Enhancing Satisfied User Ratio (SUR) Prediction for VMAF Proxy through Video Quality Metrics

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

Enhancing Satisfied User Ratio (SUR) Prediction for VMAF Proxy through Video Quality Metrics

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

***Abstract*—In adaptive video streaming, optimizing the selection of representations for the encoding bitrate ladder has a significant impact on the quality and economics of media delivery. An efficient way to select representations for the bitrate ladder of a given clip is to consider the Satisfied User Ratio (SUR) of the perceived quality of consecutive representations. This ensures that only representations with one Just Noticeable Difference (JND) are encoded and streamed by avoiding encoding similar-quality representations. VMAF (Video Multi-method Assessment Fusion) presently stands as the most commonly utilized quality metric for constructing bitrate ladders. Hence, the precise determination of JND-optimal encoding step-sizes for the VMAF proxy holds paramount importance; nevertheless, this task is intricate and can present considerable challenges.**

In this paper, we evaluate the effectiveness of different Video Quality Metrics (VQMs) in predicting SUR for the VMAF proxy to better capture content-specific characteristics. Our experimental results provide evidence that incorporating VQMs can improve the precision of the SUR prediction for the VMAF proxy. Compared to a state-of-the-art approach that utilizes video complexity metrics, our proposed approach, which incorporates two quality metrics—specifically, VMAF and SSIM calculated at an optimized quantization parameter (QP)—achieves a substantially reduced Mean Absolute Error (MAE) of 1.67. In contrast, the state-of-the-art approach yields an MAE of 2.01. Hence, we recommend using the above quality metrics to improve the accuracy of the SUR prediction for the VMAF proxy.

***Index Terms*—Video quality, SUR, JND, bitrate ladder, VMAF**

I. Introduction

Video streaming platforms like Netflix, YouTube, and Amazon Prime Video have become an essential part of our daily lives. *HTTP Adaptive Streaming* (HAS) is the dominant technique utilized in both live and Video-on-Demand (VoD) streaming applications. It relies on Adaptive Bitrate Streaming (ABR) methods [1], where video content is encoded at multiple bitrate-resolution pairs known as representations. These representations are used to construct a bitrate ladder [2], enabling the dynamic adjustment of video quality according to the viewer's available bandwidth and device type.

Traditionally, a fixed set of representations, such as the HLS bitrate ladder [3], is used for all video content. However, this "one-size-fits-all" approach may not be optimal for different types of videos. To address this, the per-title encoding approaches were introduced, where an optimized bitrate ladder is created for each video content, resulting in improved Quality of Experience (QoE). In per-title encoding [4], [2], [5], various encoding parameters, including resolution, frame rate, and others, are assessed by encoding the videos using all possible combinations of these parameters. Subsequently, an optimized bitrate ladder is constructed by selecting representations from a convex-hull based on the quality measurements of the encoded representations. Video Multi-method Assessment Fusion (VMAF) [6] is widely embraced as an objective quality metric that exhibits a strong correlation with human-perceived quality. It is frequently employed to evaluate the quality of representations and guide the bitrate laddering process [7].

Selecting a subset of representations from the convex-hull is a crucial step in constructing an optimized bitrate ladder. This selection process involves considering various factors, such as available network bandwidth, device capabilities, and perceptual quality metrics like VMAF. While some methods focus on selecting representations based on the probability of clients requesting specific bitrate versions [8], [9], other approaches prioritize the selection of representations to minimize perceptual similarity [10]. These methods aim to avoid including representations in the bitrate ladder that have similar perceptual qualities, as this redundancy may lead to inefficient resource utilization. By diversifying the quality levels of the representations, the bitrate ladder construction aims to provide a wider range of viewing options and enhance the overall streaming experience for users.

The human visual system (HVS) has limitations in detecting small distortions in videos due to psychological and physiological mechanisms. HVS can differentiate only a few discrete-scale distortion levels within a wide range of quality levels. The minimum visual difference that can be perceived by the HVS of *each individual* is defined as the Just Noticeable Difference (JND). The first JND point indicates the transition from perceptually lossless to perceptually lossy coding.

To account for variations in HVS *between viewers*, a metric called Satisfied User Ratio (SUR) was defined, representing the proportion of the population that could distinguish a given distortion level [11], [12]. SUR is defined with Eq.(1):

$$\mathrm{SUR}(x) = 1 - \int_{-\infty}^{+\infty} f(x)dx, \tag{1}$$

where, $f(x)$ corresponds to the Probability Density Function (PDF) that has been appropriately fitted to the individual

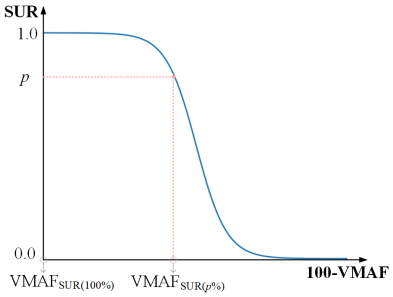

Fig. 1: Demonstration of SUR curve for the VMAF proxy.

distribution of JND obtained from subjective testing. Notably, SUR can be regarded as the inverse of the Cumulative Distribution Function (CDF) associated with the individual JND distribution. In other words, SUR effectively represents the Complementary Cumulative Distribution Function (CCDF). The variable x serves as a proxy for the SUR curve, encompassing encoding parameters (such as QP, CRF), VMAF scores, bitrates, and other relevant factors tailored to diverse application scenarios [13].

In this paper, we focus on VMAF as the proxy of the SUR curve, because of its codec-independence and superior performance, which is highly correlated with human perception. As shown in Fig.1, the transition from perceptually lossless to perceptually lossy coding of the VMAF proxy for $p\%$ of SUR value is expressed as:

$$\Delta\text{VMAF}_{\text{SUR}(p\%)} = \left|\text{VMAF}_{\text{SUR}(100\%)} - \text{VMAF}_{\text{SUR}(p\%)}\right|, \quad (2)$$

where $\text{VMAF}_{\text{SUR}(p\%)}$ is the VMAF proxy where $p\%$ of viewers cannot see the transition and is defined as:

$$\text{VMAF}_{\text{SUR}(p\%)} = \arg\min_{x} \left|\text{SUR}(100-x) - p\%\right|. \quad (3)$$

This paper focuses on investigating content-dependencies of $\Delta\text{VMAF}_{\text{SUR}(p\%)}$ estimation for the VMAF proxy. By leveraging VQMs such as VMAF and SSIM, we aim to explore and leverage the content-specific characteristics that affect $\Delta\text{VMAF}_{\text{SUR}(p\%)}$ of a given content clip.

II. Dataset Description

A comprehensive video quality dataset called VideoSet was introduced in a research publication by Wang *et al.* [14]. The dataset comprises 220 source video sequences with a duration of 5 seconds, featuring frame rates of either 24 fps or 30 fps. These sequences were encoded using the constant quantization parameter (CQP) rate control mode of the H.264/AVC video coding standard [15], with QPs ranging from 1 to 51. The subjective evaluation of JND of individuals was conducted across multiple universities, utilizing 58 dedicated stations for testing.

Although the original JND points were reported for fixed QPs only, we compute the VMAF scores for each encoding and determine the corresponding VMAF values for the first JND. Fig. 2 visually depicts the $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ values for all the 220 contents included in the VideoSet, where the first JND occurs for the $100-p\% = 25\%$ ($p\% = 75\%$) of viewers.

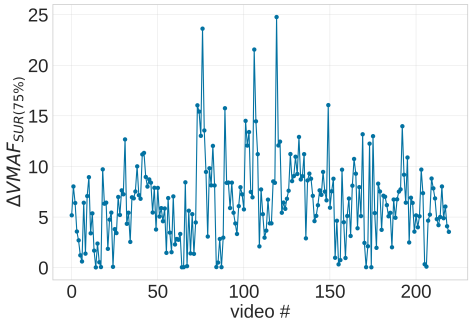

Fig. 2: $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ where the first JND occurs for for the $100-p\% = 25\%$ ($p\% = 75\%$) of viewers in VideoSet [14] for 1080p.

Please note that $p\%$ is selected equal to 75% to align with the literature [12], [14], [16], [17], [18]. Fig. 2 illustrates the substantial diversity in $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ values, highlighting the content-dependent nature of the perceptual differences. For instance, the range of $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ can vary widely, ranging from 0 to 25. This substantial diversity underscores the necessity for content-specific approaches in addressing perceptual differences.

III. State-of-the-art methods

In the existing literature, two sources provide recommendations for determining the sizing of $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ in JND-based bitrate laddering. Jan Ozer [19], in a blog post, reports an interview with an unnamed Netflix employee who suggests a constant step size of 6 without empirical evidence to support this recommendation. On the other hand, Kah *et al.* [20], in a publication, present the findings of a subjective study campaign on the acceptability and perceived quality of video content and propose a constant value of 2 for $\Delta\text{VMAF}_{\text{SUR}(75\%)}$. To further substantiate these two cases, Amirpour *et al.* [16] conducted an analysis by applying the two recommendations ($\Delta\text{VMAF}_{\text{SUR}(75\%)}$ = 2 vs. 6) to VideoSet [14], comparing the outcomes with the subjective JND ground truth provided in the dataset.

In this study, each recommendation was interpreted as a simple linear model that predicts the $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ for each video content in the dataset. The results were subsequently analyzed using common error metrics. It was shown that using the constant value of the 6 rule for $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ performs significantly better than the constant value of 2.

However, these estimators still yield high error levels. The main reason for this discrepancy is the substantial variance of $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ values across different video contents.

To effectively tackle the challenges in the current works, a content-specific framework [16] was developed to estimate $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ by utilizing features extracted from the reference video (see Fig.3(a)). These features encompass the framewise features, including: (i) Spatial Information (SI) [21], (ii) Temporal Information (TI) [21], (iii) Spatial Energy (E) [22], (iv) Temporal Energy (h) [22], (v) Brightness (L), (vi) Colourfulness (c) [23], and (vii) Frame rate (fr).

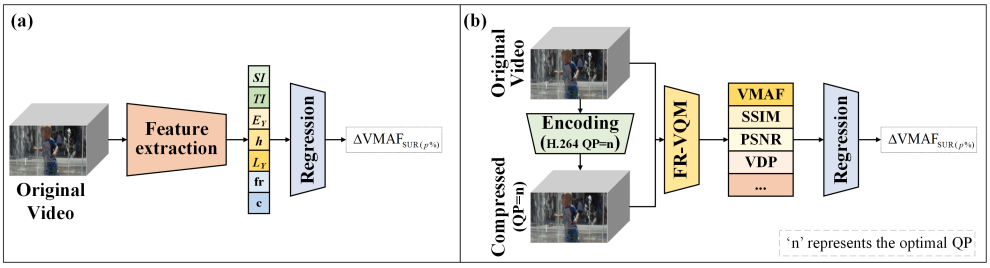

Fig. 3: Comparison of SOTA [16] (a), and our proposed pipelines (b) for $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ prediction.

By considering these features, the framework provided a more comprehensive estimation of $\Delta\text{VMAF}_{\text{SUR}(75\%)}$, taking into account the spatial and temporal aspects, as well as brightness, colourfulness, and frame rate characteristics of the videos, as:

$$\Delta\text{VMAF}_{\text{SUR}(75\%)} = f(SI, TI, E, h, L, fr, c) \quad (4)$$

However, despite the utilization of this approach, the mean absolute error (MAE) was only reduced from 2.73 down to 2.11, in comparison to using the dataset's $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ mean of 6.93 as a constant estimator. Although an improvement was observed, the remaining error levels are still not satisfactory. Therefore, in this paper, we investigate the use of VQMs to improve $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ estimation accuracy.

IV. $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ PREDICTION PIPELINE USING VQMS

Building upon the work of [16], we propose a new pipeline to further improve the estimation of $\Delta\text{VMAF}_{\text{SUR}(75\%)}$. Illustrated in Fig.3(b), we leverage VQMs on distorted video that is compressed with a fixed and optimal QP to enhance the estimation process. The rationale behind this idea is that quality metrics not only inherently incorporate video complexity features but may also provide additional insights to further improve the accuracy of the estimation. By integrating VQMs into the $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ estimation framework, we aim to achieve more refined and precise predictions of the perceptual differences in video content. The complete pipeline consists of three stages: *A)* Incorporated VQMs, *B)* Optimal QP selection, and *C)* Regression based on VQMs on the optimal QP.

A. Incorporated VQMs

We leverage the following Video Quality Metrics (VQMs) to enhance $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ estimation:

- **Peak Signal-to-Noise Ratio (PSNR)**: Measures the difference between an original and distorted video, commonly used for assessing visual quality in compressed or reconstructed videos.
- **Structural Similarity Index (SSIM)** [24]: Evaluates the similarity between reference and distorted images based on luminance, contrast, and structural aspects.
- **Multi-Scale Structural Similarity Index (MS-SSIM)** [25]: Extends SSIM by assessing structural similarity at multiple scales, offering a comprehensive evaluation.

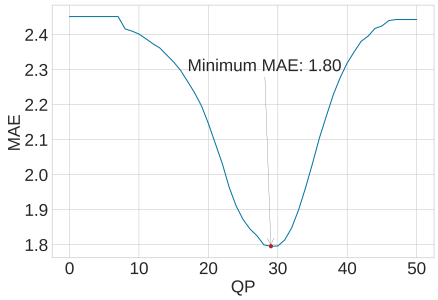

Fig. 4: MAE vs. QP for ridge regression with only VMAF as input features

- **Video Multimethod Assessment Fusion (VMAF)** [**?**]: Considers various visual factors through machine learning and large-scale subjective datasets to predict human-perceived video quality.
- **FovVideoVDP** [26]: Simultaneously addresses spatial, temporal, and peripheral perception aspects, integrating contrast sensitivity, cortical magnification, and contrast masking models.

B. Optimal QP selection

Due to the computational cost associated with encoding videos at multiple QPs and measuring the corresponding quality metrics, we propose to encode the video at a single fixed QP (qp) and calculate the VQMs specifically at that optimal QP.

The ultimate optimal QP is determined using a regression model via a brute-force approach. Among the evaluated QPs, we identify the one yielding the lowest Mean Absolute Error (MAE) after regression. As depicted in Fig. 4, illustrating the test set's MAE across various QPs utilizing a linear regression model with only VMAF as input, it becomes evident that the selection of QP notably influences the model's prediction accuracy. The results indicate that the lowest MAE is achieved at a QP value of 29 for this example.

C. Regression

Our prediction model for estimating $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ can be represented as:

$$\Delta VMAF_{SUR(75\%)} = f(\text{VMAF}_{qp}, \text{SSIM}_{qp}, \text{PSNR}_{qp}, \text{MS-SSIM}_{qp}, \text{VDP}_{qp}), \quad (5)$$

where qp is the optimal QP previously selected.

Initially, a foundational regression model, *i.e.*, ridge regression, with the complexity parameter $\alpha = 0.5$ was applied. Furthermore, a more optimized machine learning regression

model, XGBoost (eXtreme Gradient Boosting) [27], was employed using parameters $n\_estimators = 100$, $max\_depth = 1$, and $booster = gbtree$.

To evaluate the contribution and dependency of these features in predicting $\Delta\text{VMAF}_{\text{SUR}(75\%)}$, we systematically eliminate the least important feature one by one until only one feature remains. We additionally conducted a comparative analysis of feature importance between the video complexity features outlined in [16] and the optimal VQM features for predicting $\Delta\text{VMAF}_{\text{SUR}(75\%)}$. Further details of the outcomes are provided in Section V.

V. Experimental Results

The evaluation of the proposed method is conducted on VideoSet, which has been described in Section II. Prior to commencing the experiment, a preliminary data cleaning process was undertaken. As shown in Fig. 5 the box plot of the $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ of raw data in VideoSet, which shows that there are few data points that deviate significantly from the majority of the dataset. These points have been removed to prevent these extreme values from distorting the overall analysis and interpretation of the data.

The remaining data points were divided into an 80% training set and a 20% test set. Train test split is conducted 20 times using 20 different random seeds. The reported results are the mean values obtained from these multiple splits.

The results are summarized in Table I. When utilizing a fixed $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ of 2 [20], the MAE is 4.29. Increasing the fixed $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ to 6 [19] results in a reduced MAE of 2.59. The state-of-the-art method [16] achieves an MAE of 2.01. The results indicate that our proposed pipeline both ridge and XGBoost regression leads to lower MAE than SOTA. Utilizing XGBoost with a subset of VMAF_{28}, SSIM_{28} can further reduce the MAE to 1.67. The overall results presented in Table I demonstrate that incorporating quality metrics of compressed video with optimal QP can significantly reduce the MAE in $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ modeling.

Fig 6 portrays the feature importance of both the optimal VQM features and the video complexity features outlined in [16]. This importance is derived from the absolute values of the coefficients associated with each feature within the ridge regression model. The results clearly indicate that VMAF_{28} exerts the most substantial influence on the prediction model. Moreover, the VQM features demonstrate notably greater impact on $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ prediction compared to the video complexity features. This finding underscores the notion that

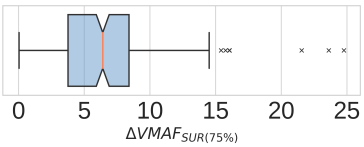

Fig. 5: The box plot of the $\Delta\text{VMAF}_{\text{SUR}(75\%)}$. Data points falling below the lower threshold ($Q_1 - 1.5IQR$) or exceeding the upper threshold ($Q_3 + 1.5IQR$) are considered outliers and have been excluded, as indicated by the 'x' marker.

TABLE I: Experiment results of different $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ estimation models on VideoSet 1080p.

Model	Features	*qp*	MAE
Jan *et al.* [19]	$\Delta\text{VMAF}_{\text{SUR}(75\%)}$= 2	-	4.29
Kah *et al.* [20]	$\Delta\text{VMAF}_{\text{SUR}(75\%)}$= 6	-	2.59
Amirpour *et al.* [16]	SI, TI, E, h, L, c, fr	-	2.01
Ridge regression	vmaf, ssim, vdp, psnr, ms-ssim	30	1.77
Ridge regression	vmaf, ssim, vdp, psnr	30	1.77
Ridge regression	vmaf, ssim, vdp	29	1.78
Ridge regression	vmaf, vdp	29	1.78
Ridge regression	vmaf	29	1.80
XGBoost	vmaf, ssim, vdp, psnr, ms-ssim	28	1.73
XGBoost	vmaf, ssim, vdp, psnr	28	1.74
XGBoost	vmaf, ssim, vdp	28	1.73
XGBoost	vmaf, ssim	28	**1.67**
XGBoost	vmaf	28	1.72

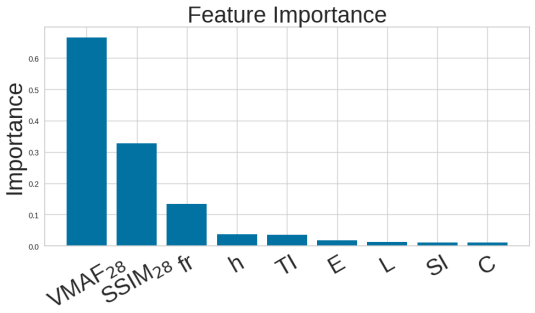

Fig. 6: Feature importance of optimal VQM features and video complexity features in [16]

VQMs not only inherently encompass video complexity features, but also furnish supplementary information, enhancing the predictive accuracy.

VI. Conclusion

In this paper, we focused on modeling the $\Delta\text{VMAF}_{\text{SUR}(75\%)}$. The debate surrounding the use of a fixed $\Delta\text{VMAF}_{\text{SUR}(75\%)}$value of 2 or 6 prompted us to explore the content-dependent nature of $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ as relying on a fixed value can lead to significant MAE. Additionally, while the state-of-the-art approach considers video complexity, our study delved into the influence of various VQMs, including VMAF, SSIM, VDP, PSNR, and MS-SSIM, on predicting the $\Delta\text{VMAF}_{\text{SUR}(75\%)}$ of the first JND. Our findings indicated that calculating these VQMs at a QP of the range 28-30 (depending on the regression model and the selected features) yielded the lowest MAE. We also investigated the interdependency of these features when modeling the $\Delta\text{VMAF}_{\text{SUR}(75\%)}$and discovered that using a subset of VMAF_{28}, SSIM_{28} resulted in the lowest MAE. This finding suggests that the computation of MS-SSIM, VDP and PSNR may not be necessary to accurately predict the $\Delta\text{VMAF}_{\text{SUR}(75\%)}$. Based on these results, we therefore recommend utilizing VMAF and SSIM quality metrics calculated at a QP of 28 to predict the content-dependent $\Delta\text{VMAF}_{\text{SUR}(75\%)}$.

VII. Acknowledge

We would like to express our sincere gratitude to Amazon Prime, Capacités, and Christian Doppler Laboratory ATHENA for their sponsorship.

REFERENCES

[1] Abdelhak Bentaleb, Bayan Taani, Ali C. Begen, Christian Timmerer, and Roger Zimmermann, "A Survey on Bitrate Adaptation Schemes for Streaming Media Over HTTP," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 562–585, 2019.

[2] Hadi Amirpour, Christian Timmerer, and Mohammad Ghanbari, "PSTR: Per-Title Encoding Using Spatio-Temporal Resolutions," in *IEEE ICME*, July 2021, pp. 1–6.

[3] Apple, "HTTP Live Streaming (HLS) Authoring Specification for Apple Devices | Apple Developer Documentation," 2015.

[4] Jan De Cock, Zhi Li, Megha Manohara, and Anne Aaron, "Complexity-Based Consistent-Quality Encoding in the Cloud," in *IEEE ICIP*, Sept. 2016, pp. 1484–1488.

[5] Hadi Amirpour, Mohammad Ghanbari, and Christian Timmerer, "DeepStream: Video Streaming Enhancements using Compressed Deep Neural Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2022.

[6] Zhi Li, Anne Aaron, Ioannis Katsavounidis, Anush Moorthy, and Megha Manohara, "Toward A Practical Perceptual Video Quality Metric," *The Netflix Tech Blog*, vol. 6, pp. 2, 2016.

[7] Angeliki V. Katsenou, Fan Zhang, Kyle Swanson, Mariana Afonso, Joel Sole, and David R. Bull, "VMAF-Based Bitrate Ladder Estimation for Adaptive Streaming," in *PCS*, June 2021, pp. 1–5.

[8] Yuriy A. Reznik, Karl O. Lillevold, Abhijith Jagannath, Justin Greer, and Jon Corley, "Optimal Design of Encoding Profiles for ABR Streaming," in *PV Workshop*, June 2018, pp. 43–47.

[9] Farzad Tashtarian, Abdelhak Bentaleb, Hadi Amirpour, Babak Taraghi, Christian Timmerer, Hermann Hellwagner, and Roger Zimmermann, "LALISA: Adaptive Bitrate Ladder Optimization in HTTP-based Adaptive Live Streaming," in *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, 2023.

[10] Vignesh V Menon, Hadi Amirpour, Mohammad Ghanbari, and Christian Timmerer, "Perceptually-Aware Per-Title Encoding for Adaptive Video Streaming," in *2022 IEEE International Conference on Multimedia and Expo (ICME)*, July 2022, pp. 1–6.

[11] Xinfeng Zhang, Chao Yang, Haiqiang Wang, Wei Xu, and C.-C. Jay Kuo, "Satisfied-User-Ratio Modeling for Compressed Video," *IEEE Transactions on Image Processing*, vol. 29, pp. 3777–3789, 2020, Conference Name: IEEE Transactions on Image Processing.

[12] Jingwen Zhu, Patrick Le Callet, Anne-Flore Perrin, Sriram Sethuraman, and Kumar Rahul, "On the benefit of parameter-driven approaches for the modeling and the prediction of satisfied user ratio for compressed video," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 4213–4217.

[13] Jingwen Zhu, Hadi Amirpour, Vignesh V Menon, Raimund Schatz, and Patrick Le Callet, "Elevating Your Streaming Experience with Just Noticeable Difference (JND)-based Encoding," in *Proceedings of the 2nd Mile-High Video Conference*, Denver CO USA, May 2023, pp. 128–129, ACM.

[14] Haiqiang Wang, Ioannis Katsavounidis, Jiantong Zhou, Jeonghoon Park, Shawmin Lei, Xin Zhou, Man-On Pun, Xin Jin, Ronggang Wang, Xu Wang, Yun Zhang, Jiwu Huang, Sam Kwong, and C. C. Jay Kuo, "VideoSet: A Large-scale Compressed Video Quality Dataset based on JND Measurement," *Journal of Visual Communication and Image Representation*, vol. 46, pp. 292–302, July 2017.

[15] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.

[16] Hadi Amirpour, Raimund Schatz, and Christian Timmerer, "Between Two and Six? Towards Correct Estimation of JND Step Sizes for VMAF-Based Bitrate Laddering," in *QoMEX*, Sept. 2022, pp. 1–4.

[17] Yun Zhang, Huanhua Liu, You Yang, Xiaoping Fan, Sam Kwong, and C. C. Jay Kuo, "Deep Learning Based Just Noticeable Difference and Perceptual Quality Prediction Models for Compressed Video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1197–1212, Mar. 2022, Conference Name: IEEE Transactions on Circuits and Systems for Video Technology.

[18] Haiqiang Wang, Ioannis Katsavounidis, Qin Huang, Xin Zhou, and C.-C. Jay Kuo, "Prediction of Satisfied User Ratio for Compressed Video," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 6747–6751, ISSN: 2379-190X.

[19] Jan Ozer, "Finding the Just Noticeable Difference with Netflix VMAF," https://www.linkedin.com/pulse/finding-just-noticeable-difference-netflix-vmaf-jan-ozer/, 2017.

[20] Andreas Kah, Christopher Friedrich, Thomas Rusert, Christoph Burgmair, Wolfgang Ruppel, and Matthias 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*, San Diego, United States, Aug. 2021, p. 38, SPIE.

[21] ITU-T, "P.910 : Subjective Video Quality Assessment Methods for Multimedia Applications," Nov. 2021.

[22] Vignesh V Menon, Christian Feldmann, Hadi Amirpour, Mohammad Ghanbari, and Christian Timmerer, "VCA: Video Complexity Analyzer," in *Proceedings of the 13th ACM Multimedia Systems Conference*, June 2022, pp. 259–264.

[23] David Hasler and Sabine E. Suesstrunk, "Measuring Colorfulness in Natural Images," in *Human Vision and Electronic Imaging VIII*. June 2003, vol. 5007, pp. 87–95, SPIE.

[24] Zhou 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, Apr. 2004, Conference Name: IEEE Transactions on Image Processing.

[25] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multiscale Structural Similarity for Image Quality Assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, Pacific Grove, CA, USA, 2003, pp. 1398–1402, IEEE.

[26] Rafał K. Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul 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.

[27] Tianqi Chen and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, Aug. 2016, KDD '16, pp. 785–794, Association for Computing Machinery.