Research proposal: Just noticeable difference (JND) and satisfied user ratio (SUR) prediction for compressed video

Jingwen Zhu, Patrick Le Callet

Research proposal: Just Noticeable Difference (JND) and Satisfied User Ratio (SUR) prediction for compressed video

Jingwen Zhu
jingwen.zhu@etu.univ-nantes.fr
Nantes Université
Ecole Centrale Nantes
CNRS, LS2N, UMR 6004
Nantes, France

Patrick Le Callet
patrick.lecallet@univ-nantes.fr
Nantes Université
Ecole Centrale Nantes
CNRS, LS2N, UMR 6004
Nantes, France

ABSTRACT

This paper contains the research proposal of Jingwen ZHU that was presented at the MMSys 2022 doctoral symposium. Just noticeable difference (JND) is the minimum amount of distortion from which human eyes can perceive difference between the original stimuli and distorted stimuli. With the rapid raise of multimedia demand, it is crucial to apply JND into the visual communication systems to use the least resources (*e.g.*, bandwidth and storage) but without damaging the Quality of Experience (QoE) of end-users. In this thesis, we focus on the JND prediction for compressed video to guide the choice of optimal encoding parameters for video streaming service. In this paper, we analyse the limitations of the current JND prediction models and present five main research questions to address these challenges.

CCS CONCEPTS

• **Applied computing**; • **Human-centered computing**;

KEYWORDS

Just noticeable difference (JND), Satisfied User Ratio (SUR), video quality assessment

1 INTRODUCTION

Human Visual System (HVS) cannot perceive small distortions, Just Noticeable Difference (JND) threshold is the minimum amount by which stimulus intensity must be changed to produce a noticeable variation for HVS. Nowadays, with the increasing multimedia demand such as video streaming, JND plays an important role to reduce the resources (*e.g*, bandwidth, storage) consumption without decreasing the Quality of Experience (QoE) for end-users. In

addition, JND has been widely employed in many other vision applications, including digital image/video processing, visual signal restoration/enhancement, and watermarking. To find the JND video in a series of videos with different distortion levels for a given anchor/reference, video with every distortion levels need to be compared with the anchor. Wang *et al.* [24] proposed a binary search method to conduct the JND search which can help to reduce the number of comparison during the JND search. Lin *et al.* [9] proposed two methods to improve the efficiency of binary search: slider and keystroke. However, compared to ordinary subjective tests such as Absolute Category Rating (ACR) or Degradation Category Rating (DCR) [6], subjective JND test is much more time consuming. As subjective test is time/money consuming, it is important to develop JND prediction models.

JND depends on 3 factors: (1) display setting, *e.g.,* viewing distance [17], monitor profiling, *etc.*; (2) subjects; (3) image/video contents [11]. In this thesis, we focus on how different video contents impact the location of JND in the context of video compression. Specifically, when the anchor is the no-compression version, *e.g.,* sources(SRCs) video, the JND is called *1st* JND, and if we take the *1st* JND as the new anchor, the next JND is then the *2nd* JND, etc. In practice, it is far enough to measure the *1st*, *2nd* and *3nd* JND, because the video quality worse than the *3nd* JND is too poor in current internet video streaming applications [24].

The prediction of JND has been investigated in both images [2, 10, 13, 27, 28] and videos [23, 25, 29, 30]. However, most of the current JND prediction methods are application/codec specific. With the rapid development of multimedia compression standards, it is not practical to collect large-scale dataset for every new codec. Therefore it is important to develop codec agnostic JND prediction methods or a general model which could be adapted easily for new codecs. Moreover, current JND prediction models are not practical, because they take the SRC and its corresponding Processed Video Sequences (PVSs) with many different encoding parameters as input. Therefore before applying these models, SRC need to be compressed several times, which is computation and time consuming. In this thesis, we present five main research questions to target the above mentioned issues in the video-wise JND prediction for compressed video. We plan to address them using existing publicly available JND datasets and machine learning methods.

2 RELATED WORK

The JND prediction models in literature can be divided into: (1) Pixel-wise JND models; (2) Sub-band domain JND models; (3) picture-wise JND model and (4) Video-wise JND model. Pixel-wise JND models

predict either a JND map, which concretely contains per pixel JND in a given image [27, 28], or a probability map that contains for each pixel the probability that an average observer will notice a difference between the reference image and the distorted image [16]. Sub-band domain JND models calculate the JND threshold for each sub-band coefficient in transform domain (*e.g.*, Discrete cosine transform (DCT)) [15, 21, 26]. These models are "generic" models which are application-free (*i.e.*, they are not limited to a certain image compression methods *e.g.*, JPEG). However, even if the pixel-wise JND prediction is accurate enough, it is still challenging to mimic HVS for a global JND threshold for images, especially for small spatial or temporal shifts. Moreover, pixel-wise JND prediction models cannot be used directly to guide the image compression. The sub-band domain models consider each sub-band blocks separately, therefore the cross sub-band blocks information is not taken into consideration.

For a better understanding of the global JND for picture, several picture-wise JND datasets have been collected from subjective JND test under different image compression methods and for different image types [3, 7, 14, 19]. For a given content, the JND annotation can vary for different subjects. Satisfied User Ratio (SUR) was introduced in [24] to measure the disagreement between subjects. SUR curve is the Complementary Cumulative Distribution Function (CCDF). For a given distortion level D, the corresponding value of SUR on the SUR curve is the ratio of subjects who cannot perceive any difference between the reference video and the distorted video with distortion level D. By fixing a threshold $p\%$ of SUR, the corresponding distortion level is defined as $p\%$SUR or $p\%$JND. The most common used threshold is 75%. Based on these picture-wise JND datasets, some deep-learning base picture-wise JND prediction models have been proposed. Liu *et al.* [13] first transferred the multi-class classification problem to a binary classification (lossy/lossless) problem and used a sliding window search strategy to predict JND points. The binary classifier is a lossy/lossless predictor based on convolutional neural networks (CNN), which is trained and evaluated on the MCL-JCI [7] dataset, and the sliding window search strategy is designed to find the first JND point using the output of the classifier. SUR-Net [2] is based on a Siamese CNN to predict the SUR value for different distortion levels of the image. The discrete points obtained by the predictor are used to fit the SUR curve by least squares assuming that the JND annotations of a group observers follow normal distribution. The first JND point is derived by the fitted SUR curve with 75% as threshold. Transfer learning and data augmentation are used to avoid over-fitting caused by the small-scale training dataset MCL-JCI [7]. Based on SUR-Net, Lin *et al.* [10] proposed SUR-FeatNet to predict the SUR curve and JND. Maximum likelihood estimation and Anderson-Darling test are used to obtain a more accurate probability distribution model for the JND distribution for different observers. The ground-truths are generated by fitting the SUR curve in order to train the network which is based on Inception-V3 [20]. There are two stages to train the network. During the first stage, it is trained to predict the fixed full-reference (FR)-IQA score of the input distorted image versus the input reference image. While during the second stage, the SUR value will be predicted instead. The least-squares will be used to fit the final SUR curve using the discrete SUR points obtained in the second stage.

Video-wise JND is less investigated compared to picture-wise JND. Two video-wise JND datasets have been collected in [22, 24] under H.264 video compression. Wang *et al.* [23] proposed a learning-based model to predict SUR curve by using the Support Vector Regression (SVR) under the assumption that the individual JND points of different subjects follow a normal distribution. This SUR predictor is trained on VideoSet [24] using Quality Degradation Features using VMAF [8] and Masking Features [4, 5]. The 75%SUR (75%JND) can be derived from the predicted SUR curve. Similar method is used in [25] to predict the *2nd* and *3rd* JND points using 3 different settings in which the reference inputs of the predictor are different. Instead of predicting encoding parameter QP of H.264 as JND profile, Zhang *et al.* [29] proposed a novel perceptual model to predict SUR versus bitrate, which is more widely used in practice. Three kinds of features, Masking features, re-compression features, and basic attribute features, are extracted from SRC and several PVSs. Gaussian Processes Regression (GPR) is applied to predict SUR. Zhang *et al.* [30] proposed a deep-learning based SUR prediction model that extracts spatial and temporal features via CNNs and a weighted average pooling to predict SUR value of a given PVS.

However, these picture-wise and video-wise JND prediction models are not application-free, they are limited to a certain image/video codec. For instance, the video-wise JND prediction models proposed in [24, 30] are learning-base models, they are both trained on VideoSet [24] which is compressed by H.264, and the JND proxy is the quantization parameter (QP) value. However with the rapid development of video compression techniques, it is neither practical nor efficient to collect a large-scale datasets each time for a new codec and to train a new model which is compatible with the new codec. Moreover, these JND prediction models usually need a large number of PVSs, *e.g.*, compressed videos with QP from 1 to 51 in [23, 30], to predict the SUR curve. However, for a video streaming service provider, it is very resource-consuming to first compress the SRC with every possible encoding parameters and then predict the JND from them. Furthermore, the leaning-based JND prediction models are not only sensitive to codecs, but also to the video resolutions [30], *i.e.*, cross resolution decreases the prediction accuracy.

3 RESEARCH QUESTIONS

To address the previous issues in related work, we plan to investigate the following research questions (RQs):

RQ1: **How to design JND subjective test methodology to collect JND dataset more efficiently?**
The current large-scale JND datasets VideoSet [24] include 220 5-second SRCs in 4 encoding resolutions. Each SRC is encoded with H.264 codec with QP from 1 to 51. More than 30 subjects participated into the viewer group for each SRC. However, it is well know that JND search is extremely time consuming, we try to find a new subjective test methodology to collect JND more efficiently. This research question could be divided into several sub-questions : (i) how to select contents to cover a wide-range of characteristics? (ii) how to define the candidate PVS list instead of all possible PVS from

QP = 1 to 51 in VideoSet? (iii) how to design cross resolution JND search?

RQ2: How to predict SUR/JND for a given video codec (_e.g._, H.264) only based on SRC?
This research question aims to address the issue that in previous JND/SUR prediction models, a large number of PVSs are required as input, which is time and storage consuming for the video streaming service provider. We will try to explore a method that takes only SRC as input without re-compression for SUR curve prediction with a given codec. This question is based on the assumption that the SUR curve is mainly determined by the content features (_e.g._, masking effect of spatial and temporal randomness [23]). It should be mentioned that the distortion caused by the codec can also affect the SUR curve. To see how the quality degradation feature from PVSs can improve the accuracy, we can investigate the trade-off problem between the prediction accuracy of the SUR and the number of input PVSs.

RQ3: Can we interpret JND for compressed videos with objective Video Quality Assessment (VQA) (e.g., VMAF)?
Objective of this research question is to address the issue that the current JND prediction models are not application-free, they are usually limited to one certain codec. However, the well developed objective VQA such as VMAF [8] is codec agnostic to some degree. Based on this, we want to know if we can interpret JND for compressed videos with on-hand objective VQA, _i.e._, for two same content videos with different encoding recipes, how large is the different of their VMAF values when human eye can perceive difference between them?

RQ4: How to predict SUR/JND for a certain codec but with different encoding resolutions?
Current SUR/JND prediction models are limited to one single resolutions. For instance, the SRC are encoded in 4 different resolutions respectively in VideoSet [24], and 1st, 2nd and 3rd JND are researched in each resolution separately. However in practice, it is interesting to know the JND point drop in which encoding resolution. This is based on one observation in practice: encoding one SRC with two different resolutions (_e.g._, 1080p and 720p) and with different distortion levels (_e.g._, QP from 1 to 51) in each resolution, if we conduct two JND search in 1080p and 720p with the SRC as anchor, two JND can be obtained in 1080p and 720p respectively. Usually the QP value of the JND in 720p is smaller than the QP of JND in 1080p. We can considered that the two JND in different encoding resolutions have the same quality for observers. Determining the optimal encoding resolution is a key question to investigate to reduce the bitrate.

RQ5: How to design a codec agnostic SUR/JND prediction model or a general model which could be adapted easily for new codecs?
This research question also addresses the issue that the current learning-based JND/SUR prediction models are not codec-agnostic. It is well known that subjective JND test is much more time-consuming than traditional subjective tests such as ACR or DCR. It is not practical to collect at each time a large-scale JND datasets such as VideoSet when it comes
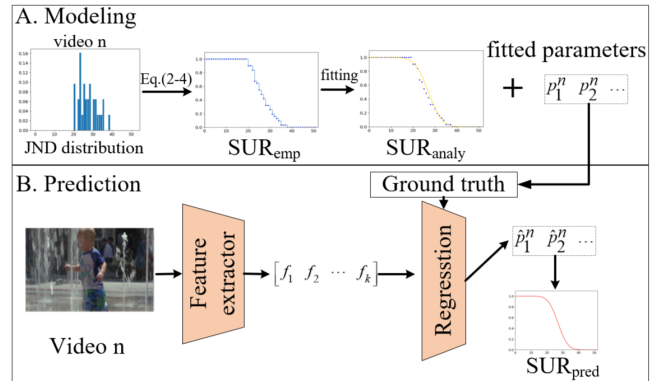

Figure 1: Illustration of the pipeline of SUR and JND modeling (A) and prediction (B)

to a new codec and to train a new prediction model based on it. How can we use the existing JND datasets and with few JND annotations of the new codec to make predictions? This transfer could be addressed with few-shot learning and meta-learning methods.

4 WORK IN PROGRESS

In this section, a brief description of the work already done and a tentative plan for the ongoing and future work for the five main research questions will be presented. We have already investigated in RQ1 and RQ2 since the beginning of the project and for the RQ3 and RQ4, general ideas and plans will be presented. This thesis started in May 2021 and is scheduled to end in mid-2024.

RQ1 We plan to collect a new large-scale JND datasets for compressed video under H.265 for HD, UHD and HDR contents. The proxy of the JND is set to be Constant Rate Factor (CRF). We use the content selection features in [12] and add compression-base features: including the number of I, P, and B frames with respective average sizes, and the clip bitrate as features to select contents with wide-range of characteristic. Each SRC is encoded with 13 PVSs in each encoding resolution, 1080p and 720p respectively. JND searches are not limited in one resolution, SRCs are used as anchor to make JND search in both encoding resolutions in order to find the best encoding parameters. 20 observers who have either normal or corrected-to-normal visual acuity were included in the experiment, and A 55-inch calibrated "UHD Grundig Finearts 55 FLX 9492 SL" was employed as display screen. The viewing distance is set to 1.5H for UHD videos and 3H for HD as recommended in ITU-R BT.2022 [1], where H is the height of the screened video. Binary search [24] is used to find JND for each content. This new JND datasets gives us opportunities to address the cross-resolution issues in RQ3 and develop codec agnostic prediction model in RQ5 along with VideoSet [24] under H.264 compression.

RQ2 We proposed a pipeline for the prediction of SUR and JND based only on SRC as illustrated in Fig.1. There are two main steps: (i) mathematical modeling of the SUR curve to obtain the parameters as ground truth that describe the SUR curve; (ii) prediction of the parameters and SUR accordingly from the features of SRC. We
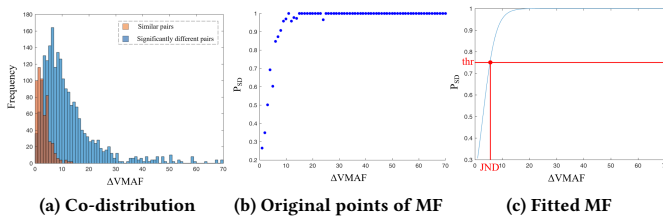
(a) Co-distribution **(b) Original points of MF** **(c) Fitted MF**

Figure 2: How to obtain the mapping function (MF)

compared different candidate functions with the monotonic non-increasing constraint to find the best-fit function. We extracted features that characterize the "complexity" spatial and temporal, which will mask the capacity of human eyes to detect distortion [23]. Features are not extracted directly from the entire video, but the spatial-temporal (S-T) video patches in the eye fixation level [18]. The distribution of features from each S-T patches of a entire video will be fed into a Support Vector Regression (SVR) to predict the SUR curve parameters. JND can be calculated with a given threshold (*e.g.*, 75%) for SUR. Moreover, we also compared the prediction accuracy between the model that takes only SRC features as input and the model that takes also PVS features as input. Experiment results show that adding quality degradation information from PVSs will slightly improve the prediction of both SUR and JND. The MAE of the our SUR prediction model is 0.0446 ± 0.0021 in VideoSet.

RQ3 To address the RQ2, we proposed mapping functions of the VMAF [8] residual and the probability of JND for a given anchor. In other words, with a given video as anchor (anchor could be SRC or PVS), and the VMAF of this anchor video $VMAF(A)$, we can predict the residual of the VMAF ($\Delta VMAF$) between $VMAF(A)$ and $VMAF(J)$ of JND video. The mapping functions are obtained from the subjective annotations of the DCR. The process is demonstrated in Fig. 2. All videos (SRCs and their PVSs with different encoding recipes) are grouped into pairs of two (videos in a pair are always the same content), and each video pair will be classified into two categories: (i) significantly different pairs (SDP); (ii) similar pairs (SP). This classification is based on the significant t-test with the DCR annotations. Afterwards, we can get the two histograms of the SDP and SP, respectively, with the same number and interval of bins of Δ(VMAF), namely, co-distribution (Fig. 2a). For each Δ(VMAF) bin, the probability of being SDP could be calculated by $f_{SDP}/(f_{SDP} + f_{SP})$, where f is the frequency in histograms. These probabilities along with their Δ(VMAF) are the original points of the mapping function as shown in Fig.2b. The final mapping function is obtained by fitting these original points with Generalized Linear Model (GLM). As illustrated in Fig.2c, the JND is the corresponding value Δ(VMAF) in the mapping function with a given threshold (*thr*). The mapping functions varies with the objective quality range of the anchor video. Experiment result shows that the MAE of Δ(VMAF) of JND is 2.9808 ± 0.3120 in VideoSet [24] (VMAF$\in$ [1, 100]).

RQ4 As shown in Fig.3, SRC is encoded with different encoding resolutions with a given codec (blue and green curve for 1080p and orange curve for 720p). The blue point is the 1st JND with SRC as anchor. If we make the 2nd JND search in the 1080p (green

curve) and in the 720p (orange curve), respectively, with the 1st JND (blue point) as anchor, there will be one JND in 720p (orange point) and one JND in 1080p (green point). The question is which resolution is optimal for the 2nd JND. The purpose of JND prediction is to reduce the resources without affecting the user experience, hence, the optimal encoding resolution should be the one that has a lower bitrate. For example, the optimal encoding resolution for Fig.3a is 720p while for Fig.3b is 1080p. However, the current works conduct JND research separately in each encoding resolution. To our acknowledge, there is no JND datasets taking cross resolution into consideration. We plan to first make a JND datasets in which the JND searches are not limited to only one encoding resolutions, and afterward develop methods to predict cross-resolution JND.
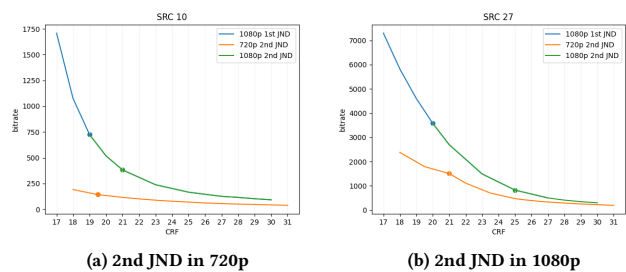

(a) 2nd JND in 720p **(b) 2nd JND in 1080p**

Figure 3: Cross resolution JND analyse

RQ5 The current learning-based JND prediction models are specific to a certain codec such as H.264. When updating the codec, it is costly to collect a new large-scale datasets and train a new model based on it. We plan to use meta-learning and few-shot learning methods to find solutions. Few-shot learning is to handle new data with a few training examples. Similarly, meta-learning aims at learning how to learn. It seeks to learn a general model with available problems/data, which could be then adapted quickly to solve other unrelated/novel problems with limited samples.

5 CONCLUSION

In this thesis, we focus on five mains research questions about the JND prediction for compressed video to guide the choice of encoding parameters for a high quality of experience with limited storage and internet bandwidth. Each research questions are based on the existing problems in related work, and we try to address these questions on the benefit of on-hand VQA methods and machine learning methods. We present briefly the methodologies and conclusions of work already done and moreover introduce the guidelines for the ongoing and future work.

REFERENCES

[1] ITUR BT. 2012. General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. *International Telecommunication Union* (2012).

[2] Chunling Fan, Hanhe Lin, Vlad Hosu, Yun Zhang, Qingshan Jiang, Raouf Hamza-oui, and Dietmar Saupe. 2019. 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, 1–6.

[3] Chunling Fan, Yun Zhang, Huan Zhang, Raouf Hamzaoui, and Qingshan Jiang. 2019. 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* 62 (2019), 140–151.

[4] Sudeng Hu, Lina Jin, Hanli Wang, Yun Zhang, Sam Kwong, and C-C Jay Kuo. 2015. Compressed image quality metric based on perceptually weighted distortion. *IEEE Transactions on Image Processing* 24, 12 (2015), 5594–5608.

[5] Sudeng Hu, Lina Jin, Hanli Wang, Yun Zhang, Sam Kwong, and C-C Jay Kuo. 2016. Objective video quality assessment based on perceptually weighted mean squared error. *IEEE Transactions on Circuits and Systems for Video Technology* 27, 9 (2016), 1844–1855.

[6] P ITU-T RECOMMENDATION. 1999. Subjective video quality assessment methods for multimedia applications. (1999).

[7] Lina Jin, Joe Yuchieh Lin, Sudeng Hu, Haiqiang Wang, Ping Wang, Ioannis Katsavounidis, Anne Aaron, and C-C Jay Kuo. 2016. Statistical study on perceived JPEG image quality via MCL-JCI dataset construction and analysis. *Electronic Imaging* 2016, 13 (2016), 1–9.

[8] Zhi Li, Anne Aaron, Ioannis Katsavounidis, Anush Moorthy, and Megha Manohara. 2016. Toward a practical perceptual video quality metric. *The Netflix Tech Blog* 6, 2 (2016).

[9] Hanhe Lin, Guangan Chen, Mohsen Jenadeleh, Vlad Hosu, Ulf-Dietrich Reips, Raouf Hamzaoui, and Dietmar Saupe. 2022. Large-scale crowdsourced subjective assessment of picturewise just noticeable difference. *IEEE Transactions on Circuits and Systems for Video Technology* (2022).

[10] Hanhe Lin, Vlad Hosu, Chunling Fan, Yun Zhang, Yuchen Mu, Raouf Hamzaoui, and Dietmar Saupe. 2020. SUR-FeatNet: Predicting the satisfied user ratio curve for image compression with deep feature learning. *Quality and User Experience* 5, 1 (2020), 1–23.

[11] Joe Yuchieh Lin, Lina Jin, Sudeng Hu, Ioannis Katsavounidis, Zhi Li, Anne Aaron, and C-C Jay Kuo. 2015. Experimental design and analysis of JND test on coded image/video. In *Applications of digital image processing XXXVIII*, Vol. 9599. International Society for Optics and Photonics, 95990Z.

[12] Suiyi Ling, Yoann Baveye, Deepthi Nandakumar, Sriram Sethuraman, and Patrick Le Callet. 2020. Towards better quality assessment of high-quality videos. In *Proceedings of the 1st Workshop on Quality of Experience (QoE) in Visual Multimedia Applications*. 3–9.

[13] Huanhua Liu, Yun Zhang, Huan Zhang, Chunling Fan, Sam Kwong, C-C Jay Kuo, and Xiaoping Fan. 2019. Deep learning-based picture-wise just noticeable distortion prediction model for image compression. *IEEE Transactions on Image Processing* 29 (2019), 641–656.

[14] Xiaohua Liu, Zihao Chen, Xu Wang, Jianmin Jiang, and Sam Kowng. 2018. JND-Pano: Database for just noticeable difference of JPEG compressed panoramic images. In *Pacific Rim Conference on Multimedia*. Springer, 458–468.

[15] Lin Ma, King Ngi Ngan, Fan Zhang, and Songnan Li. 2011. Adaptive block-size transform based just-noticeable difference model for images/videos. *Signal Processing: Image Communication* 26, 3 (2011), 162–174.

[16] Rafał Mantiuk, Kil Joong Kim, Allan G Rempel, and Wolfgang Heidrich. 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on graphics (TOG)* 30, 4 (2011), 1–14.

[17] E Nakasu, K Aoki, R Yajima, Y Kanatsugu, and K Kubota. 1996. A statistical analysis of MPEG-2 picture quality for television broadcasting. *SMPTE journal* 105, 11 (1996), 702–711.

[18] Alexandre Ninassi, Olivier Le Meur, Patrick Le Callet, and Dominique Barba. 2009. Considering temporal variations of spatial visual distortions in video quality assessment. *IEEE Journal of Selected Topics in Signal Processing* 3, 2 (2009), 253–265.

[19] Xuelin Shen, Zhangkai Ni, Wenhan Yang, Xinfeng Zhang, Shiqi Wang, and Sam Kwong. 2020. Just Noticeable Distortion Profile Inference: A Patch-Level Structural Visibility Learning Approach. *IEEE Transactions on Image Processing* 30 (2020), 26–38.

[20] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2818–2826.

[21] Wenfei Wan, Jinjian Wu, Xuemei Xie, and Guangming Shi. 2017. A novel just noticeable difference model via orientation regularity in DCT domain. *Ieee Access* 5 (2017), 22953–22964.

[22] Haiqiang Wang, Weihao Gan, Sudeng Hu, Joe Yuchieh Lin, Lina Jin, Longguang Song, Ping Wang, Ioannis Katsavounidis, Anne Aaron, and C-C Jay Kuo. 2016. MCL-JCV: a JND-based H. 264/AVC video quality assessment dataset. In *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1509–1513.

[23] Haiqiang Wang, Ioannis Katsavounidis, Qin Huang, Xin Zhou, and C-C Jay Kuo. 2018. Prediction of satisfied user ratio for compressed video. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6747–6751.

[24] Haiqiang Wang, Ioannis Katsavounidis, Jiantong Zhou, Jeonghoon Park, Shawmin Lei, Xin Zhou, Man-On Pun, Xin Jin, Ronggang Wang, Xu Wang, et al. 2017. VideoSet: A large-scale compressed video quality dataset based on JND measurement. *Journal of Visual Communication and Image Representation* 46 (2017), 292–302.

[25] Haiqiang Wang, Xinfeng Zhang, Chao Yang, and C-C Jay Kuo. 2018. Analysis and prediction of JND-based video quality model. In *2018 Picture Coding Symposium (PCS)*. IEEE, 278–282.

[26] Zhenyu Wei and King N. Ngan. 2009. Spatio-Temporal Just Noticeable Distortion Profile for Grey Scale Image/Video in DCT Domain. *IEEE Transactions on Circuits and Systems for Video Technology* 19, 3 (2009), 337–346. https://doi.org/10.1109/TCSVT.2009.2013518

[27] Jinjian Wu, Leida Li, Weisheng Dong, Guangming Shi, Weisi Lin, and C-C Jay Kuo. 2017. Enhanced just noticeable difference model for images with pattern complexity. *IEEE Transactions on Image Processing* 26, 6 (2017), 2682–2693.

[28] Jinjian Wu, Guangming Shi, Weisi Lin, Anmin Liu, and Fei Qi. 2013. Just noticeable difference estimation for images with free-energy principle. *IEEE Transactions on Multimedia* 15, 7 (2013), 1705–1710.

[29] Xinfeng Zhang, Chao Yang, Haiqiang Wang, Wei Xu, and C-C Jay Kuo. 2020. Satisfied-user-ratio modeling for compressed video. *IEEE Transactions on Image Processing* 29 (2020), 3777–3789.

[30] Yun Zhang, Huanhua Liu, You Yang, Xiaoping Fan, Sam Kwong, and CC Jay Kuo. 2021. Deep learning based just noticeable difference and perceptual quality prediction models for compressed video. *IEEE Transactions on Circuits and Systems for Video Technology* (2021).