Bridge the Gap between Visual Difference Prediction Model and Just Noticeable Difference Subjective Datasets

Jingwen Zhu, Patrick Le Callet, Jiawen Liu

▶ **To cite this version:**

Jingwen Zhu, Patrick Le Callet, Jiawen Liu. Bridge the Gap between Visual Difference Prediction Model and Just Noticeable Difference Subjective Datasets. IEEE International Workshop on Multimedia Signal Processing, IEEE, Sep 2023, Poiters, France. hal-04197899

HAL Id: hal-04197899
https://hal.science/hal-04197899

Submitted on 6 Sep 2023

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.

Bridge the Gap between Visual Difference Prediction Model and Just Noticeable Difference Subjective Datasets

Jiawen Liu[1,2], Jingwen Zhu[2], Patrick Le Callet[2]
[1]College of Information Science and Engineering, Ocean University of China, Qingdao, China
[2]Nantes Université, Ecole Centrale Nantes, CNRS, LS2N, UMR 6004, IUF, Nantes, France

***Abstract*—In video compression applications, the term 75%SUR (Satisfied User Ratio) is used to describe the compression parameter with which only 75% of the users can not notice the difference between one compressed media and its source. 75%SUR is widely used in JND (Just Noticeable Difference) modeling as a common threshold to standardize differences in users' perceptions of JND location. Visible difference detection is an essential step in JND prediction. However, Visible Difference Predictors (VDP), as objective quality metrics, are usually calibrated and applied on media quality datasets, no study has yet trained or applied the VDP on JND datasets. In this work, we will explore the feasibility of using the VDP model in predicting SUR and JND. We focus on Video Wise JND(VW-JND) and propose the model Extend-FvVDP, which maps the continuous quality scores output from the current best-performing VDP model, the FovVideoVDP, to VW-JND ground truth. Finally, Extend-FvVDP got a mean SUR prediction error of 0.0624, a mean JND prediction error of 1.9318. Our results show that VDP still performs on the JND datasets, and the JND prediction using VDP has the potential to exceed that of pure deep learning models.**

***Index Terms*—just noticeable difference, quality metric, visible difference predictor, visual perception**

I. Introduction

Video quality decreases with increasing compression. However, in cases where compression has to be applied, it is imperative to ensure the highest possible perceived quality for the user. We can do this by relying on a psychophysics study conclusion, which states that the human visual system has a limited sensitivity in recognizing minor video distortions [11]–[13], [15]. As we increase the compression on the source video, the level of distortion also escalates. When the subject first observes the difference between the compressed video and its source, the compression parameter at this point is called the first JND. The 1st JND is the optimal compression parameter that guarantees the perceived quality. While even for the same video, the JND location varies from person to person. SUR proves to be an effective method for aggregating the individual VW-JNDs of each observer [6]–[8].

We can obtain JND data through subjective test experiments, in which the subjects point out the location of JND by their own observation on the source and compressed videos. We can derive the SUR curve from the JND data. The SUR curve depicts the ratio of users who are satisfied with the compressed video (users who cannot see the video difference) as the compression level changes. Fig.1 shows an example SUR curve. The ordinate represents the satisfied user ratio, and the abscissa represents the video compression Quantization Parameter(QP). It can be observed that as the compression level increases, more and more users perceive the distortion. Previous works [6], [7], [15] take the QP corresponding to the threshold SUR of 75% in order to satisfy a great majority of viewers. This QP is the result of harmonizing the JNDs of different users, it helps us make decisions on the trade-off between the compression level and the video perception quality. In this paper, we use the QP corresponding to 75% SUR as the prediction target and call this QP value the harmonized JND of the video. The streaming providers can certainly opt for distinct thresholds tailored to various use cases.

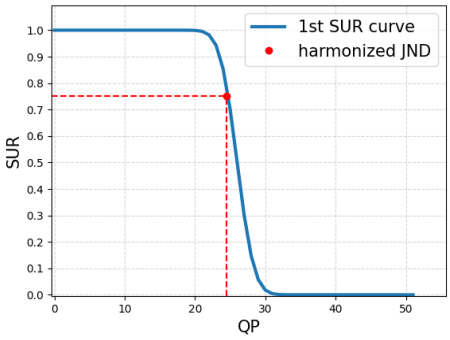

Fig. 1. Example of a SUR curve. The intersection point of the 75% threshold and the SUR curve is the harmonized JND point [9]

II. Related works

Obtaining JND data from subjective test experiments, while ensuring data accuracy, is expensive and inefficient. Therefore, the prediction of JND becomes a crucial task. Existing video JND prediction models can be divided into two categories, pixel-wise models and video-wise models. Many studies [2]–[5] work on pixel-wise models. These models focus on the effect of visual perceptual factors such as masking effects or subband frequencies on media. Pixel-wise JND modeling is more challenging since it is difficult to model all the influencing factors and there are still many unknown factors. In

addition, since it is difficult for the human eyes to recognize the small differences in pixels, it is more the experience of the entire picture or video. Therefore, most of the currently available JND prediction models are image/video wise models [6]–[9]. In [6], a perceptual model for compressed video SUR prediction was proposed by exploiting the bitrate changes in the compressed video, and the spatial-temporal features extracted from both compressed video as well as its source. In [7], Wang et al. proposed a method to predict the SUR curves using VMAF (Video Multi-method Assessment Fusion) [10] quality degradation features followed masking effect features. After that, a SVM model is used to regress the features to complete the prediction of JND. Zhu et al. proposed a source based VW-JND prediction model in [8], which improves the model in [7] by extracting the masking effect features from only the source video and then using SVM to regress the features to complete the JND prediction. Zhang et al. proposed a deep learning model for H.264 videos' JND prediction in [9]. The authors used a deep learning model for extracting spatio-temporal features of the reference and test videos and used them to predict the SUR curve. Then the QP corresponding to 75%SUR is taken as JND. This model outperforms in both SUR and JND prediction as compared with the state-of-the-art.

Besides, the JND predictors are designed to find the QP at which visible differences are just perceived by users, which brings us to the VDP. Mantiuk et al. proposed a calibrated image visual metric called HDR-VDP2 in [11]. This metric is based on a new visual model for all luminance conditions, especially the High Dynamic Range (HDR) images. It was calibrated against several contrast discrimination datasets, and image quality databases [16], [17]. In [12], the authors further proposed HDR-VDP2.2 to improve HDR-VDP2 to support the quality analysis for both Standard Dynamic Range (SDR) signals and HDR signals. Mantiuk et al. also proposed a quality metric specific for videos called FovVideoVDP in [13]. For computational complexity reasons, FovVideoVDP removes the modeling of the Human Visual System. FovVideoVDP adds temporal characteristics and contrast masking analysis to better detect video visibility. FovVideoVDP has been calibrated on 3 independent foveated video datasets, and on a large image quality dataset to evaluate its validity.

Visibility detection is an indispensable step in JND prediction, and usually implicitly included in the neural network structure of JND predictors. While the standalone VDPs are calibrated and applied on media quality datasets, no study has yet trained or applied the VDP on JND datasets. In this work, we will bridge the gap between VDP and JND datasets. Our main contributions are:

- Validated VDP on the JND datasets;
- Proposed the VW-JND prediction model **Extend-FvVDP** to explore the mapping relation between VDP score and JND location;
- Analysed the room for improvement of VDP on JND prediction.

The rest of this paper is organized as follows: In section III, we will describe our experiment data and model's implementation details. In section IV, we will examine the prediction accuracy and robustness of Extend-FvVDP. The final section V will summarize the whole paper and point out the room for improvement of this study.

III. Proposed framework for VW-JND prediction

In this section we will describe the implementation details of our proposed model Extend-FvVDP. Since the environmental factors have a significant impact on the output of VDP and our study conclusion, we will pay attention to the values and acquisition methods of the environmental parameters used in this paper and describe them in section III.A. In section III.B, we will describe Extend-FvVDP's framework structure.

A. Dataset and its environmental parameters

This paper uses 220 source videos with the resolution of 1920x1080 and their compressed versions in the VW-JND dataset VideoSet [15]. Each source video has 51 compressed versions with H.264, corresponding to the QP range from 1 to 51. When QP is at the edge, the video cannot be substantially compressed. QP 1-7 product the same video, so do the QP 47-51. Therefore, this paper only uses compressed videos with QP of 7-47, a total of 9240 videos, as our experiment data. In this paper, QP appears many times in the charts as the abscissa. Without loss of generality, all QP values in the chart in this paper are the result of subtracting 7 from the real value. For example, when QP shows 0 in the chart, it represents QP 7.

The environmental parameters of the subjective test experiment have a serious impact on the perceived quality and the output of the VDP. Therefore, we have made careful considerations in the selection of parameters. The parameters required in this paper are specifically:

- Y_peak: peak luminance of the display in cd/m^2
- contrast: display contrast
- gamma: standard gamma-encoding
- E_ambient: ambient light in lux
- k_ref: reflectivity of the display
- ppd: pixels per degree

VideoSet is a huge video dataset, thus the subjective test on it was assigned to different universities in different cities. Every experiment site conducted the subjective test experiment in different environment. VideoSet's authors didn't record the assignment of the experiment work. Thus, we can only use the mean of all the experiment site as our parameter. Besides, the environmental parameters we need are only very vaguely represented in the dataset paper [15]. We begin with the **Y_peak**. The paper does not precise the parameter values but only a chart showing them. So we added gridlines to the chart. Fig.2 shows the original chart with our gridlines. Each blue cross represents the Y-peak value of one experiment site. We try to record the value of each cross as accurately as possible, and then calculate their mean. The parameter **contrast** is in the same situation, we did the same process. For the parameter

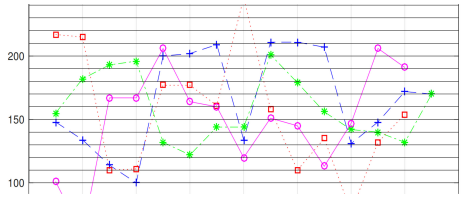

Fig. 2. VideoSet chart of peak luminance with gridlines [15]

TABLE I
VDP ENVIRONMENT PARAMETERS

Parameter	Value
Y_peak	165.8
contrast	435
gamma	2.2
E_ambient	100
k_ref	0.005
ppd	60.8

E_ambient, it is mentioned in the paper that the ambient light of the subjective test experiment is the brightness of usual offices, so we set it to 100 lux. The other two parameters **gamma** and **k_refl** are not mentioned in the paper at all. As these two parameters have only a weak effect on VDP output, we leave them as the default value of FovVideoVDP. Finally, the paper clearly informs that distance_h(viewing distance/active display height) equal to 3.2, and the display height resolution is 1080. We use the equation (1) to calculate **ppd**. The final environment parameter settings are shown in Table I.

$$ppd = \frac{height\_resolution}{arctan(\frac{1}{distance\_h})} \tag{1}$$

B. Extend-FvVDP, JND prediction using FovVideoVDP

We propose Extend-FvVDP, which extends FovVideoVDP for JND prediction to demonstrate the existence of the mapping relation between VDP quality score and JND. Fig.3 shows the framework structure of Extend-FvVDP. This section goes on to explain the implementation details of each step.

Calculation of VDP scores

At the end of the FovVideoVDP model, the authors pool the visible differences of each frame of the video to obtain a pooling score. This pooling score is then regressed to map to the final quality score. The purpose of this step is to scale the pooling score to make the final output more closely match the quality score ground truth. Since there is no quality score data to be fit and only the visible differences in quantification are needed in this study, we dropped the final regression and directly use the pooling scores as the output score of FovVideoVDP.

We first input all the 9020 compressed videos as well the 220 source videos into FovVideoVDP. After display model, temporal & multiscale decomposition, contrast sensitivity & masking modules and the final pooling, we get the output scores. For every source video, we learned the mapping of each QP to VDP score. From the subjective test experiments, for every source video, we learned the mapping of each QP to SUR value. In the next step, we will find the correspondence between VDP score and SUR value using QP as the proxy.

Preprocessing of VDP score

For each source video, its VDP score increases with increasing QP in the form of a concave curve. VDP score achieves a minimum value of 0 at QP equal to 0, where the source video is compared with itself, so the visible difference is 0. But the maximum value has a considerable range of variation from 8000 to 17500. We calculated the correlation between the maximum VDP score and JND ground truth of each source video. We got Pearson Correlation Coefficient (PCC) = 0.0076 and Spearman Correlation Coefficient (SRCC) = 0.0363. They are both approximately 0. This shows that the range of scores has no effect on the JND prediction. To fit these curves better, we normalized them using the maximum score of each source video.

Besides, we found some points out of the trend of the VDP score curve. The objective quality of the video will gradually decrease with the increase of the QP. These outliers are unnormal and will interfere with the following regression. So we replace the outliers with the mean of the before and the after points. Among the 220 1080P source videos in VideoSet, we find 8 videos with outlier scores, they are SRC10, SRC40, SRC41, SRC42, SRC43, SRC45, SRC53 and SRC124.

Feature Extraction and Regression Model

Using QP as a proxy, we can obtain the correspondence between VDP score and SUR ground truth. Fig.4 gives the corresponding relation between the VDP score and SUR of all the 9020 compressed videos in our dataset. Each blue dot represent a compressed video with its VDP score and SUR value. We can observe that when we fix the score, for example, when score = 1250 as the red vertical line shown in Fig.4, on different videos, the score corresponds to a big range of SUR values. Therefore, our regression work cannot simply regress the (VDP score, SUR) point set, but consider the VDP score curve where each point is located. Therefore, for each source video and its compressed versions, we performed polynomial fitting between the VDP score and QP value. After that, we use the fitting parameters as features, and splicing them with the QP index and score value of each point to jointly predict the SUR value. Through this method, we provide the regressor with the score of each QP, and also the information of the source video where this score point is from. Our feature structure is as equation (2):

$$[fit\_para1...fit\_paraN, QP\_index, vdp\_score] -> SUR \tag{2}$$

The left side of the arrow is the input vector, the output of the model is the SUR value corresponding to the source video and QP index.

In order to select the appropriate number of fitting parameters and the type of regressor, we tested the performance

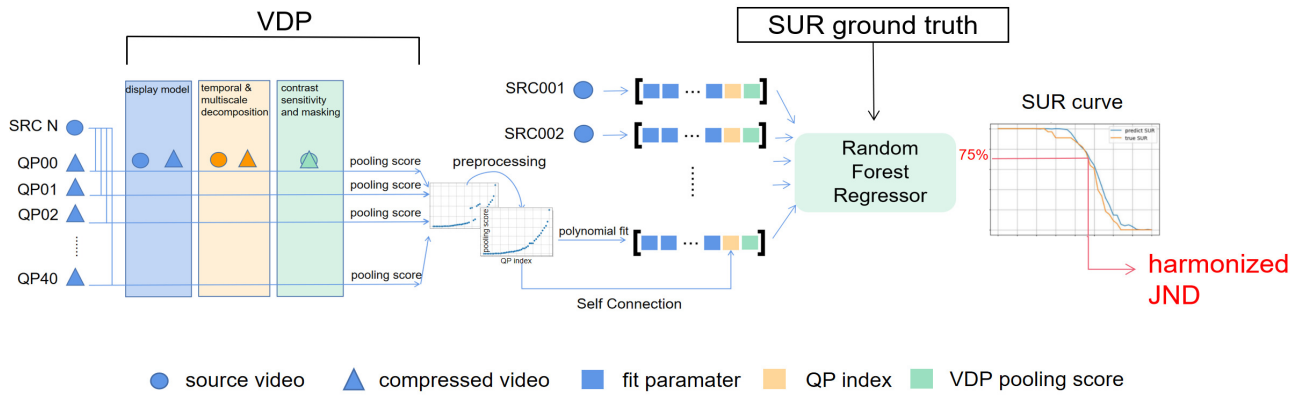

Fig. 3. Extend-VDP framework structure [13]

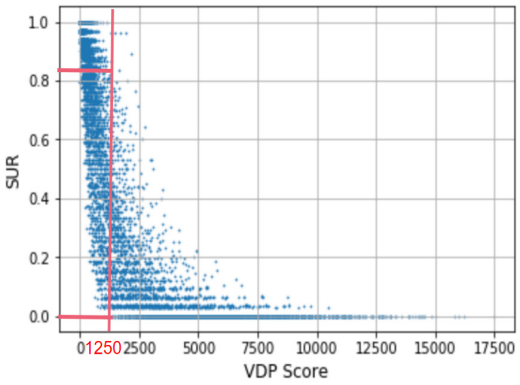

Fig. 4. VDP score and SUR corresponding map of all the 9020 compressed video in our dataset. This figure shows that the same VDP score may correspond to a very large range of SURs on different videos. Therefore, there is no uniform mapping between VDP score and SUR, but rather it is related to the video as a whole.

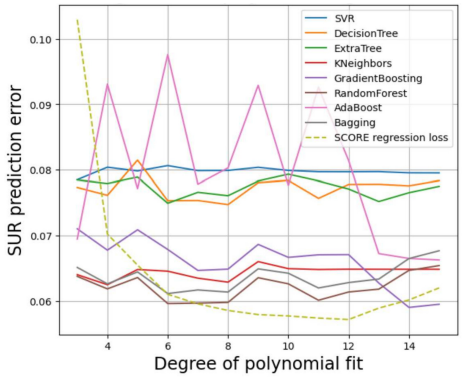

Fig. 5. Polynomial fit degree and regressor type test results

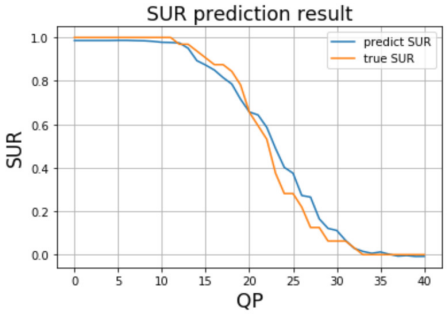

Fig. 6. SUR prediction result example: SUR prediction result of the source video SRC220

of 8 different regressors under different numbers of curve fitting degree, the results are shown in Fig.5. The abscissa represents the degree of polynomial fitting, and the ordinate represents the mean absolute error of SUR prediction out from different regression models after 4 cross-tests. We consider firstly the score curve fit. We can see that when the degree equals to 12, the score curves are fitted the best. So we set the degree of the polynomial fit to 12. Then we selected the best performing model when degree equals to 12, that is random forest regressor. Fig.6 gives an example of the SUR curve prediction results of source video SRC220. After obtaining the SUR curve, we only need to find the QP corresponding to 75% SUR to complete the harmonized JND prediction.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

To evaluate the performance of Extend-FvVDP, we compared our prediction results with the results of the model **VW-STSUR-QF** proposed in [9]. VW-STSUR-QF is implemented entirely based on deep learning methods and the authors also use VideoSet [15] for training and evaluation. It is suitable as our comparison object. The dataset division scheme used in [9] is 60% for training, 20% for validation, and 20% for testing. For a fairer comparison, we applied the same dataset split scheme.

We first tuned our model using the train and validation set. In section III-B, we have given test results for the degree of polynomial fit and the type of regressor in Fig.5, and we finally chose the degree of 12 and the random forest regressor. Table II shows the specific training error for each fold of the random forest regressor.

We then apply this tuned model to the test set. The SUR

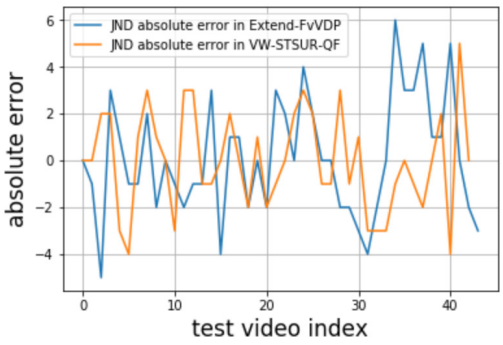

Fig. 7. The absolute prediction error of Extend-FvVDP and VW-STSUR-QF

TABLE II
EXTEND-FVVDP TRAINING ERROR FOR SUR PREDICTION

fold	MAE	MSE	R2
1	0.0598	0.0106	0.9391
2	0.0628	0.0123	0.9296
3	0.0672	0.0130	0.9250
4	0.0559	0.0089	0.9473
mean	0.0614	0.0112	0.9353

and JND prediction results are given in table III.

Compared with the results given in [9]: SUR MAE = 0.049, harmonized JND MAE = 1.69, the accuracy of our model in predicting SUR and JND was reduced by 1.3% and 0.58% respectively according to equation (3), where $\Delta acc(x)$ represents the increment in the prediction accuracy of the variable x, $\Delta MAE(x)$ represents the incremental MAE of the variable x, and $MAX(x)$ represents the maximum possible value of the variable x.

$$\Delta acc(x) = \frac{\Delta MAE(x)}{MAX(x)} \tag{3}$$

The authors of [9] also give out the specific absolute error of the test set JND, we also give ours in the Fig.7. We observed that both the JND prediction performance of our model and that of VW-STSUR-QF are unstable. For our model, the absolute error range from -5 to 6. And the absolute error of VW-STSUR-QF range from -4 to 5. In summary, our model has similar performance as VW-STSUR-QF in terms of accuracy and robustness, but slightly inferior.

V. CONCLUSION AND FURTHER WORK

In this work, we innovatively calibrated the VDP using the JND dataset and subsequently applied it to JND prediction. Our prediction results are comparable to the current state of research, which proves the feasibility of utilizing VDP for JND prediction. But unfortunately, our model did not surpass the current state of research in terms of both accuracy and robustness.

But there is a controversial point in this work. That is the environmental parameter which we input to the model. As we mentioned before, environmental parameters have a significant impact on the output of the VDP. However, the precise values of these parameters remain elusive due to inadequate recording.

We have consistently maintained confidence in the capability of VDP for JND prediction. In our future endeavors, we intend to employ more rigorous and dependable environmental parameters and experimental designs to further enhance the accuracy of our predictions.

TABLE III
EXTEND-FVVDP PREDICTION RESULTS ON TEST SET

	MAE	MSE	R2
SUR	0.062479	0.009501	0.945483
harmonized JND	1.931818	5.931818	0.292291

REFERENCES

[1] Lin, J. Y. , et al. "Experimental design and analysis of JND test on coded image/video." International Society for Optics and Photonics (2015).
[2] Wu, J. , Lin, W. , Shi, G. , Wang, X. , & Li, F. . (2013). Pattern masking estimation in image with structural uncertainty. IEEE Transactions on Image Processing.
[3] Wang, Shiqi, Gao, Wen, Ma, & Lin, et al. (2016). Just noticeable difference estimation for screen content images. IEEE Transactions on Image Processing.
[4] Wei, Z. , Member, S. , IEEE, Fellow, & IEEE. (2009). Spatio-temporal just noticeable distortion profile for grey scale image/video in dct domain. IEEE Transactions on Circuits & Systems for Video Technology, 19(3), 337-346.
[5] Bae, S. , & Kim, M. . (2014). A novel generalized dct-based jnd profile based on an elaborate cm-jnd model for variable block-sized transforms in monochrome images. IEEE Transactions on Image Processing, 23(8), 3227-3240.
[6] Zhang, X. , Yang, C. , Wang, H. , Xu, W. , & Kuo, C. C. J. . (2020). Satisfied-user-ratio modeling for compressed video. IEEE Transactions on Image Processing, PP(99), 1-1.
[7] Wang, H. , Katsavounidis, I. , Huang, Q. , Zhou, X. , & Kuo, C. C. J. . (2017). Prediction of Satisfied User Ratio for Compressed Video. 10.1109/ICASSP.2018.8461571.
[8] Zhu, J. , Callet, P. L. , Perrin, A. F. , Sethuraman, S. , & Rahul, K. . On The Benefit of Parameter-Driven Approaches for the Modeling and the Prediction of Satisfied User Ratio for Compressed Video. 2022 IEEE International Conference on Image Processing (ICIP). IEEE.
[9] Zhang, Y. , et al. "Deep Learning Based Just Noticeable Difference and Perceptual Quality Prediction Models for Compressed Video." IEEE Transactions on Circuits and Systems for Video Technology (2022).
[10] Li, Z. , Aaron, A. ,& Manohara, M. . (2016). Toward A Practical Perceptual Video Quality Metric.
[11] Mantiuk, R. , et al. "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions." ACM Transactions on Graphics 30.4(2011):40.
[12] Narwaria, M. , Mantiuk, R. K. , Da Silva, M. P. , & Le Callet, P. . (2014). Hdr-vdp-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images. Journal of Electronic Imaging, 24(1), 010501.
[13] Rafał K. Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. 2021. FovVideoVDP: a visible difference predictor for wide field-of-view video. ACM Trans. Graph. 40, 4, Article 49 (August 2021), 19 pages.
[14] https://github.com/gfxdisp/FovVideoVDP
[15] Wang, H. , et al. "VideoSet: A large-scale compressed video quality dataset based on JND measurement." Journal of Visual Communication & Image Representation 46.jul.(2017):292-302.
[16] Sheikh, H. R. , Sabir, M. F. , & Bovik, A. C. . (2006). A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Transactions on Image Processing, 15.
[17] Ponomarenko, N. , Battisti, F. , Egiazarian, K. , Astola, J. , & Lukin, V. . (2009). Metrics performance comparison for color image database. proceedings of international workshop on video processing & quality metrics.