Beyond Curves and Thresholds - Introducing Uncertainty Estimation to Satisfied User Ratios for Compressed Video

Jingwen Zhu, Hadi Amirpour, Raimund Schatz, Patrick Le Callet, Christian Timmerer

▶ **To cite this version:**

Jingwen Zhu, Hadi Amirpour, Raimund Schatz, Patrick Le Callet, Christian Timmerer. Beyond Curves and Thresholds - Introducing Uncertainty Estimation to Satisfied User Ratios for Compressed Video. 2024 Picture Coding Symposium (PCS), Jun 2024, Taichung, France. pp.1-5, 10.1109/PCS60826.2024.10566451 . hal-04684728

HAL Id: hal-04684728
https://hal.science/hal-04684728

Submitted on 3 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Beyond Curves and Thresholds - Introducing Uncertainty Estimation to Satisfied User Ratios for Compressed Video

Jingwen Zhu*, Hadi Amirpour†, Raimund Schatz‡, Patrick Le Callet*§ and Christian Timmerer†
*Nantes Université, Ecole Centrale Nantes, CNRS, LS2N, UMR 6004, Nantes, France
†Christian Doppler Laboratory ATHENA, Alpen-Adria-Universität, Klagenfurt, Austria
‡AIT Austrian Institute of Technology, Austria
§Institut universitaire de France (IUF)

***Abstract*—Just Noticeable Difference (JND) establishes the threshold between two images or videos wherein differences in quality remain imperceptible to an individual. This threshold, collectively known as the Satisfied User Ratio (SUR), holds significant importance in image and video compression applications, ensuring that differences in quality are imperceptible to the *majority* ($p\%$) of users, known as $p\%$SUR. While substantial efforts have been dedicated to predicting the $p\%$SUR for various encoding parameters (*e.g.*, QP) and quality metrics (*e.g.*, VMAF), referred to as proxies, systematic consideration of the prediction uncertainties associated with these proxies has hitherto remained unexplored.**

In this paper, we analyze the uncertainty of $p\%$SUR through Confidence Interval (CI) estimation and assess the consistency of various Video Quality Metrics (VQMs) as proxies for SUR. The analysis reveals challenges in directly using $p\%$SUR as ground truth for training models and highlights the need for uncertainty estimation for SUR with different proxies.

***Index Terms*—Satisfied User Ratio (SUR), Just Noticeable Difference (JND), VQM**

I. Introduction

In adaptive video streaming, content nowadays is encoded at multiple bitrates to accommodate a wide range of network conditions and end-user device types. The number of bitrates and their values, collectively known as the bitrate ladder, are of utmost importance for the video streaming ecosystem. It is crucial to ensure that perceptually similar bitrates are not added to the bitrate ladder, particularly avoiding the addition of bitrates that are perceptually indistinguishable and would only increase costs without providing tangible benefits. In this context, the concept of Video Wise Just Noticeable Difference (VW-JND) is introduced, which represents the smallest perceptible difference in quality between two visual stimuli that can be perceived by an *individual*. For example, the first VW-JND point indicates the transition from perceptually lossless to perceptually lossy coding.

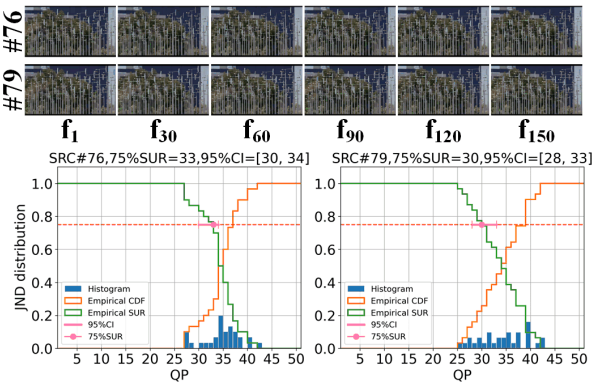

Fig. 1: The sampled frames of SRC#76 and #79 and the corresponding VW-JND distribution in VideoSet [1].

To account for variations in the perception of video quality within the human visual system (HVS) among different viewers, a metric known as the Satisfied User Ratio (SUR) was introduced [1]. SUR quantifies the portion of the population that cannot perceive distortion when a video is compared to a reference at a specific distortion level [1], [2]. This level is referred to as the proxy of the SUR curve, which can be defined using encoding parameters such as QP (Quantization Parameter) or VQMs such as VMAF (Video Multimethod Assessment Fusion) [3].

To investigate SUR on a larger scale, Wang *et al.* [1] proposed an extensive VW-JND dataset known as VideoSet. This dataset includes JND assessments from individuals for 220 source (SRC) video sequences, each with a duration of 5 seconds. Individuals were

tasked with identifying the JND point between the reference video and encoded videos at different QPs (proxy). Previous studies have aimed to predict the SUR curve, specifically targeting p%SUR, representing the point where $p\%$ of users cannot perceive differences between the distorted video and the reference. The most commonly used p is 75 in literatures [1], [2], [4]. For example, [2], [4]–[6] have introduced machine learning (ML) or deep learning (DL) models to predict the SUR curve and 75%SUR using QP as a proxy based on VideoSet. Other research efforts have used VMAF [7]–[11] and bitrate [12] as a proxy to model the SUR.

Upon revisiting the original annotations in VideoSet, we discovered instances where certain SRCs depicted nearly identical scenes, yet their respective 75%SUR values exhibited considerable disparity. As depicted in Fig.1, we present sample frames from SRC#76 and #79, both featuring nearly identical video content, along with the distributions of original VW-JND annotations provided by the individuals. Notably, the 75%SUR QP values for these two SRCs are 33 and 30, respectively, indicating a significant difference. However, when performing the ANOVA [13] analysis on the two distributions, no statistically significant differences emerged. Furthermore, when examining the 95% confidence interval (95%CI) ranges for the 75%SUR values, they appear relatively close, despite the significant disparity in the 75%SUR values themselves. It is worth highlighting that previous works [4]–[6] have employed the 75%SUR as ground truth for training their models, aiming to predict two distinct values for what are essentially the same video contents, which can cause ambiguity for model training. Therefore, it becomes imperative to analyze the uncertainty associated with the SUR derived from subjective tests.

Following the estimation of uncertainty in SUR, another pertinent question arises: Are current VQMs reliable proxies for SUR? A prior study [9] has revealed a significant variance of the first 75%SUR when using VMAF as a proxy in VideoSet, ranging from 75.22 to 99.96. This implies that, for some videos, 75% of viewers perceive no differences until the VMAF score drops to 99.96, while for other videos, they don't perceive differences until the score drops to 75.22. This observation highlights that VMAF does not serve as a perfect proxy for SUR and JND, because the VMAF value is not consistent for different video contents in terms of SUR. To address these problems, we conducted an analysis of SUR, along with its associated uncertainty, for six widely used VQMs in current practice.

In the next section, we will start by estimating the uncertainty of SUR through CI estimation. Then, we will compare the consistency of different VQMs as proxies.

II. Uncertainty Estimation

In this section, we define the SUR for different proxies and introduce a mathematical method for estimating uncertainty in SUR through CI estimation. In statistical analysis, confidence interval can be estimated by Bayesian model [14] using probability distributions [15] or Non-Bayesian model. In this study, we didn't rely on any distribution assumptions regarding the raw distribution of individual VW-JND values. Hence, we use the term empirical SUR (SUR_{emp}) instead of SUR.

For a given video content clip m, assuming that there are VW-JND annotations from N reliable subjects, we denote the VW-JND of these N subjects as J^m, which is defined as follows:

$$J^m = [j_1^m, j_2^m, ..., j_N^m]$$

Here, j_n^m represents the individual annotation of each subject, which can be QP or any other proxy capable of representing the distortion level, such as VMAF. Considering J^m as a discrete random variable, the Probability Mass Function (PMF) of J^m is defined as:

$$p^m(x) = Pr(\text{JND} = x) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left(j_i^m = x\right), \quad (1)$$

where $\mathbf{1}(c)$ is an indicator function that equals to 1 if the specified binary clause c is true. Thus, the empirical Cumulative Distribution Function (CDF) can be calculated from the PMF as follows:

$$\text{CDF}_{\text{emp}}^{\text{m}}(x) = Pr(\text{VW-JND} \leq x) = \sum_{\omega<x} p^m(\omega). \quad (2)$$

In Fig.1, empirical CDFs are represented in orange. Considering SUR_{emp} to depend on the polarity of the chosen proxy, it is defined as follows:

$$\text{SUR}_{\text{emp}}(x) = \begin{cases} 1 - \text{CDF}_{\text{emp}}(x), \text{for case 1} \\ \text{CDF}_{\text{emp}}(x), \text{ for case 2} \end{cases} \quad (3)$$

In case 1, where quality decreases with an increase in the proxy (*e.g.*, using QP as the proxy as shown in Fig.1), the empirical SUR corresponds to the complementary empirical CDF. In contrast, in case 2, such as using VMAF as the proxy, where quality increases with the proxy increases, the empirical SUR is equals to the empirical CDF. Finally, $p\%\text{SUR}_{\text{emp}}$ is defined as:

$$p\%\text{SUR}_{\text{emp}} = \begin{cases} \min\left\{x \,|\text{SUR}_{\text{emp}}\left(x\right) \leqslant p\%\right\}, \text{for case 1,} \\ \max\left\{x \,|\text{SUR}_{\text{emp}}\left(x\right) \leqslant p\%\right\}, \text{for case 2.} \end{cases} \quad (4)$$

Fig.1 showcases the $75\%\text{SUR}_{\text{emp}}$ (represented by the pink point) for the QP proxy.

We can determine $p\%\text{SUR}_{\text{emp}}$ for a specific video content using individual VW-JND annotations collected from a sampled population through subjective test. However, if we were to replicate the same test with a different group of subjects, would we obtain the same $p\%\text{SUR}_{\text{emp}}$ results? Fig.1 has shown that the $75\%\text{SUR}_{\text{emp}}$ for almost same contents can be very different. Therefore, assessing the uncertainty of the $p\%\text{SUR}_{\text{emp}}$ data obtained from the collected datasets is very important.

Using statistical theory, we can estimate the true $p\%$SUR of the entire population based on the $p\%\text{SUR}_{\text{emp}}$ obtained from a sample of N subjects. If we assume that the true $p\%$SUR is equal to s, and we randomly select one subject from the population with their VW-JND denoted as j_n^m, we can calculate the probability of j_n^m being less than s using Eq.(5), in accordance with the definition of the $p\%$SUR.

$$Pr(j_n^m \leqslant s) = \begin{cases} (1-p)\%, \text{for case 1,} \\ p\%, \text{for case 2.} \end{cases} \tag{5}$$

Taking case 2 as an example, we define the random variable A as equal to 1 (event success) when $j_n^m \leqslant s$ and 0 (event failure) when $j_n^m > s$. Consequently, the random variable A conforms to a Bernoulli distribution [16], as presented in Table I.

TABLE I: The random variable A follows a Bernoulli distribution (this table serves as an example for case 2)

Event	A	Probability
$j_n^m \leqslant s$	1 (success)	$p\%$
$j_n^m > s$	0 (fail)	$(1-p)\%$

A subjective test involving N subjects can be understood as N times independently sampling the population. The count of event successes, denoted as X, conforms to a binomial distribution [17]:

$$X \sim B(N, p\%). \tag{6}$$

The PMF of X can be obtained by:

$$f(x, N, p\%) = \Pr(X = x) = C_N^x p\%^x (1-p\%)^{N-x} \tag{7}$$

Where $C_N^x = \frac{N!}{x!(N-x)!}$ and $x \in [0, N]$. Fig.2 shows the PMF of the binomial distribution with parameters $N = 34$ and $p = 75$. When the count of event successes is 26, the probability is calculated as 0.1564. This indicates that if we were to conduct a subjective test with 34 subjects, there is a 15.64% probability that 26 of these subjects would have VW-JND values smaller than or equal to s. If we can determine the lower and upper bounds, denoted as l and u, respectively, such that the cumulative probability between them encompasses approximately 95%, we can confidently assert that there

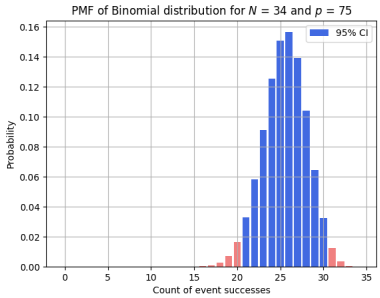

Fig. 2: PMF of binomial distribution for N=34, p=75, and the 95%CI of 75%SUR

is a 95% probability that the number of subjects with $j_n^m \leqslant s$ falls within the interval $[l, u]$.

We employed the Near-symmetric Algorithm [18] to derive l and u for the desired CI. Once l and u are determined, we arrange the values of J^m in ascending order. Subsequently, the CI range for $p\%\text{SUR}_{\text{emp}}$ is defined as $CI_l = J^m_{\text{ordered}}[l]$ and $CI_u = J^m_{\text{ordered}}[u]$, where J^m_{ordered} represents the ordered values of J^m. The 95%CI range can be interpreted as follows: if we were to replicate the subjective test multiple times, there is a 95% probability that the $p\%\text{SUR}_{\text{emp}}$ falls within this range.

III. CI Validation and Comparison

A. CI validation

After computing the 95%CI ranges as presented in Sec. II, we conduct bootstrapping on the original annotations to validate the CI estimation. For each bootstrap sample, we computed $p\%\text{SUR}_{\text{emp}}$ and calculated the percentage of $p\%\text{SUR}_{\text{emp}}$ values that fell within the estimated 95%CI, denoted as Avg CI. We performed 1,000,000 bootstrapping iterations, each with sample sizes of 0.25, 0.5, and 0.75 of the original annotations. Table II shows the Avg CI values for 95%CI estimation on VideoSet in 1080p.

TABLE II: Avg CI with 1,000,000 bootstrapping iteration with different sample sizes

Sample size	0.25	0.5	0.75
Avg CI	0.8331	0.9790	0.9998

In VideoSet, each SRC is annotated by 25 to 34 subjects. Consequently, when the sample size is reduced to 0.25, we observe a decrease in Avg CI. However, on average, the bootstrapped CI closely aligns with the mathematically based CI estimation presented in Sec. II, confirming the validity of our proposed mathematical CI estimation method.

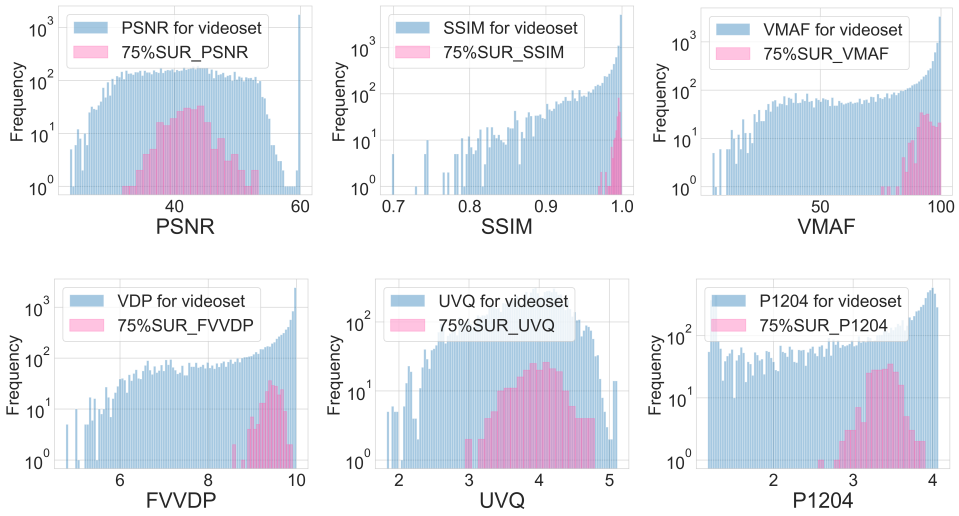

Fig. 3: Distributions of $75\%\text{SUR}_{\text{emp}}$ and the distribution of VQMs on the entire datasets on VideoSet for six VQMs

TABLE III: Benchmark FR-VQM and NR-VQM on 70%SUR, 75%SUR and 80%SUR on VideoSet 1080p for first JND

VQM (min-max)	Mean	COV	Avg 95%CI lower_b	Avg 95%CI upper_b	NAvg 95%CI range	Mean	COV	Avg 95%CI lower_b	Avg 95%CI upper_b	NAvg 95%CI range	Mean	COV	Avg 95%CI lower_b	Avg 95%CI upper_b	NAvg 95%CI range
	80%SUR					75%SUR					70%SUR				
PSNR (23.4-60)	42.6826	0.0825	41.7589	44.7754	0.0824	42.2025	0.0823	41.2626	43.8400	0.0704	41.8065	0.0836	40.9990	43.3017	0.0629
SSIM (0.7-1)	0.9947	**0.0037**	0.9931	0.9972	**0.0135**	0.9939	**0.0043**	0.9919	0.9964	**0.0149**	0.9932	**0.0047**	0.9912	0.9956	**0.0147**
VMAF (5.3-100)	94.5579	0.0399	92.6803	97.1454	0.0471	93.6156	0.0427	91.4264	96.2902	0.0514	92.7761	0.0473	90.6152	95.5342	0.0519
FVVDP (4.8-10)	9.4835	0.0215	9.3727	9.6782	0.0585	9.4287	0.0231	9.3069	9.6044	0.0569	9.3795	0.0248	9.2667	9.5492	0.0541
UVQ (1.8-5)	3.9824	0.0890	3.9448	4.0541	0.0333	3.9644	0.0880	3.9221	4.0291	0.0326	3.9473	0.0884	3.9048	4.0069	0.0311
P.1204 (1.2-4.1)	3.4229	0.0544	3.3152	3.6251	0.1071	3.3723	0.0560	3.2482	3.5450	0.1025	3.3219	0.0606	3.2129	3.4872	0.0948

B. Consistency and uncertainty of VQMs

VQMs capture video quality on a continuous scale, aiming to exhibit a strong correlation with human visual perception. In this section, we employ two types of VQMs: Full Reference VQM (FR-VQM) and No Reference VQM (NR-VQM), as proxies for SUR. Specifically, we investigate the consistency and uncertainty associated with the following VQMs at a specific SUR threshold: four **FR-VQMs**: PSNR, SSIM [19], VMAF [3] (*v0.6.1*), FVVDP [20](*v1.2.0, L_Peak=165.8, contrast = 435, gamma = 2.2, E_ambient = 100, ppd[pix/deg]=60.8, k_ref = 0.005*), and two **NR-VQMs**: P.1204 [21], and UVQ [22] (using *compression_content_distortion* as the output score). These VQMs were applied to the QP range from 0 to 51 for 220 SRC in VideoSet 1080p.

The original VW-JND annotations for each subject J^m are provided in terms of QP values. We then convert these J^m values into their corresponding VQM scores, respectively, and calculate the $p\%\text{SUR}_{\text{emp}}$ for each VQM. We present the mean values of $80\%\text{SUR}_{\text{emp}}$, $75\%\text{SUR}_{\text{emp}}$, and $70\%\text{SUR}_{\text{emp}}$ across the entire dataset in Table III. The mean value of $75\%\text{SUR}_{\text{emp}}$ on VideoSet for VMAF is 93.62, in line with previous studies [8], [9] that suggest a first 75%SUR of approximately 6 for VMAF. It can be observed that $p\%\text{SUR}_{\text{emp}}$ for the first JND increases with higher values of p for all six VQMs.

To measure the consistency of different VQMs in terms of SUR, we calculate the Coefficient Of Variation (COV) for $p\%\text{SUR}_{\text{emp}}$. COV is the ratio of the standard deviation to the mean, serving as an indicator of variability. In this context, it is utilized because different VQMs operate on different scales. A larger COV indicates lower consistency for $p\%\text{SUR}_{\text{emp}}$. Table III reveals that, for p values of 80, 75, and 70, SSIM exhibits the highest level of consistency among the six VQMs.

To visually represent the consistency of different VQMs, as shown in Fig. 3, we plot the distributions using blue bars for the quality scores across the entire dataset (comprising 220 SRCs with QP values ranging from 1 to 51). Additionally, we use pink bars to represent the distributions of $75\%\text{SUR}_{\text{emp}}$ for each SRC. The y-axis uses a logarithmic scale. The distributions of

$75\%\text{SUR}_{\text{emp}}$ for PSNR, UVQ, and P.1204 appear relatively wide compared to the entire dataset. In contrast, the distributions for VMAF and FVVDP are relatively narrower. SSIM exhibits the narrowest distribution range for $75\%\text{SUR}_{\text{emp}}$, in line with its COV values presented in Table III.

We also compute the 95%CI of $p\%\text{SUR}_{\text{emp}}$, as presented in Table III. We calculate the mean of the lower bound and upper bound for each VQM. Notably, the lower bound and the upper bound exhibit the same trend as the mean of $p\%\text{SUR}_{\text{emp}}$. To account for the varying scales, we normalized the CI range using the minimum and maximum values observed for each VQM across the entire VideoSet dataset:

$$Norm(95\%\text{ CI range}) = \frac{95\%CI_u - 95\%CI_l}{\max(VQM) - \min(VQM)}. \quad (8)$$

The mean of the normalized CI range is listed in Table III under the column 'NAvg 95%CI range'. Notably, SSIM exhibits the smallest CI range.

IV. Conclusion

In this paper, we demonstrated the significance of uncertainty estimation for SUR and introduced a method to estimate related CIs. Additionally, we performed a comparative analysis of the consistency and uncertainty associated with different VQMs when serving as SUR proxies. The incorporation of uncertainty estimation not only contributes to a better design of subjective test methodologies but also facilitates a more nuanced interpretation of a SUR prediction model, leading to more robust and informed decision-making in video streaming.

V. Acknowledgment

We would like to express our sincere gratitude to Amazon Prime, Capacités, and Christian Doppler Laboratory ATHENA for their sponsorship. We thank Dr. Sriram Sethuraman and Dr. Kumar Rahul for helpful discussions and insights into this topic.

References

[1] H. Wang, I. Katsavounidis, J. Zhou, J. Park, S. Lei, X. Zhou, M.-O. Pun, X. Jin, R. Wang, X. Wang *et al.*, "Videoset: A large-scale compressed video quality dataset based on jnd measurement," *Journal of Visual Communication and Image Representation*, vol. 46, pp. 292–302, 2017.

[2] J. Zhu, P. Le Callet, A.-F. Perrin, S. Sethuraman, and K. Rahul, "On the benefit of parameter-driven approaches for the modeling and the prediction of satisfied user ratio for compressed video," in *2022 IEEE ICIP*. IEEE, 2022, pp. 4213–4217.

[3] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," *The Netflix Tech Blog*, vol. 6, no. 2, 2016.

[4] H. Wang, I. Katsavounidis, Q. Huang, X. Zhou, and C.-C. J. Kuo, "Prediction of satisfied user ratio for compressed video," in *2018 IEEE ICASSP*. IEEE, 2018, pp. 6747–6751.

[5] H. Wang, X. Zhang, C. Yang, and C.-C. J. Kuo, "Analysis and prediction of jnd-based video quality model," in *2018 Picture Coding Symposium (PCS)*. IEEE, 2018, pp. 278–282.

[6] Y. Zhang, H. Liu, Y. Yang, X. Fan, S. Kwong, and C. J. Kuo, "Deep learning based just noticeable difference and perceptual quality prediction models for compressed video," *IEEE TCSVT*, vol. 32, no. 3, pp. 1197–1212, 2021.

[7] A. Kah, C. Friedrich, T. Rusert, C. Burgmair, W. Ruppel, and M. Narroschke, "Fundamental relationships between subjective quality, user acceptance, and the vmaf metric for a quality-based bit-rate ladder design for over-the-top video streaming services," in *Applications of Digital Image Processing XLIV*, vol. 11842. SPIE, 2021, pp. 316–325.

[8] J. Ozer, "Finding the Just Noticeable Difference with Netflix VMAF," https://streaminglearningcenter.com/codecs/finding-the-just-noticeable-difference-with-netflix-vmaf.html.

[9] H. Amirpour, R. Schatz, and C. Timmerer, "Between two and six? towards correct estimation of jnd step sizes for vmaf-based bitrate laddering," in *2022 14th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2022, pp. 1–4.

[10] J. Zhu, S. Ling, Y. Baveye, and P. Le Callet, "A framework to map vmaf with the probability of just noticeable difference between video encoding recipes," in *2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. IEEE, 2022, pp. 1–5.

[11] J. Zhu, H. Amirpour, R. Schatz, C. Timmerer, and P. L. Callet, "Enhancing satisfied user ratio (sur) prediction for vmaf proxy through video quality metrics," in *2023 IEEE VCIP*, 2023, pp. 1–5.

[12] X. Zhang, C. Yang, H. Wang, W. Xu, and C.-C. J. Kuo, "Satisfied-user-ratio modeling for compressed video," *IEEE TIP*, vol. 29, pp. 3777–3789, 2020.

[13] R. A. Fisher, *Statistical methods for research workers*. Springer, 1992.

[14] N. M. Laird and T. A. Louis, "Empirical bayes confidence intervals based on bootstrap samples," *Journal of the American Statistical Association*, vol. 82, no. 399, pp. 739–750, 1987.

[15] I. Manivannan, "A comparative study of uncertainty estimation methods in deep learning based classification models," 2020.

[16] D. Bertsekas and J. N. Tsitsiklis, *Introduction to probability*. Athena Scientific, 2008, vol. 1.

[17] W. Feller, *An introduction to probability theory and its applications, Volume 2*. John Wiley & Sons, 1991, vol. 81.

[18] G. J. Hahn and W. Q. Meeker, *Statistical intervals: a guide for practitioners*. John Wiley & Sons, 2011, vol. 92.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[20] R. K. Mantiuk, G. Denes, A. Chapiro, A. Kaplanyan, G. Rufo, R. Bachy, T. Lian, and A. Patney, "FovVideoVDP: a visible difference predictor for wide field-of-view video," *ACM Transactions on Graphics*, vol. 40, no. 4, pp. 1–19, Aug. 2021. [Online]. Available: https://dl.acm.org/doi/10.1145/3450626.3459831

[21] R. R. Ramachandra Rao, S. Göring, W. Robitza, A. Raake, B. Feiten, P. List, and U. Wüstenhagen, "Bitstream-based model standard for 4k/uhd: Itu-t p.1204.3 – model details, evaluation, analysis and open source implementation," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, Athlone, Ireland, May 2020.

[22] Y. Wang, J. Ke, H. Talebi, J. G. Yim, N. Birkbeck, B. Adsumilli, P. Milanfar, and F. Yang, "Rich features for perceptual quality assessment of ugc videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 13 435–13 444.