From Just Noticeable Differences to Image Quality

Ali Ak, Andreas Pastor, Patrick Le Callet

From Just Noticeable Differences to Image Quality

Ali Ak
Andreas Pastor
Patrick Le Callet
ali.ak@univ-nantes.fr
andreas.pastor@univ-nantes.fr
patrick.lecallet@univ-nantes.fr
Nantes University, Ecole Centrale Nantes, CNRS, LS2N, UMR 6004
Nantes, France

ABSTRACT

Distortions can occur due to several processing steps in the imaging chain of a wide range of multimedia content. The visibility of distortions is highly correlated with the overall perceived quality of a certain multimedia content. Subjective quality evaluation of images relies mainly on mean opinion scores (MOS) to provide ground-truth for measuring image quality on a continuous scale. Alternatively, just noticeable difference (JND) defines the visibility of distortions as a binary measurement based on an anchor point. By using the pristine reference as the anchor, the first JND point can be determined. This first JND point provides an intrinsic quantification of the visible distortions within the multimedia content. Therefore, it is intuitively appealing to develop a quality assessment model by utilizing the JND information as the fundamental cornerstone. In this work, we use the first JND point information to train a Siamese Convolutional Neural Network to predict image quality scores on a continuous scale. To ensure generalization, we incorporated a white-box optical retinal pathway model to acquire achromatic responses. The proposed model, D-JNDQ, displays a competitive performance on cross dataset evaluation conducted on TID2013 dataset, proving the generalization of the model on unseen distortion types and supra-threshold distortion levels.

CCS CONCEPTS

• **Computing methodologies** → **Perception**; **Neural networks**; *Image compression.*

KEYWORDS

image quality assessment, siamese convolutional neural networks, just noticeable difference, visual difference predictor

ACM Reference Format:
Ali Ak, Andreas Pastor, and Patrick Le Callet. 2022. From Just Noticeable Differences to Image Quality. In *Proceedings of the 2nd Workshop on Quality of Experience in Visual Multimedia Applications (QoEVMA '22), October 14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3552469.3555712

1 INTRODUCTION

Subjective assessment of image quality commonly often relies on collecting Mean Opinion Scores (MOS) from a set of observers, as it provides a continuous image quality measurement. Just Noticeable Difference (JND) provides a binary measurement to quantify the perceptual differences between a pair of images and thus could serve as a potential workaround for quality prediction. It is defined as the smallest intensity change of a stimulus that can be noticed by the human visual system (HVS). Concisely, when the 1^{st} JND is obtained by using the pristine reference as an anchor, it also represents the minimum visible distortion intensity. In other words, it measures sub-threshold and near-threshold distortions. Without loss of generality, the following JNDs, *i.e.,* the 2^{nd}, 3^{rd}, ..., n^{th} JND, are the perceptual difference obtained by utilizing the previous JND point as the anchor. Since the anchor points are distorted images, these JND points provide information regarding to supra-threshold distortions. Only the first JND measures to which extent observers may start to notice the distortions when degrading the quality, while the following JNDs only provide preference information between different distortion levels. Since the first JND point [8] indicates directly the minimum noticeable distortion intensity, it may also reveal how our HVS perceives the distortions quantitatively. It is thus of great potential to be explored for the development of perceptual-based quality assessment metrics.

Recently, several JND datasets were published that measures the visual differences between different distortion levels [8, 11, 16, 17]. Nonetheless, the collected JND points of the same content may have high variation among observers. Therefore, additional models are often used to fuse observer responses into one single JND point [7].

Identifying the JND points of certain content from one observer requires a series of comparisons between pairs of images. Limited by the budget, existing datasets contain only a handful of SouRce Content (SRC) and distortion types, *i.e., the Hypothetical Reference Circuit (HRC)*[8, 11]. There may not be sufficient data to directly develop a learning-based model. Therefore, it is inevitable to adopt alternative approaches to overcome the problem of lack of training data. For instance, transfer learning was adopted in [4] to predict Satisfied User Ratio (SUR) using the MCL-JCI dataset [8]. Siamese Convolutional Neural Network (CNN) was frequently adopted in

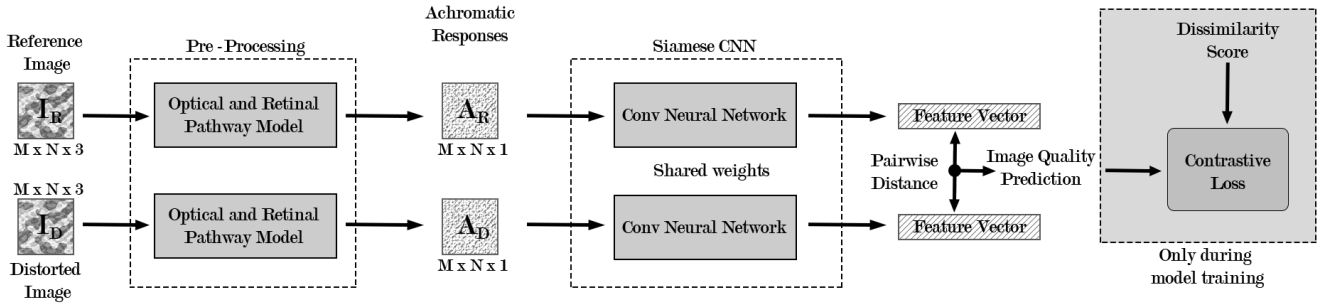

Figure 1: Diagram of the proposed model where $I_{R/D}$ indicates the input images, and $A_{R/D}$ denotes the achromatic responses. Reference and distorted image pairs inputted to the Siamese CNN after the pre-processing. Siamese CNN outputs is used for image quality prediction. During training, dissimilarity scores derived from MCL-JCI dataset is used for training via contrastive loss.

the quality domain for predicting the quality-ranking of the stimuli, where pairwise or triplet-based inputs artificially augment the limited data.

Moreover, existing white-box methods can be utilized to reduce the dimensionality of the input images while preserving related cues. Ultimately, an accurate model of the early HVS can greatly benefit not only image quality assessment but also other computer vision tasks. To this end, to model the HVS for distortion perception scenarios, Visual Difference Predictors (VDP) have been developed [3, 13]. They model the HVS in a modular fashion including optical, retinal and neural pathways based on the psychophysical studies. Although VDPs are often used as standalone models to assess the visibility of distortions, individual modules can be beneficial as feature extractors.

To this end, we propose a Siamese CNN by exploiting the perceptual information provided by the first JND step. A similarity score is determined from the first JND step opinions of observers in MCL-JCI dataset. We hypothesize that the similarity scores acquired this way are correlated with the image quality. Therefore, acquired similarity scores were used to train the a Siamese CNN to predict image quality of a distorted image in comparison to the reference. To further ensure model generalization with limited data, we also exploit the intermediate perceptual representation introduced in the Optical and Retinal Pathway model[13] to bridge the gap between the perceptual distortion space within the HVS and the latent representation output by our Siamese network. This allow us to generalize the model to unseen distortions despite training the model only on compression distortions. According to experimental results on the TID-2013 dataset [14], the proposed model achieves competitive performances compared to state-of-the-art image quality metrics. It was also verified via experiments that the model generalizes well for unseen distortion types in both sub-threshold and supra-threshold ranges. Contribution of the work can be summarized as follows:

- Incorporating white-box achromatic response model from HDR-VDP 2 [13] as pre-processing step to ensure the generalization of the model in image quality assessment of unseen distortion types and supra-threshold distortions.
- Proposing a simple similarity score derivation from the first JND points by capturing the JND point variance as the distance between reference and distorted image pair.
- Providing a publicly available image quality assessment metric[1] that displays competitive performance without any pre-training on cross dataset evaluations.

2 RELATED WORKS

There are several JND datasets in the literature for image compression and video compression with varied pre-process approaches to obtain accurate and representative JNDs. MCL-JCI dataset [8] is composed of 50 source images (SRCs) with varying numbers of JND points on JPEG compression levels. After getting the raw JND points, a Gaussian Mixture Model (GMM) was adapted to generate a staircase quality function from a set of JND points [7]. MCL-JCV dataset was released from [16], which contains JND data obtained from 50 observers prepossessed by a similar staircase quality function designed for H.264/AVC. JND-PANO dataset contains JND samples for 40 reference panoramic images over JPEG compression levels[11]. VideoSet [18] is a large-scale dataset that provides JND samples for H.264 compression levels at varying resolution. PWJNDInfer consist of JND samples over 202 reference images over compression levels.

The booming of JND subjective studies over the past several years has sparked a lot of interest and have spurred a lot of interesting ideas for the development of JND prediction models. Liu *et al.* proposed a picture-wise binary JND prediction model by defining JND prediction as a multi-label classification task and reducing it to a series of binary classification problems [10]. Fan *et al.* proposed a model to predict the satisfied user ratio and the first JND point over MCL-JCI dataset. Analogously, Zhang et al. proposed a satisfied user ratio (SUR) prediction model for video compression distortions [19]. SUR curve, for a lossy compression scheme (e.g. JPEG), aims to characterize the probability distribution of the JND levels. Consequently, SUR prediction models only aims to impact of compression algorithms on the perceived quality. They do not generalize to other type of distortions. Our work differentiates from the

[1] model weights and code available at https://github.com/kyillene/D-JNDQ

SUR prediction task by directly predicting image quality on unseen distortion types and levels. According to our best knowledge, this is the first work that utilizes the first JND points to predict overall image quality on a continuous scale for various distortion types in both sub-threshold and supra-threshold ranges.

3 PROPOSED MODEL

HVS is a very complex system and not yet fully understood. Based on relevant studies [5], it can mainly be divided into four major parts: optic, retinal, lateral geniculate nucleus, and visual cortex processing. In our proposed framework, we simplify the approach by dividing this complex process into two. We first use an existing Optical and Retinal Pathway model to pre-process input images, *i.e.,* the Optical and Retinal Pathway proposed by Mantiuk *et al.* This module provides an estimation of the achromatic responses for displayed images. Optical and Retinal processing of HVS highly affects the visibility of distortions. Hence, including this module as a pre-processing tool simplifies the similarity prediction task and allows the generalization of the model into unseen distortions. After acquiring achromatic responses of both the reference and distorted images, the remaining task is to predict the similarity between the achromatic response inputs.

Regarding its proven success in visual similarity, and pairwise ranking prediction tasks [12, 15], Siamese CNN was employed to predict the similarity between input pairs. In general, Siamese networks are equipped with two or more identical networks with shared weights to learn the embedding between a pair or triplet of input data. More concretely, we aim to learn the first JND point distributions of the observers using the Siamese network. A detailed explanation regarding to similarity score calculation is given in Section 4.

The overall structure of the proposed model is shown in Fig. 1. All the achromatic responses are acquired by pre-processing input RGB images with the optical and retinal pathway model from HDR-VDP 2. Then, these achromatic responses are fed into the Siamese CNN to extract their latent representation, *i.e., feature vectors.* Afterward, the pairwise distance between outputted feature vectors is calculated to compute a similarity score. In the following subsections, detailed information is given regarding the pre-processing stage and the Siamese CNNs.

3.1 Optical and Retinal Pathway Model

Optical and Retinal Pathways are modeled as a combination of 4 sub-modules in the HDR-VDP 2 [13]. The first module accounts for the light scattering that occurs in the cornea, lens, and retina. It is defined by a modulation transfer function estimated via psychophysical studies. The second module calculates the probability of a photo-receptor sensing a photon at a corresponding wavelength. It outputs cone and rod responses of the input image. The third module mimics the non-linear response to light of the photoreceptors. It is modeled as a non-linear transducer function. The final module converts the non-linear responses into joint cone and rod achromatic responses by simple summation.

By incorporating Optical and Retinal Pathways into the pre-processing stage, masking effects occurring at this stage of the visual pipeline could be well accounted for. By enhancing or masking the distortions visibility with existing knowledge in the domain, the training complexity of the similarity network could be much simplified and accelerated. Nevertheless, it improves the generalization of the model to tackle not only unseen distortion types, but also supra-threshold distortion values.

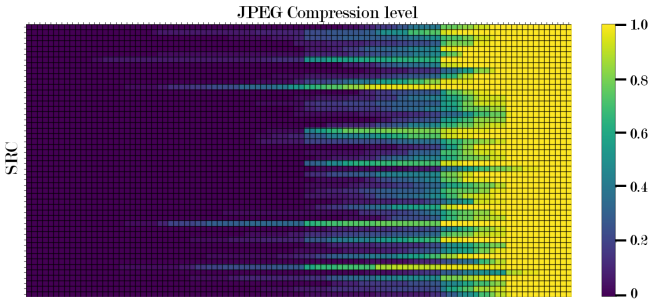

Figure 2: Dissimilarity scores acquired by using first JND steps of each observer. Each row represents an SRC. Columns are ordered from the highest QP level to lowest, left to right.

3.2 Siamese Convolutional Neural Network

The Siamese network is utilized as a feature extractor without any fully connected layers. On top of this backbone, we directly compute the pairwise distances. This architecture facilitates arbitrary input resolutions. We design our Siamese network from the scratch. It contains 5 convolutional layers with batch normalization and ReLu activation layers. To reduce the spatial resolution, a stride of 2 was adapted at the first 4 convolutional layers with appropriate zero padding.

For the last layer of the network, a Sigmoid activation function is employed without stride. After flattening the output feature vector, they are then used to calculate the similarity score between the reference and distorted images.

4 DATASET AND TRAINING DETAILS

MCL-JCI dataset [8] was used to train our network. MCL-JCI dataset contains 50 SRCs with a resolution of 1920×1080. Each SRC is encoded using JPEG encoder [1] with varying Quantization Parameter (QP) levels in a range of $[0, 100]$, where 100 corresponds to the highest quality. This results in a total of 5050 images including the reference images. 30 subjects provided 3 to 8 JND points for each SRC image with the bisection method. Individual JND points were fused for each SRC [7].

As described in Sec. 3.1, reference and distorted images from the MCL-JCI dataset are first converted into achromatic responses using the Optical and Retinal Pathway model from HDR-VDP 2. The obtained achromatic responses share the same spatial resolution as the input images. However, pixel values are represented in a single channel, resulting in an array of size $1920 \times 1080 \times 1$.

After experimenting on the MCL-JCI dataset, it was observed that the task of detecting the first JND point and following JND points are different. While identifying the first JND point, an observer tries to identify the difference between the reference and distorted image.

However, for the later JND points, this task gradually turned into a preference task, *i.e., which stimulus is preferred compared to the other*. More specifically, instead of "at which QP level the distortion becomes visible", the question evolved into "at which QP level the distortion becomes more disturbing". This observation encourages us to utilize only the first JND point for labeling the training dataset. In order to capture the uncertainty of the observers, we utilized the individual observer scores as done in [2, 4]. For each SRC, an uncompressed image was paired with 100 compressed images with different QP levels. Each pair was assigned with a dissimilarity score ranging from 0 to 1. It is defined by the number of observers, whose JND points are beyond the corresponding QP level. For a given QP level i, the dissimilarity score d_i is calculated simply as;

$$d_i = \frac{s_i}{n} \tag{1}$$

where, s_i is the number of observers with first JND point, JND_1, greater than i. n is the total number of observers.

For each SRC in the MCL-JCI dataset, dissimilarity scores for every QP level is calculated this way. In Fig. 2, each SRC is represented by a line. With a decreasing QP level, the dissimilarity between the reference and distorted image increases.

During the training, a contrastive loss function[6] is used as defined below:

$$L_\sigma = \frac{\sum_{i=1}^{N}(1 - S_i) \times D_i^2 + S_i \times (M - D_i)^2}{N} \tag{2}$$

where D_i is the euclidean distance between the two output feature vector, S_i is the ground truth similarity score between the two inputs, and M is the margin as defined in [6].

We conducted hyperparameter tuning for the Siamese network. The contrastive loss function was used with a batch size of 32 during training. We found out that a 0.03 learning rate with the Adam optimizer provides us with the best convergence speed and lowest validation loss with the final network structure. Finally, the Siamese network was trained for 100 epochs over the training dataset with the optimal hyperparameters. We also experimented with weight decay and regularization terms during hyperparameter searches. However, we observed no improvement in training convergence or model accuracy.

5 EVALUATION AND RESULTS

It is worth mentioning that our model was trained only on JPEG distortions with the first JND. To prove the generalization capabilities of the proposed model on unseen distortions and novel suprathreshold distortion levels, we conducted a cross-dataset evaluation on the TID-2013 dataset [14]. TID-2013 dataset contains 24 different distortions including but not limited to noise, blur, transmission error, and compression distortions. They are categorized into 6 overlapping groups. In total, there are 3000 distorted images with varying distortion intensity and distortion types.

We tested the model on all 3000 images without any pre-training. We used the scripts provided by the authors to calculate the correlation between the predicted results and the MOS. As such, correlation results are directly comparable with other metric correlations acquired by the authors. Table 1 reports the Spearman rank-order correlation coefficients of the proposed model and the other

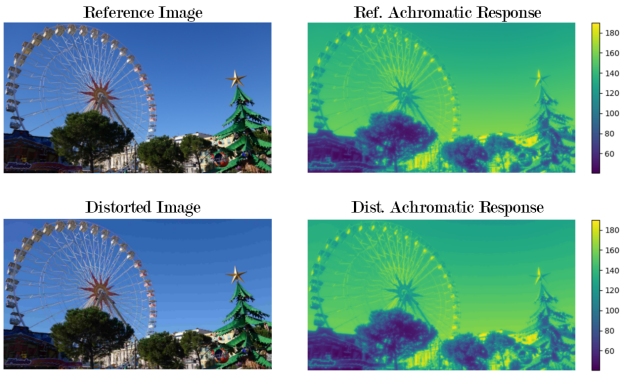

Figure 3: Reference and distorted image (QP=17) with corresponding achromatic responses for SRC-7 in MCL-JCI.

Table 1: SROCC values for selected metrics in TID-2013

	Noise	Actual	Simple	Exotic	New	Color	Full
D-JNDQ	0.851	0.881	0.894	0.315	**0.842**	**0.813**	0.589
HDR-VDP 3	0.829	0.847	0.929	0.822	0.679	0.635	0.772
FSIM	0.897	0.911	0.949	**0.844**	0.649	0.565	0.801
FSIMc	0.902	0.915	0.947	0.841	0.788	0.755	0.851
PSNR	0.822	0.825	0.913	0.597	0.618	0.535	0.640
PSNRc	0.769	0.803	0.876	0.562	0.777	0.734	0.687
PSNRHA	**0.923**	**0.938**	**0.953**	0.825	0.701	0.632	0.819
SSIM	0.757	0.788	0.837	0.632	0.579	0.505	0.637
MSSSIM	0.873	0,887	0.905	0.841	0.631	0.566	0.787
VIFP	0.784	0.815	0.897	0.557	0.589	0.506	0.608

methodologies provided by [14]. The proposed model, *i.e.,* D-JNDQ, provides competitive results with the compared metrics in Noise, Actual and Simple categories and provides better results in distortion categories "New" and "Color" compared to other evaluated metrics. The proposed model achieved the lowest performance in the subset "Exotic". This is mainly due to the preferential nature of the distortions in this category. Fig 5 presents two failure examples from this category. As it can be seen, according to MOS, image on the right has higher quality but our model predicts the opposite. This is due to model being trained on the visibility of distortions. However, for distortions, such as local block-wise distortion, detecting the distortion plays a minimal role since the distortions are visible at all levels with different variations rather than different intensities. Therefore, we expected a poor prediction performance in this category, which also reduces the overall correlation results in the 'full' category.

In order to measure the impact of the pre-processing stage of the model, we conducted an ablation study. Table. 2 depicts the result of this study. The best model parameters for each input type were trained for the same amount of iterations. Results show that the model with achromatic response input, *i.e. with pre-processing*, has a higher correlation with the MOS compared to the one using RGB inputs in all categories.

In addition to Spearman correlation evaluation, we also conducted an analysis on the performance of identifying significant

Table 2: SROCC values with and without pre-processing.

	Noise	Actual	Simple	Exotic	New	Color	Full
A.R. Input	0.851	0.881	0.894	0.315	0.842	0.813	0.589
RGB Input	0.742	0.750	0.801	0.141	0.703	0.734	0.446

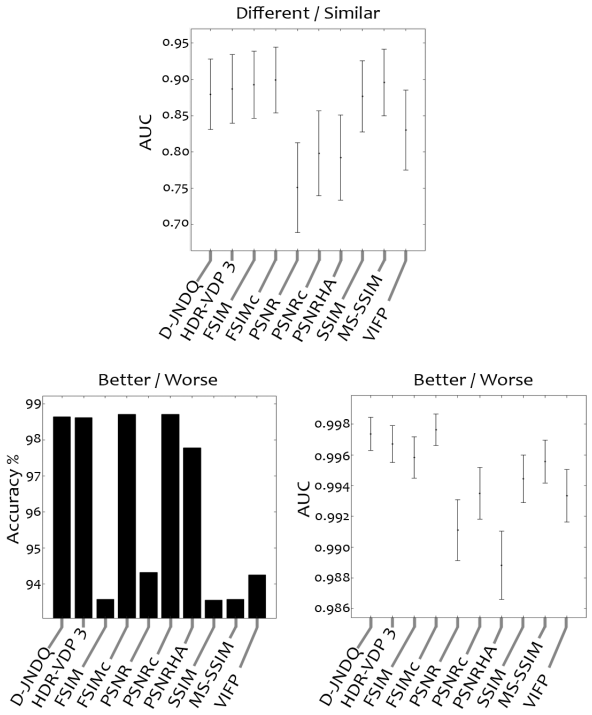

Figure 4: Metric performances on TID-2013 dataset excluding part of the "Exotic" category.

pairs. In this analysis, we have excluded the aforementioned 4 distortion types (out of the 24 total types) from the "Exotic" category. We followed the strategy proposed in [9] to stress out the performances of considered models, readers are recommended to refer to [9] for more details. In Fig. 4, the left sub-figure presents the area under curve (AUC) values for each metric at identifying significant and non-significant pairs. Similarly, the right figure shows AUC values for each metric in identifying better or worse image pairs, while the middle figure indicates the accuracy of the metric in terms of distinguishing better or worse images in significant pairs. Although there is no significant difference in many of the metric performances, the proposed metric (D-JNDQ) has a competitive performance in identifying significant versus similar pairs. For better/worse analysis, all metrics seem to perform well overall. D-JDNQ, HDR-VDP 3, FSIM, and FSIMc have a significantly better performance than the rest of the evaluated metrics in terms of AUC values. D-JNDQ, HDR-VDP 3, FSIMc, and PSNRc have more than 98% accuracy in identifying whether a stimulus within a significant pair is significantly better or worse than another.

Figure 5: Failure cases from 'Exotic' category. For both images, MOS and D-JNDQ predictions are given below. While, higher numbers indicate better quality for MOS, lower number indicates better quality for D-JNDQ.

6 CONCLUSION

We propose a learning-based metric, D-JNDQ, trained using the first JND point information. The optical and retinal pathway model from HDR-VDP 2 is used as a pre-processing module to improve the performance of the metric. Ablation study proves that the pre-processing stage is crucial for the generalization of the model. Our experimental results show that the metric can generalize well for the quality assessment of various types of distortions in both sub and supra-threshold intensities. The competitive performance of the model also demonstrated that the first JND points provide rich information for image quality assessment. More specifically, incorporating the first JND point variance among observers into similarity scores shown to be beneficial for the training of the Siamese CNN.

On another front, D-JNDQ showed poor performance for certain distortion types (e.g. Figure 5), where the image quality task is related to distortion preference rather than distortion visibility. Since we utilized a distortion visibility database (*i.e., JND dataset, MCL-JCI*) to develop the metric, this is not a surprising outcome.

At its current state, JND datasets are widely available for 2D lossy compression algorithms. With the increasing popularity of JND datasets for other multimedia types, we believe that the proposed metric and rest of our contributions can be extended to other multimedia types.

ACKNOWLEDGMENTS

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement No. 765911 (RealVision)

REFERENCES

[1] 2021. Independent JPEG Group, JPEG image compression software. http://http://www.ijg.org. Accessed: 2021-01-22.
[2] Ali Ak and Patrick Le Callet. 2019. Towards Perceptually Plausible Training of Image Restoration Neural Networks. In *International Conference on Image Processing Theory, Tools and Applications*. Istanbul, Turkey. https://doi.org/10.1109/IPTA.2019.8936096
[3] Scott Daly. 1993. *The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity.* MIT Press, Cambridge, MA, USA, 179–206.
[4] C. Fan, H. Lin, V. Hosu, Y. Zhang, Q. Jiang, R. Hamzaoui, and D. 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)*. 1–6. https://doi.org/10.1109/QoMEX.2019.8743204

[5] Xinbo Gao, Wen Lu, Dacheng Tao, and Wei Liu. 2010. Image quality assessment and human visual system. *Proceedings of SPIE - The International Society for Optical Engineering* 7744 (07 2010). https://doi.org/10.1117/12.862431

[6] R. Hadsell, S. Chopra, and Y. LeCun. 2006. Dimensionality Reduction by Learning an Invariant Mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2. 1735–1742. https://doi.org/10.1109/CVPR.2006.100

[7] S. Hu, H. Wang, and C. . J. Kuo. 2016. A GMM-based stair quality model for human perceived JPEG images. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1070–1074. https://doi.org/10.1109/ICASSP.2016.7471840

[8] Lina Jin, J. Lin, Sudeng Hu, H. Wang, P. Wang, I. 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 (2016), 1–9.

[9] L. Krasula, K. Fliegel, P. Le Callet, and M. Klíma. 2016. On the accuracy of objective image and video quality models: New methodology for performance evaluation. In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*. 1–6. https://doi.org/10.1109/QoMEX.2016.7498936

[10] H. Liu, Y. Zhang, H. Zhang, C. Fan, S. Kwong, C. . C. J. Kuo, and X. Fan. 2020. Deep Learning-Based Picture-Wise Just Noticeable Distortion Prediction Model for Image Compression. *IEEE Transactions on Image Processing* 29 (2020), 641–656. https://doi.org/10.1109/TIP.2019.2933743

[11] Xiaohua Liu, Zihao Chen, Xu Wang, Jianmin Jiang, and Sam Kowng. 2018. JND-pano: Database for just noticeable difference of JPEG compressed panoramic images. https://doi.org/10.1007/978-3-030-00776-8_42 19th Pacific-Rim Conference on Multimedia (PCM 2018) ; Conference date: 09-2018.

[12] Xialei Liu, Joost Van De Weijer, and Andrew D Bagdanov. 2017. Rankiqa: Learning from rankings for no-reference image quality assessment. In *Proceedings of the IEEE international conference on computer vision*. 1040–1049.

[13] Rafal 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 Trans. Graph.* 30, 4, Article 40 (July 2011), 14 pages. https://doi.org/10.1145/2010324.1964935

[14] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. Jay Kuo. 2015. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication* 30 (2015), 57 – 77. https://doi.org/10.1016/j.image.2014.10.009

[15] S. Roy, M. Harandi, R. Nock, and R. Hartley. 2019. Siamese Networks: The Tale of Two Manifolds. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 3046–3055. https://doi.org/10.1109/ICCV.2019.00314

[16] H. Wang, W. Gan, S. Hu, J. Y. Lin, L. Jin, L. Song, P. Wang, I. Katsavounidis, A. Aaron, and C. . J. Kuo. 2016. MCL-JCV: A JND-based H.264/AVC video quality assessment dataset. In *2016 IEEE International Conference on Image Processing (ICIP)*. 1509–1513. https://doi.org/10.1109/ICIP.2016.7532610

[17] Haiqiang Wang, Ioannis Katsavounidis, Jiantong Zhou, Jeong-Hoon 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. 2017. VideoSet: A Large-Scale Compressed Video Quality Dataset Based on JND Measurement. *CoRR* abs/1701.01500 (2017). arXiv:1701.01500 http://arxiv.org/abs/1701.01500

[18] 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. 2017. VideoSet: A Large-Scale Compressed Video Quality Dataset Based on JND Measurement. arXiv:1701.01500 [cs.MM]

[19] X. Zhang, C. Yang, H. Wang, W. Xu, and C. . C. J. Kuo. 2020. Satisfied-User-Ratio Modeling for Compressed Video. *IEEE Transactions on Image Processing* 29 (2020), 3777–3789. https://doi.org/10.1109/TIP.2020.2965994