.270 0.143 0.333 1.030 0.434 0.641 0.610 0.522 0.444 1.400 0.504 0.857 1.250 0.593 0.367 0.900 0.422 0.000 1.210 0.470 0.800 1.140 0.502 0.557 1.070 0.456 0.605	1.070 0.425 0.846 0.910 0.429 0.657 0.850 0.466 0.267 0.530 0.143 0.333 0.920 0.424 0.630 0.600 0.522 0.444 1.240 0.487 0.714 1.150 0.587 0.400 0.950 0.417 0.000 1.070 0.465 0.800 1.040 0.495 0.586 0.960 0.447 0.618	1.150 0.416 0.769 0.920 0.421 0.600 0.850 0.468 0.200 0.730 0.155 0.333 0.960 0.419 0.576 0.570 0.516 0.333 1.420 0.480 0.714 1.080 0.585 0.333 0.900 0.403 0.500 1.090 0.461 0.640 1.020 0.489 0.486 0.990 0.442 0.539	1.300 0.445 0.923 0.990 0.434 0.643 0.970 0.470 0.400 0.420 0.145 0.333 1.020 0.431 0.652 0.600 0.519 0.444 1.260 0.501 0.714 1.180 0.580 0.400 0.980 0.420 0.500 1.130 0.469 0.640 1.080 0.498 0.529 1.050 0.453 0.599	1.170 0.429 0.769 0.990 0.430 0.614 1.030 0.475 0.400 0.500 0.146 0.333 1.010 0.426 0.598 0.660 0.520 0.556 1.340 0.494 0.714 1.190 0.581 0.367 0.950 0.416 0.000 1.110 0.461 0.800 1.080 0.493 0.557 1.050 0.449 0.579	1.300 0.443 0.846 1.040 0.441 0.600 0.960 0.478 0.333 0.400 0.147 0.000 1.070 0.437 0.598 0.620 0.520 0.222 1.300 0.499 0.857 1.160 0.587 0.333 0.960 0.423 0.000 1.200 0.471 0.760 1.110 0.500 0.514 1.090 0.458 0.572	1.170 0.423 0.846 0.940 0.431 0.586 0.860 0.466 0.133 0.650 0.151 0.333 0.970 0.425 0.565 0.630 0.511 0.556 1.140 0.475 0.857 1.140 0.573 0.400 0.940 0.430 0.500 1.090 0.461 0.640 0.990 0.485 0.529 0.990 0.444 0.553

Table 21: Video-LLaVA performance across different question types

Noise Type	CP	FP-S	FP-C	HL	Mean	LR	AR	GPT SBERT Acc	RR	CSR	TR	Mean2	Overall																								
Clean	0.990	0.405	0.615	0.980	0.427	0.543	0.730	0.470	0.067	0.150	0.143	0.000	0.930	0.420	0.500	0.460	0.503	0.222	1.070	0.487	0.429	1.000	0.569	0.367	0.680	0.419	0.000	0.860	0.453	0.480	0.860	0.484	0.400	0.920	0.441	0.467	
Bit Error	0.910	0.392	0.462	0.900	0.422	0.543	0.680	0.461	0.067	0.160	0.143	0.000	0.860	0.413	0.478	0.500	0.507	0.222	1.070	0.480	0.429	0.980	0.571	0.367	0.840	0.420	0.000	0.780	0.447	0.480	0.850	0.481	0.400	0.870	0.435	0.454	
Bright Transform	0.930	0.400	0.462	0.910	0.423	0.557	0.680	0.461	0.067	0.180	0.147	0.000	0.870	0.415	0.489	0.480	0.511	0.222	1.040	0.480	0.429	1.060	0.573	0.367	0.650	0.414	0.000	0.850	0.444	0.520	0.860	0.480	0.414	0.880	0.436	0.467	
Color Quantized	0.900	0.394	0.462	0.940	0.426	0.571	0.720	0.461	0.133	0.190	0.146	0.000	0.880	0.416	0.511	0.440	0.507	0.222	1.010	0.481	0.429	0.980	0.572	0.367	0.650	0.414	0.000	0.830	0.444	0.520	0.830	0.481	0.414	0.880	0.437	0.480	
Color Shift	0.910	0.402	0.462	0.890	0.420	0.557	0.690	0.461	0.133	0.210	0.143	0.000	0.850	0.413	0.500	0.400	0.503	0.222	1.140	0.479	0.429	1.030	0.574	0.367	0.620	0.409	0.000	0.820	0.453	0.560	0.860	0.482	0.429	0.870	0.435	0.480	
Contrast	1.010	0.404	0.538	0.950	0.426	0.557	0.720	0.467	0.133	0.160	0.143	0.000	0.920	0.418	0.511	0.430	0.506	0.222	1.100	0.483	0.429	1.070	0.575	0.367	0.770	0.421	0.000	0.850	0.450	0.520	0.880	0.483	0.414	0.920	0.439	0.480	
Defocus Blur	0.980	0.396	0.615	0.900	0.422	0.543	0.690	0.463	0.133	0.150	0.143	0.000	0.870	0.414	0.511	0.500	0.507	0.222	1.010	0.483	0.286	1.020	0.567	0.367	0.620	0.411	0.000	0.830	0.447	0.520	0.840	0.481	0.400	0.870	0.435	0.474	
Edge Sawtooth	0.970	0.399	0.385	0.940	0.421	0.543	0.720	0.459	0.067	0.240	0.144	0.000	0.900	0.413	0.467	0.530	0.507	0.222	1.060	0.479	0.429	0.990	0.569	0.367	0.690	0.404	0.000	0.850	0.448	0.520	0.860	0.479	0.414	0.900	0.435	0.454	
Elastic	1.030	0.400	0.538	1.010	0.424	0.543	0.840	0.465	0.067	0.210	0.143	0.000	0.970	0.415	0.489	0.680	0.505	0.222	1.090	0.477	0.429	1.080	0.567	0.367	0.880	0.412	0.000	0.900	0.453	0.480	0.940	0.481	0.400	0.970	0.436	0.461	
Flicker	0.980	0.402	0.615	0.930	0.422	0.557	0.680	0.460	0.067	0.260	0.148	0.000	0.890	0.415	0.511	0.480	0.502	0.222	1.110	0.484	0.429	1.020	0.573	0.367	0.590	0.415	0.000	0.860	0.451	0.520	0.860	0.482	0.414	0.890	0.436	0.480	
Foggy	0.970	0.406	0.462	0.930	0.426	0.557	0.710	0.462	0.133	0.160	0.141	0.000	0.890	0.418	0.500	0.460	0.499	0.222	1.170	0.482	0.429	1.070	0.573	0.367	0.730	0.408	0.000	0.800	0.448	0.480	0.880	0.480	0.400	0.890	0.438	0.467	
Frame Drop	0.780	0.381	0.231	0.750	0.406	0.500	0.620	0.449	0.133	0.160	0.146	0.000	0.730	0.398	0.435	0.520	0.505	0.222	0.900	0.460	0.286	1.200	0.569	0.367	0.650	0.402	0.000	0.790	0.448	0.440	0.850	0.475	0.371	0.770	0.422	0.414	
Frame Repeat	0.790	0.372	0.462	0.730	0.401	0.471	0.590	0.436	0.067	0.080	0.136	0.000	0.710	0.392	0.424	0.560	0.515	0.222	0.880	0.459	0.286	1.020	0.563	0.367	0.590	0.391	0.000	0.850	0.444	0.520	0.830	0.473	0.400	0.760	0.418	0.421	
Frame Replace	1.070	0.409	0.615	0.940	0.427	0.571	0.690	0.468	0.133	0.190	0.146	0.000	0.920	0.420	0.533	0.500	0.509	0.222	1.040	0.489	0.286	1.040	0.575	0.367	0.840	0.423	0.000	0.820	0.450	0.560	0.870	0.486	0.414	0.910	0.441	0.493	
Frost	0.880	0.398	0.308	0.900	0.419	0.571	0.640	0.462	0.067	0.190	0.143	0.000	0.850	0.412	0.478	0.450	0.505	0.222	1.020	0.475	0.429	1.050	0.569	0.367	0.740	0.408	0.000	0.870	0.447	0.520	0.860	0.478	0.414	0.860	0.433	0.461	
Gaussian	0.920	0.400	0.385	0.890	0.415	0.514	0.670	0.453	0.133	0.110	0.145	0.000	0.840	0.409	0.457	0.430	0.504	0.222	1.020	0.468	0.429	0.980	0.568	0.367	0.700	0.409	0.000	0.770	0.448	0.560	0.810	0.477	0.414	0.840	0.431	0.454	
Glass Blur	0.790	0.371	0.462	0.780	0.410	0.514	0.600	0.446	0.133	0.210	0.147	0.000	0.750	0.400	0.467	0.480	0.499	0.222	1.050	0.454	0.286	0.900	0.562	0.367	0.570	0.406	0.000	0.830	0.446	0.560	0.790	0.471	0.400	0.770	0.423	0.454	
H265 Artifacts	0.840	0.379	0.385	0.920	0.418	0.543	0.540	0.457	0.133	0.180	0.147	0.000	0.860	0.410	0.478	0.580	0.504	0.222	1.010	0.481	0.286	0.930	0.574	0.367	0.580	0.405	0.000	0.650	0.446	0.520	0.730	0.481	0.386	0.840	0.433	0.454	
Impulse	0.720	0.368	0.385	0.780	0.405	0.557	0.600	0.437	0.133	0.150	0.146	0.000	0.750	0.397	0.478	0.580	0.497	0.222	0.760	0.447	0.286	0.930	0.564	0.367	0.570	0.405	0.000	0.730	0.444	0.480	0.750	0.470	0.386	0.760	0.420	0.447	
Jpeg Artifact	0.860	0.391	0.308	0.870	0.418	0.557	0.650	0.455	0.133	0.190	0.151	0.000	0.740	0.395	0.489	0.590	0.507	0.222	0.810	0.447	0.286	0.980	0.564	0.367	0.480	0.390	0.000	0.830	0.440	0.560	0.790	0.468	0.414	0.770	0.418	0.467	
Motion Blur	0.760	0.376	0.385	0.800	0.407	0.529	0.760	0.451	0.133	0.300	0.145	0.000	0.820	0.409	0.478	0.540	0.511	0.222	0.990	0.470	0.429	0.980	0.567	0.367	0.620	0.407	0.000	0.810	0.441	0.480	0.830	0.476	0.400	0.830	0.430	0.454	
Overexposure	0.940	0.403	0.462	0.880	0.419	0.557	0.730	0.458	0.133	0.190	0.145	0.000	0.860	0.412	0.500	0.450	0.510	0.222	1.010	0.465	0.429	1.010	0.566	0.367	0.600	0.413	0.000	0.800	0.443	0.440	0.800	0.470	0.371	0.800	0.422	0.434	
Poisson	0.860	0.388	0.385	0.820	0.412	0.557	0.610	0.444	0.133	0.110	0.149	0.000	0.780	0.403	0.480	0.510	0.508	0.222	0.910	0.463	0.429	1.020	0.566	0.367	0.650	0.411	0.000	0.810	0.442	0.520	0.810	0.476	0.414	0.800	0.426	0.467	
Rainy	0.930	0.398	0.462	0.920	0.423	0.557	0.690	0.462	0.133	0.190	0.145	0.000	0.880	0.414	0.500	0.450	0.506	0.222	1.030	0.485	0.429	0.980	0.572	0.367	0.740	0.430	0.000	0.840	0.453	0.560	0.840	0.486	0.429	0.870	0.437	0.480	
Random Block	0.920	0.392	0.462	0.890	0.421	0.543	0.670	0.458	0.067	0.100	0.142	0.000	0.840	0.412	0.478	0.580	0.505	0.222	1.060	0.486	0.429	1.040	0.567	0.367	0.740	0.416	0.000	0.780	0.448	0.520	0.870	0.482	0.414	0.860	0.434	0.461	
Reflect	0.900	0.400	0.385	0.910	0.423	0.543	0.710	0.459	0.133	0.180	0.149	0.000	0.860	0.414	0.478	0.480	0.508	0.222	1.050	0.472	0.429	0.930	0.572	0.367	0.730	0.420	0.000	0.810	0.449	0.560	0.830	0.481	0.429	0.860	0.435	0.467	
Resolution Degrade	0.930	0.400	0.462	0.870	0.417	0.557	0.680	0.463	0.133	0.210	0.147	0.000	0.840	0.411	0.500	0.560	0.503	0.222	1.050	0.481	0.286	1.030	0.577	0.367	0.590	0.402	0.000	0.780	0.443	0.480	0.830	0.480	0.386	0.860	0.433	0.461	
Rolling Shutter	0.940	0.399	0.462	0.870	0.416	0.529	0.590	0.455	0.133	0.130	0.142	0.000	0.830	0.409	0.478	0.450	0.521	0.222	1.010	0.475	0.429	1.070	0.579	0.367	0.840	0.421	0.000	0.810	0.447	0.480	0.860	0.484	0.400	0.840	0.433	0.454	
Shadow	0.930	0.397	0.385	0.910	0.422	0.543	0.720	0.465	0.133	0.180	0.142	0.000	0.870	0.414	0.478	0.460	0.504	0.222	1.050	0.477	0.429	0.980	0.571	0.367	0.650	0.412	0.000	0.800	0.450	0.520	0.830	0.481	0.414	0.870	0.436	0.461	
Snow	0.950	0.404	0.615	0.940	0.423	0.543	0.800	0.467	0.133	0.180	0.145	0.000	0.900	0.416	0.511	0.420	0.504	0.222	1.050	0.477	0.429	0.980	0.571	0.367	0.680	0.419	0.000	0.900	0.451	0.560	0.860	0.482	0.429	0.900	0.437	0.487	
Speckle	0.880	0.385	0.462	0.840	0.416	0.514	0.700	0.454	0.133	0.130	0.149	0.000	0.810	0.406	0.467	0.560	0.513	0.222	1.010	0.480	0.470	0.286	0.980	0.569	0.367	0.830	0.425	0.000	0.790	0.440	0.520	0.840	0.478	0.386	0.830	0.429	0.447
Stretch Squish	0.740	0.372	0.385	0.820	0.410	0.543	0.700	0.453	0.133	0.210	0.147	0.000	0.780	0.400	0.478	0.490	0.505	0.222	0.850	0.457	0.286	0.950	0.566	0.367	0.520	0.392	0.000	0.770	0.441	0.520	0.770	0.471	0.400	0.790	0.423	0.454	
Target Block	0.940	0.393	0.462	0.920	0.420	0.557	0.700	0.459	0.133	0.180	0.146	0.000	0.880	0.411	0.500	0.540	0.508	0.222	1.030	0.472	0.2																