



**HAL**  
open science

# A Framework to Map VMAF with the Probability of Just Noticeable Difference between Video Encoding Recipes

Jingwen Zhu, Suiyi Ling, Yoann Baveye, Patrick Le Callet

► **To cite this version:**

Jingwen Zhu, Suiyi Ling, Yoann Baveye, Patrick Le Callet. A Framework to Map VMAF with the Probability of Just Noticeable Difference between Video Encoding Recipes. IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop, Jun 2022, Nafplio, Greece. hal-03788610

**HAL Id: hal-03788610**

**<https://hal.science/hal-03788610v1>**

Submitted on 26 Sep 2022

**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.

# A Framework to Map VMAF with the Probability of Just Noticeable Difference between Video Encoding Recipes

Jingwen Zhu<sup>\*§</sup>, Suiyi Ling<sup>\*§</sup>, Yoann Baveye<sup>\*</sup>, Patrick Le Callet<sup>\*</sup>

<sup>\*</sup>Nantes Université, Ecole Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France

**Abstract**—Just Noticeable Difference (JND) model developed based on Human Vision System (HVS) through subjective studies is valuable for many multimedia use cases. In the streaming industries, it is commonly applied to reach a good balance between compression efficiency and perceptual quality when selecting video encoding recipes. Nevertheless, recent state-of-the-art deep learning based JND prediction model relies on large-scale JND ground truth that is expensive and time consuming to collect. Most of the existing JND datasets contain limited number of contents and are limited to a certain codec (e.g., H264). As a result, JND prediction models that were trained on such datasets are normally not agnostic to the codecs. To this end, in order to decouple encoding recipes and JND estimation, we propose a novel framework to map the difference of objective Video Quality Assessment (VQA) scores, i.e., VMAF, between two given videos encoded with different encoding recipes from the same content to the probability of having just noticeable difference between them. The proposed probability mapping model learns from DCR test data, which is significantly cheaper compared to standard JND subjective test. As we utilize objective VQA metric (e.g., VMAF that trained with contents encoded with different codecs) as proxy to estimate JND, our model is agnostic to codecs and computationally efficient. Throughout extensive experiments, it is demonstrated that the proposed model is able to estimate JND values efficiently.

## I. INTRODUCTION

It is well studied in the cognitive community through psychophysical experiments that our Human Visual System has a limited resolution. Just Noticeable Difference, defined as the minimum amount by which stimulus intensity must be adjusted in order to produce a noticeable difference, is commonly applied to represent this limitation. JND model developed based on HVS study has become the model of choice in many multimedia applications, especially for streaming industries, as it serves as a threshold for video compression platform. With the JND threshold, one can achieve high coding efficiency by eliminating perceptual redundancy without sacrificing the perceived quality to ensure best user experience.

**Subjective study:** To develop robust JND model, subjective studies need to be conducted to collect reliable JND data for the target usecase. Recently, several subjective studies have been conducted. Some JND datasets [1]–[4] were released publicly for Picture Wise JND (PW-JND), and a few for Video Wise JND (VW-JND) [5], [6]. However, most of the existing subjective JND studies applied the ‘one-direction’ JND search protocol, i.e., by only decreasing/increasing the perceived quality of the compressed stimuli.

**Objective JND models:** Since it is time consuming and costly to conduct subjective tests, especially for video contents, objective JND models that predict the JND values automatically are more desirable. One of the most common ways to predict JND is to first predict the Satisfied User Ratio (SUR) [7]–[10]. For a given image/video content, the individual JND of different subjects are different. SUR is the Complementary Cumulative Distribution Function (CCDF) of the group-based JND annotations collected from subjective test [6]. In the literature, the proxy of JND is usually the encoding parameters (e.g., QF, QP or the CRF values that decide the encoding setting). Explicitly, given a certain encoding parameter  $P$ , SUR is defined as the percentage of subjects who are satisfied with the the compressed content. In another word, SUR is the ratio of subjects that cannot perceive any difference between the reference/anchor and the content encoded with  $P$ . JND can be derived from the SUR curve giving a certain threshold, and the most commonly used JND threshold is 75%. Earlier, image JND/SUR models were developed for streaming contents. Liu *et al.* [11] has proposed a deep learning based PW-JND model by formulating the task as a multi-category classification problem; Later, using also CNN, another PW-JND algorithm was presented in [12]; To overcome the challenge of training a deep PW-JND model with limited data, an effective JND model was proposed by Shen [4] using patch-level structural visibility learning; In [10], a Siamese network was utilized to predict the SUR curve, which was later improved in [9]. Apart from the JND models dedicated for images, there are also some models were developed for videos. Zhang *et al.* modeled the SUR curve via Gaussian Processes Regression [7], [13]; A new discrete cosine transform-based energy-reduced JND was demonstrated in [14]. Similar to image SUR model, a video SUR model was proposed by Zhang *et al.* in [15]; Using deep learning, Zhang *et al.* developed the Video Wise Spatial SUR method (VW-SSUR) for predicting the SUR value for compressed video along with the Video Wise Spatial-Temporal SUR (VW-STSUR) to boost the prediction accuracy. Nevertheless, most of the aforementioned SUR prediction methods (or JND approaches based on SUR estimation) suffer from 2 drawbacks: (1) the subjective JND datasets is indispensable. Specifically, large-scale JND datasets are necessary for deep learning-based method for SUR prediction [8]–[10]. (2) these SUR/JND models are not codec agnostic. For instance, the SUR prediction model proposed by Wang *et al.* [7] that predicts the QP value based on H.264 encoding datasets [6] cannot be directly applied for other codec.

To address these issues in existing works, rather than using encoding parameter as the JND proxy, which leads to

<sup>§</sup>Equal contribution

codec dependence, we propose to exploit objective quality metric, (e.g., VMAF [16]) to estimate the JND. Intuitively, if an objective quality metric correlates well with human’s opinions and is able to distinguish non-ambiguous pair confidently, it could be utilized to tell to which point the difference of a pair of stimuli becomes noticeable. Therefore, given a content as anchor with a certain perceived quality, the goal is to estimate the JND by calculating the difference of the objective quality scores between the anchor and another candidate with worse/better quality. With ground truth JND data, we can learn to which extend by increasing/decreasing the objective quality scores, we will notice the difference. Then, we can use this learned objective scores’ difference (e.g., difference in VMAF) as the new proxy for predicting the JND. The anchor is not limited to the Source sequence (SRC), it can also be a Processed Video Sequence (PVS) obtained with different encoding recipes. By increasing the quality of the anchor to a certain point (i.e., the 1<sup>st</sup> JND obtained via increasing the perceived quality), observer may be able to notice the quality got improved. Vice versus, we can also get the JND through by decreasing the quality. In piratical application, given the anchor, the proposed model output the difference of objective quality score  $\Delta obj$  indicating to which point the change of quality become noticeable. One can then estimate objective quality score of the increased/decreased JND point  $obj_{inc-JND}$  or  $obj_{dec-JND}$ , by simply adding/subtracting the  $\Delta obj$  to/from the objective quality of the anchor  $obj_{anchor}$ . With the predicted  $obj_{inc-JND}$  or  $obj_{dec-JND}$ , one can then map/interpolate the optimized codec configuration that generate the PVS with an objective quality score that equals (or close) to  $obj_{inc-JND}$  or  $obj_{dec-JND}$ .

The contribution of this study is two-fold: (1) we collected a JND dataset with a novel two-direction (with both increased and decreased quality) JND search subjective protocol, where each stimulus is also annotated with subjective DCR rating; (2) we presented a framework to map the difference/residual of an objective quality metric (e.g., VMAF) between the anchor and a PVS with another encoding recipes by estimating the probability of the PVS being considered with noticeable quality difference compared to the anchor with a mapping function. It has to be highlighted that the proposed framework takes subjective DCR test annotations as input but not the subjective JND data, as JND subjective test is more resources and time consuming. Interestingly, it is showcased that the calculated difference of objective scores for JND estimation depends on the quality range, where the objective score of the anchor content  $obj_{anchor}$  falls in.

## II. METHODOLOGIES

### A. Subjective studies

1) *Two-directions JND search*: The objective of the two-direction (2-d) JND search is to provide the JND ground truth for validating the proposed JND mapping function. As shown in Fig.1, the proposed 2-d JND search not only the JND with worse quality (*dec-JND*) compared to the anchor (*JND-dec-anchor*) via decreasing the perceived quality, but also the JND with better quality (*inc-JND*) compared to an anchor (*JND-inc-anchor*) by increasing the perceived quality. In our study, given an content, *JND-dec-anchor* is defined as the PVS/SRC with best quality, and *JND-inc-anchor* is the PVS that with the worse perceived quality (points from the same Rate-Quality

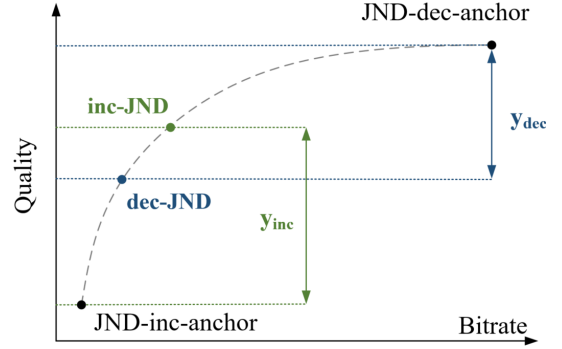


Fig. 1: Illustration of the proposed 2-D JND search

(R-Q) curve). For each direction JND search, the bisection search process was applied, readers could refer to [6] for more details. By considering both direction, we can cover more streaming usecases, where *JND-dec-anchor* and/or *JND-inc-anchor* are/is required.

20 HD SRCs and 10 UHD SRCs were selected from 229 pristine videos from Amazon Prime Video streaming platform based on the content selection strategy proposed in [17] such that the selected contents are with wide-range of different characteristics and ambiguities in terms of quality. Each SRC was encoded with 39 ( $3 \times 13$ ) recipes (3 encoding resolution and each resolution with 13 levels of distortion) using high efficiency video coding (HEVC). The subjective test of 2-d JND search was conducted in a controlled lab environment. A 55-inch calibrated ‘UHD Grundig Finearts 55 FLX 9492 SL’ was employed as the display screen. The viewing distance was set as  $1.5H$  for UHD and  $3H$  for HD contents as recommended in ITU-R BT.2022 [18], where  $H$  is the height of the screened video. Five experts, namely the ‘golden eyes’, participated in the 2-d JND search. They are familiar with encoding artifacts and thus the collected JND data is more accurate. Furthermore, less resources (time and observers) were required. The 2-d JND searches were conducted for each resolution respectively, (including 1080p, 720p and 540p for HD). It is worth noting that only the *JND-dec-anchors* for 1080p of HD contents is the SRC, and the *JND-dec-anchors* for 720p and 540p of HD contents are the PVS with best perceived quality. Similar setting was also applied for the UHD set. For each resolution, the two anchors (i.e., both the anchor for increased and decreased quality) and their corresponding JNDs were selected, which ended up to total 12 ( $4 \times 3$ ) videos that are with significantly different qualities for each content (11 PVS + 1 SRC).

2) *DCR based on 2-d JND search*: The proposed JND mapping model is based on estimating the probability that one stimulus is significant different/similar to another stimulus or not. Thus, subjective quality experiments are also required. To this end, we conducted a subjective quality study utilizing DCR protocol according to ITU-T P.913 [19] with the same contents selected in the 2-d JND searching test. During the test, for each observer, the SRC of each content was compared to 11 corresponding PVSs encoded with different encoding recipes. In total 24 observers participated in the test, and all of them passed the pre-experiment vision check [19] to ensure that they have the correct visual acuity.

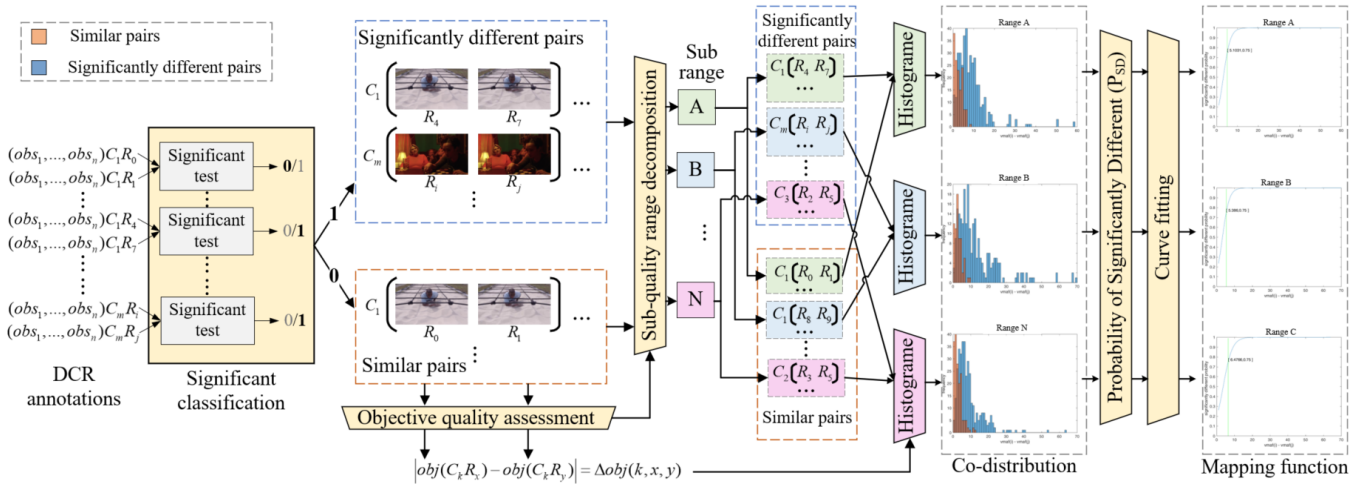


Fig. 2: Diagram of the proposed framework of the JND mapping function using DCR datasets

### B. JND mapping framework

The JND prediction model is depicted in Fig.2 (5 steps):  
**Step 1 - Significant classification:** For a given content  $C_k$ ,  $N$  JNDs (videos encoded from  $C_k$  with different encoding recipes  $R_x$ ) were selected via the proposed 2-D JND search. From the  $N$  selected videos (i.e., increased/decreased JNDs), we formed pairs that took any two videos among the  $N$  videos. At the end, there were in total  $N: \frac{N!}{d!(N-d)!}$  pairs, where  $d = 2$ ,  $N = 12$ . For each content, all the formed pair were classifier into two classes: (1) significantly different pairs; (2) similar pairs. This classification was proceeded through a significant test, e.g., t-test, by using the individual subjective scores (opinion scores from all the observers  $obs_1, \dots, obs_n$ ) from the DCR subjective test described in Section II-A2. Pairs with significant difference were labeled with  $sig = 1$ , otherwise  $sig = 0$ .

**Step 2 - Objective quality assessment:** The quality of each video was evaluated by an certain objective quality assessment model  $obj(\cdot)$  (e.g.,  $obj = \text{VMAF}$  [16]). Afterwards, the residuals/differences of objective quality scores for each pair were calculated:  $\Delta obj(k, x, y) = |obj(C_k R_x) - obj(C_k R_y)|$ .

**Step 3 - Sub-quality range decomposition** Before constructing the mapping function, video pairs  $C_k \{R_x, R_y\}$  were grouped based on which sub-quality range the objective quality score of the anchor falls in. This grouping procedure that divides the entire quality range into a set of continuous sub-ranges is named as *sub-quality range decomposition*. The motivation behind is that the perceptual distances between anchors and JND points differ along the quality range. There are many possible ways to split the entire quality range. One of the most straightforward ways is to divide the whole quality range into equal units, where the size of each bin is 5 in terms of VMAF score. Nevertheless, in this pilot test, such bins division strategy may end out into disconnected distribution (where some bins are empty) or distribution that contains only significantly different/similar pairs due to the limited numbers of tested videos. Therefore, in this study, the sub-ranges were split to have same amount of videos (balanced bins).

**Step 4 - Mapping function design:** For each sub-quality range, the significantly different pairs that belong to this sub-range, were further represented by a histogram based on the  $\Delta obj$  of the pair, i.e., the objective score differences of the pairs:  $h_{dif}(\Delta obj) = \{f_1^{dif}, \dots, f_b^{dif}, \dots, f_B^{dif}\}$ , where

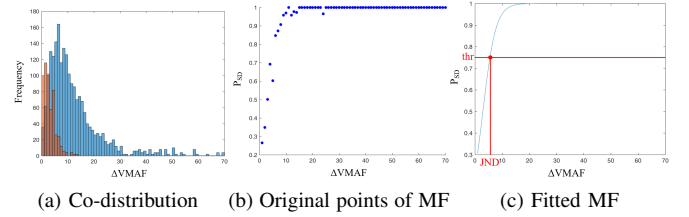


Fig. 3: How to obtain the mapping function (MF)

$f_b^{dif}$  is one bin that accumulates the frequency of pairs with objective quality residual equal to  $b$ , i.e.,  $\Delta obj(k, x, y) = b$ :

$$f_b^{dif} = \sum_k \sum_{\{x,y\}} \mathbf{1}(\Delta obj(k, x, y) = b \ \& \ sig(k, x, y) = 1), \quad (1)$$

where  $\mathbf{1}(c)$  is an indicator function that equals to 1 if the specified binary clause is true; and video pairs  $\{x, y\}$  belong to the current objective quality range. Similarly, the distribution of similar pairs  $h_{sim}(\Delta obj)$  could be obtained with:

$$f_b^{sim} = \sum_k \sum_{\{x,y\}} \mathbf{1}(\Delta obj(k, x, y) = b \ \& \ sig(k, x, y) = 0). \quad (2)$$

Afterwards, for each sub-quality range, we can get the two histograms, i.e.,  $h_{dif}(\Delta obj)$  and  $h_{sim}(\Delta obj)$ , with the same number and interval of bins of  $\Delta obj$ , namely, the co-distribution. In this study, we utilized VMAF, i.e.,  $obj = \text{VMAF}$ . As depicted in Fig.2, the orange bar is the  $h_{sim}(\Delta obj)$  and blue bar is the  $h_{dif}(\Delta obj)$ . It can be observed that similar pairs are distributed in regions where the  $\Delta \text{VMAF}$  is relatively small, and significantly different pairs are distributed in regions with larger  $\Delta \text{VMAF}$ . For two different encoding recipes  $R_x$  and  $R_y$  of the content  $C_k$ , supposing that  $\Delta \text{VMAF}(k, x, y) = b$ , the probability that these two videos are Significantly Different (SD) is calculated as:

$$P_{SD}(b) = \frac{f_b^{dif}}{f_b^{dif} + f_b^{sim}}, \quad (3)$$

where  $f_b^{dif}$  and  $f_b^{sim}$  were obtained from the co-distribution. Fig.3 illustrates how to obtain the mapping function from the co-distribution: for each bin in the co-distribution (Fig.3a), the value of  $P_{SD}$  was obtained through Eq.(3). These  $P_{SD}$  along

with its corresponding  $\Delta\text{VMAF}$  are the original points of the Mapping Function (Fig.3b). The final Mapping Function was then obtained by fitting these original points (Fig.3c).

### Step 5 - From mapping function to JND estimation:

After getting the mapping functions for different quality ranges, we can predict the JND of a given anchor with the corresponding estimated  $\Delta_{obj}$  ( $obj = \text{VMAF}$ ) by fixing a threshold. As illustrated in Fig.3c, if the blue curve is the mapping function of an anchor video  $V_a$ , and the threshold is set to  $thr = 75\%$ , the corresponding value of  $\Delta\text{VMAF}$   $d$  is then considered JND of  $V_a$ . In other words, for a PVS  $V_t$ , which was encoded from the same SRC as  $V_a$ , but encoded with a different recipe, if  $|\text{VMAF}(V_a) - \text{VMAF}(V_t)| > b$ , we assumed that observers can perceive difference between them.

## III. EXPERIMENT AND RESULTS

The proposed JND estimation model was tested on both the JND datasets collected in Section II-A1 and the publicly available JND VideoSet [6]. According to our best knowledge, we are the first to proposed such codec-independent JND estimation model using subjective quality data and objective quality metrics. Thus, as a preliminary study, we did not compared our model with any existing JND models. Before using the collected DCR annotations for the significant classification, three methods including (1) *VQEG HDTV Annex I* [20], (2) *ITU-R BT.1788* [21] and (3) *ITU-R BT.500-12* [22] were used to identify and remove outliers(5 for HD and 3 for UHD). With significant t-test, we obtained 1075 significantly different pairs for HD and 563 significantly different paris for UHD set. We processed the collected HD and UHD datasets separately, due to page limitation, in this study, we only present the results of HD for demonstration. After unequal quality range decomposition (to ensure same number of pairs within each bin), the decomposed sub-quality ranges for HD include (30, 79], (79, 86], (86, 90], (90, 95] and (95, 100]. For video pair  $\{x, y\}$ , it was classified to the sub-quality range  $A$ , if  $\text{VMAF}(x) \in A \vee \text{VMAF}(y) \in A$ .

In our experiments, four fitting functions were considered, with the constraint that the function is monotonic on the full interval: (1) The 5-parameter logistic curve (5-para) [23]; (2) The 4-parameter cubic polynomial (4-para) [23]; (3) The 2-parameter logistic curve (2-para) [23]; (4) The Generalized Linear Model fitting (GLM) (commonly used in psycho-physical studies [24], [25]). To evaluate the performance of JND prediction, two commonly used regression error metrics, Mean Average Error (MAE) and Root Mean Square Error (RMSE) between the ground truth  $y$  and the predicted label  $\hat{y}$ , were calculated.

The results on our JND dataset (Section II-A1) are summarized in Table I for increasing JND and Table II for decreasing JND. It can be observed that by applying the GLM fitting function and a threshold equals to 0.95, our model achieves best performance in predicting both *dec-JND* and *inc-JND*. To validate the generality of the proposed JND estimation model, the performance of JND prediction was also evaluated on the large-scale JND VideoSet [6]. It contains 220 5-second SRCs in 4 resolutions. Each SRC was encoded with H.264 codec with QP valus from 1 to 51. The dataset was labeled with the 1st, 2nd and 3rd JND. The results are summarized in Table III. It can be observed that the best performances are obtained

TABLE I: Performance of JND prediction ( $\Delta\text{VMAF}$ ) for increasing JND for HD contents considering different thresholds  $thr.$  and fitting functions  $fit(\cdot)$

$thr.$	$fit(\cdot)$	5-para	4-para	2-para	GLM
		Mean Absolute Error (MAE)[1, 100]			
0.75		5.1519	4.1997	5.1281	4.8915
0.8		4.1956	3.9765	4.7420	4.4463
0.85		3.6163	3.8181	4.3771	3.9757
0.9		4.2271	3.7447	4.0420	3.6114
0.95		6.7110	3.8777	3.9551	<b>3.4963</b>
		Root Mean Square Error (RMSE) [1, 100]			
0.75		7.2220	5.9346	7.0309	6.7227
0.8		5.9573	5.7270	6.6600	6.2290
0.85		5.1240	5.5412	6.2459	5.7050
0.9		5.3592	5.3984	5.8011	5.1732
0.95		7.4985	5.4512	5.4493	<b>4.8471</b>

TABLE II: Performance of JND prediction ( $\Delta\text{VMAF}$ ) for decreasing JND for HD contents considering different thresholds  $thr.$  and fitting functions  $fit(\cdot)$

$thr$	$fit(\cdot)$	5-para	4-para	2-para	GLM
		Mean Absolute Error (MAE)[1, 100]			
0.75		4.2586	3.4865	4.1996	4.0253
0.8		3.6958	3.3277	3.8736	3.6564
0.85		3.8694	<b>3.2493</b>	3.6509	3.3475
0.9		5.1067	3.3602	3.6679	3.2665
0.95		7.6473	3.6988	4.2052	3.6863
		Root Mean Square Error (RMSE) [1, 100]			
0.75		7.5868	6.2914	7.2717	7.0435
0.8		6.6783	6.1947	7.0217	6.6774
0.85		6.2428	6.1316	6.7698	6.3115
0.9		6.1184	6.1184	6.5555	5.9811
0.95		6.2585	6.2585	6.5525	<b>5.8850</b>

TABLE III: Performance of JND prediction for the first, second, and third JND in **VideoSet** (1080p) considering different thresholds  $thr.$  with GLM fitting functions.

$N^{\text{th}}$	1 <sup>st</sup> JND		2 <sup>nd</sup> JND		3 <sup>rd</sup> JND	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
0.75	3.5026	4.6771	3.5192	4.5879	2.7160	3.6242
0.8	3.1819	4.3186	3.3584	4.3428	2.5840	3.4218
0.85	<b>3.0491</b>	<b>4.1053</b>	<b>3.3242</b>	<b>4.1961</b>	<b>2.5691</b>	<b>3.3247</b>
0.9	3.2415	4.2470	3.5338	4.2794	2.7681	3.4638
0.95	3.8400	4.7528	3.5192	4.5879	3.6060	4.5246

when the threshold is 0.85 for all the three JNDs. Furthermore, the prediction errors using the proposed model on the VideoSet are slightly smaller than ones on our 2-d JND dataset, which further verifies the fact that the proposed framework is generic and codec agnostic.

## IV. CONCLUSION

In this work, a framework that maps the  $\Delta\text{VMAF}$  values to the probability of the JND between videos encoded with different encoding recipes is proposed. Given an anchor video as input, the proposed model take into account which sub-quality-range the anchor video falls in, and out put the estimated JND in terms of  $\Delta\text{VMAF}$ . With the  $\Delta\text{VMAF}$ , one can then select the right encoding recipes. The current JND prediction model could be improved by taking the content characteristics into consideration, e.g., using different mapping functions for different types of contents instead of based on quality range.

## REFERENCES

- [1] L. Jin, J. Y. Lin, S. Hu, H. Wang, P. Wang, I. Katsavounidis, A. Aaron, and C.-C. J. Kuo, "Statistical study on perceived jpeg image quality via mcl-jci dataset construction and analysis," *Electronic Imaging*, vol. 2016, no. 13, pp. 1–9, 2016.
- [2] X. Liu, Z. Chen, X. Wang, J. Jiang, and S. Kwong, "Jnd-pano: Database for just noticeable difference of jpeg compressed panoramic images," in *Pacific Rim Conference on Multimedia*. Springer, 2018, pp. 458–468.
- [3] C. Fan, Y. Zhang, H. Zhang, R. Hamzaoui, and Q. Jiang, "Picture-level just noticeable difference for symmetrically and asymmetrically compressed stereoscopic images: Subjective quality assessment study and datasets," *Journal of Visual Communication and Image Representation*, vol. 62, pp. 140–151, 2019.
- [4] X. Shen, Z. Ni, W. Yang, X. Zhang, S. Wang, and S. Kwong, "Just noticeable distortion profile inference: A patch-level structural visibility learning approach," *IEEE Transactions on Image Processing*, vol. 30, pp. 26–38, 2020.
- [5] H. Wang, W. Gan, S. Hu, J. Y. Lin, L. Jin, L. Song, P. Wang, I. Katsavounidis, A. Aaron, and C.-C. J. Kuo, "Mcl-jcv: a jnd-based h. 264/avc video quality assessment dataset," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 1509–1513.
- [6] 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.
- [7] H. Wang, I. Katsavounidis, Q. Huang, X. Zhou, and C.-C. J. Kuo, "Prediction of satisfied user ratio for compressed video," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6747–6751.
- [8] 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 Transactions on Circuits and Systems for Video Technology*, 2021.
- [9] H. Lin, V. Hosu, C. Fan, Y. Zhang, Y. Mu, R. Hamzaoui, and D. Saupé, "Sur-featnet: Predicting the satisfied user ratio curve for image compression with deep feature learning," *Quality and User Experience*, vol. 5, no. 1, pp. 1–23, 2020.
- [10] C. Fan, H. Lin, V. Hosu, Y. Zhang, Q. Jiang, R. Hamzaoui, and D. Saupé, "Sur-net: Predicting the satisfied user ratio curve for image compression with deep learning," in *2019 eleventh international conference on quality of multimedia experience (QoMEX)*. IEEE, 2019, pp. 1–6.
- [11] H. Liu, Y. Zhang, H. Zhang, C. Fan, S. Kwong, C.-C. J. Kuo, and X. Fan, "Deep learning-based picture-wise just noticeable distortion prediction model for image compression," *IEEE Transactions on Image Processing*, vol. 29, pp. 641–656, 2019.
- [12] T. Tian, H. Wang, L. Zuo, C.-C. J. Kuo, and S. Kwong, "Just noticeable difference level prediction for perceptual image compression," *IEEE Transactions on Broadcasting*, vol. 66, no. 3, pp. 690–700, 2020.
- [13] 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.
- [14] S. Ki, S.-H. Bae, M. Kim, and H. Ko, "Learning-based just-noticeable-quantization-distortion modeling for perceptual video coding," *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3178–3193, 2018.
- [15] X. Zhang, C. Yang, H. Wang, W. Xu, and C.-C. J. Kuo, "Satisfied-user-ratio modeling for compressed video," *IEEE Transactions on Image Processing*, vol. 29, pp. 3777–3789, 2020.
- [16] 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.
- [17] S. Ling, Y. Baveye, D. Nandakumar, S. Sethuraman, and P. Le Callet, "Towards better quality assessment of high-quality videos," in *Proceedings of the 1st Workshop on Quality of Experience (QoE) in Visual Multimedia Applications*, 2020, pp. 3–9.
- [18] I. BT, "General viewing conditions for subjective assessment of quality of sdtv and hdtv television pictures on flat panel displays," *International Telecommunication Union*, 2012.
- [19] I. Union, "Methods for the subjective assessment of video quality audio quality and audiovisual quality of internet video and distribution quality television in any environment," *SERIES P: TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS*, 2016.
- [20] VQEG, "Vqeg hdtv group : Test plan for evaluation of video quality models for use with high definition tv content (draft version 3.1)," <https://vqeg.github.io/software-tools/subjective%20test%20software/matlab-screening/>.
- [21] P. ITU-T RECOMMENDATION, "Subjective video quality assessment methods for multimedia applications," 1999.
- [22] R. I.-R. BT, "Methodology for the subjective assessment of the quality of television pictures," *International Telecommunication Union*, 2002.
- [23] I. Tutorial, "Objective perceptual assessment of video quality: Full reference television," *ITU-T Telecommunication Standardization Bureau*, 2004.
- [24] F. A. Wichmann and N. J. Hill, "The psychometric function: I. fitting, sampling, and goodness of fit," *Perception & psychophysics*, vol. 63, no. 8, pp. 1293–1313, 2001.
- [25] Y. Lee, J. A. Nelder, and Y. Pawitan, *Generalized linear models with random effects: unified analysis via H-likelihood*. Chapman and Hall/CRC, 2018.