https://doi.org/10.2352/ISSN.2470-1173.2021.11.HVEI-152

The effect of display brightness and viewing distance: a dataset for visually lossless image compression

Aliaksei Mikhailiuk, Nanyang Ye and Rafał K. Mantiuk; Univeristy of Cambridge; Cambridge, United Kingdom

Abstract

Visibility of image artifacts depends on the viewing conditions, such as display brightness and distance to the display. However, most image and video quality metrics operate under the assumption of a single standard viewing condition, without considering luminance or distance to the display. To address this limitation, we isolate brightness and distance as the components impacting the visibility of artifacts and collect a new dataset for visually lossless image compression. The dataset includes images encoded with JPEG and WebP at the quality level that makes compression artifacts imperceptible to an average observer. The visibility thresholds are collected under two luminance conditions: 10 cd/m^2, simulating a dimmed mobile phone, and 220 cd/m^2, which is a typical peak luminance of modern computer displays; and two distance conditions: 30 and 60 pixels per visual degree. The dataset was used to evaluate existing image quality and visibility metrics in their ability to consider display brightness and its distance to viewer. We include two deep neural network architectures, proposed to control image compression for visually lossless coding in our experiments.

Introduction

Finding a threshold at which the human eye cannot perceive changes introduced to an image can be beneficial for computer vision, computer graphics, and image processing algorithms. Such threshold can be used, for example, to adjust the image/video compression level so that the size of the bit-stream is minimized while the distortions remain mostly invisible. We will refer to such quantization or quality level of an image/video codec as a visually lossless threshold (VLT). Such a VLT depends not only on the image content, but will also vary with the viewing conditions: the viewing distance and display brightness. While all image quality metrics account for image content, very few of them account for viewing conditions [1, 13, 19, 35]. Such lack of accountability for viewing conditions makes VLT prediction unreliable, across different displays and different viewing distances.

The goal of this work is to provide a dataset[1] that could be used to evaluate quality and visibility metrics on the task of finding VLTs under different viewing conditions. We measure VLT at viewing distances corresponding to 30 ppd (pixels per degree) and 60 ppd and two peak brightness levels: 220 cd/m^2, common to computer displays, and 10 cd/m^2, replicating the brightness of a dimmed phone. We collect a dataset for two popular compression methods: JPEG [28] and WebP [26]. We then benchmark both hand-crafted and data-driven image metrics on the dataset.

The main contributions of our paper are: (i) a visually lossless image compression dataset with varying peak display brightness and viewing distance and (ii) performance evaluation of image quality and visibility metrics on the new dataset.

[1]The dataset can be downloaded from `https://doi.org/10.17863/CAM.60354`.

Related work

Image compression

Image and video compression methods can be divided into three types: lossless, lossy, and visually lossless. Lossless compression methods preserve all information in the decompressed image [6]. However, their compression rates are much lower than for lossy compression. Lossy compression methods allow for much higher compression rates, but they compromise on the visual image quality, often introducing noticeable artifacts [25,26]. Two commonly used lossy image compression standards are JPEG and WebP. Both methods have an adjustable quality factor (QF). The QF varies between 0 (lowest visual quality and highest compression rate) and 100 (highest visual quality and lowest compression rate). Visually lossless compression methods introduce distortions but it is ensured that they are unlikely to be noticed [17]. Visually lossless compression requires finding a VLT — the maximum compression level at which distortions are invisible to most observers. In this work, we choose the VLT for which at most 25% of observers can perceive the compression artifacts.

Most compression algorithms rely on hand-crafted methods with a relatively small number of adjustable parameters. With the advent of machine learning, deep neural networks, capable of learning complex relations in an automated manner without the need for explicit assumptions, have also been used for image compression [18]. Their results, however, rely on the quality and quantity of training data.

Image quality and visibility assessment

To evaluate compression algorithms, it is necessary to measure the quality of the decompressed image. Two types of metrics can be used for that task: image visibility metrics [19, 35] and image quality metrics [30, 31, 37]. The latter are also known as image fidelity or image impairment metrics and are typically used to evaluate image and video codecs. These metrics produce a single value representing the magnitude of the distortion for the entire image. They are meant to predict clearly visible (suprathreshold) distortions and can be less accurate for barely visible (near-threshold) distortions. By comparison, visibility metrics, are meant to predict near-threshold distortions. They produce a visibility map with local information about the probability of detecting a difference between two images. Similar to compression algorithms, image quality and visibility metrics can be divided into two categories: hand-crafted [30, 31, 37], and trained using machine-learning methods [4, 10, 11, 20, 24, 35, 38]. Both quality

and visibility metrics have been adapted to operate under different display luminance [1, 19, 20, 35].

Several works have focused solely on predicting the VLT. For example, authors in [3] proposed SUR-Net, a deep Siamese-CNN architecture predicting the satisfied-user-ratio (SUR) curve. For each compressed image, their model predicts the proportion of the population, which would not notice the distortion. The authors first pre-trained the network to predict image quality and then fine-tuned on a smaller dataset to predict SUR. The authors later extended the work to SUR-FeatNet [14], where the pre-trained Inception-V3-CNN is used to extract the features from the images, which are then fed into a smaller network trained to predict the SUR curve. Authors in [15] took a different approach and trained a deep binary classifier. Which, given the distorted and undistorted source image, predicts whether the distortion would be noticed or not.

Existing datasets

To train and test image metrics to predict VLT, one requires a benchmark dataset where, at each quality level, the visibility of distortions is judged by a group of observers. While several image quality and visibility datasets are available, they either do not contain VLTs [23, 34, 35], or present images compressed only with JPEG codec at a single luminance level with at a fixed viewing distance. As such, the authors in [9] collected a subjective dataset, called MCL-JCI, with visually lossless thresholds for images compressed with JPEG codec. The dataset contains 50 source images, with VLTs identified by 30 observers for each image. Overall the dataset was collected with 150 observers. Similarly, authors in [16] created a dataset with VLTs for panoramic images compressed with JPEG codec. Each of the 40 panoramic source images was inspected by at least 25 observers, using a head-mounted display.

Unlike exisiting datasets, the presented dataset in this work consist of VLTs for images compressed by not only JPEG but also WebP standard. More importantly, the proposed dataset includes VTLs based on different viewing distance and display brightness.

Proposed dataset

Our goal was to create a dataset with VLTs for images depicting a varied selection of contents and compressed with two codecs (JPEG and WebP), which were viewed on monitors of different peak luminance (10 cd/m^2 and 220 cd/m^2), and at different viewing distance (30 ppd and 60 ppd). We captured 20 images with 1920×1281 resolution, which were obtained from DSLR RAW images and stored in a lossless format. Half of the images were compressed with JPEG (libjpeg[2]) and the other half with WebP (libwebp[3]). Since it was more important for us to capture the variety of content than to compare both codecs, we did not attempt to collect VLT for the same contents and both codecs. To ensure that the observers could find the distortions in a reasonable amount of time, we cropped the stimuli to 512×512 pixels. Examples of the images from the dataset are given in Figure 4. To uniformly cover the entire range of compression quality values at a reasonable number of points, we incremented QF from 2 to 100 in steps of 2, where 100 is an image compressed with the highest quality and also the highest bit-rate.

[2] https://github.com/LuaDist/libjpeg

[3] https://github.com/webmproject/libwebp

(a) VLT experiment stage 1 (b) VLT experiment stage 2

Figure 1: Two stages of visually lossless compression experiment. The images were displayed at the native display resolution without upscaling (images are shown out of scale for better legibility).

Procedure

To improve the accuracy of the VLT measurements, we combined the method of adjustment and 4-alternative-forced-choice (4AFC) protocol, controlled by an adaptive procedure. In the first method-of-adjustment stage, observers were presented with reference and compressed images presented side-by-side and were asked to adjust the QF of the compressed image until they could not distinguish it from the reference (Figure 1a). A half-a-second blank with a middle-gray background was displayed when changing compression levels to avoid observers relying on temporal changes to guide their choice. The compression level found in the first stage was used as an initial guess for the 4AFC procedure [8], in the second stage. Observers were shown three copies of an undistorted reference image and one distorted image in a random quarter, as shown in Figure 1b. They were then asked to choose the distorted image by clicking on the spots where they could see the distortions. The location of those mouse clicks for all four viewing conditions are shown in Figure 4. The next value of the QF parameter for the distorted image was then selected with the QUEST adaptive procedure [32]. We collected between 20 and 30 4AFC trials per observer for each image. To find the VLT for each image we fitted a psychometric function to the collected data.

Display

The experiments were conducted in a dim room ($\sim$10 Lux). The screen was positioned to minimize ambient light reflections. The viewing distance was controlled with a chin-rest. Observers viewed a 27" Asus PG279Q display (2560x1440) from a distance of 80 cm (angular resolution: 60 pixels per degree) or of 40 cm (angular resolution: 30 pixels per degree). For the bright condition, display brightness was set to its maximum value (220 cd/m^2). For the dark condition, we placed a 1.2 neutral density filter (Rosco E-Colour 299) in front of the monitor and adjusted the display brightness so that the effective peak luminance was 10 cd/m^2. We verified that the display color calibration conformed with ITU-R recommendations [7] and sRGB transfer function.

Observers

We collected the data from 19 observers aged between 20 and 30 years old, with normal or corrected-to-normal vision. All

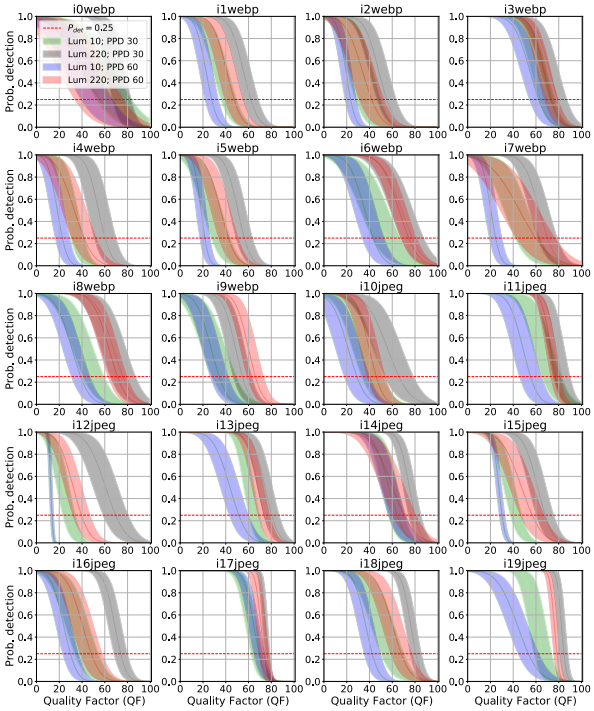


Figure 2: Distribution of visually lossless thresholds across the population. The shaded regions are 99% confidence intervals (due to limited sample size). The dashed red lines denote the detection threshold used to find the VLT.

observers were trained and paid for their participation and were naïve to the purpose of the experiment.

Data analysis

Before analyzing the data, we removed outliers. For each scene we followed the standard z-score procedure and removed VLT measurements which were more than two standard deviations away from the mean.

VLT distribution

We estimate VLT distribution across the population assuming it to be normally distributed, similarly as in [3]. The proportion of the population that can detect the distortion is, thus, described by the function:

$$P_{det}(l) = 1 - \Phi(l; \mu, \sigma^2), \tag{1}$$

where l is the JPEG/WebP quality factor and $\Phi(l; \mu, \sigma)$ is the cumulative normal distribution with the estimated mean and variance of VLT distribution for each condition. The plots of those functions for each image are shown in Figure 2.

The plots of probability of detection in Figure 2 show a clear pattern, with the brighter display (220 cd/m^2) and shorter viewing distance (30 ppd) requiring the highest quality factor. The opposite is shown for the darker display (10 cd/m^2) seen from a larger distance (60 ppd). The slope of those curves indicates how the VLT varied between the observers. The slope is similar for most tested conditions, with a few exceptions. For example, the slope is shallower for *i7webp* with 220 cd/m^2 and 60 ppd (red), indicating a higher variance between the observers. Figure 4 (red dots)

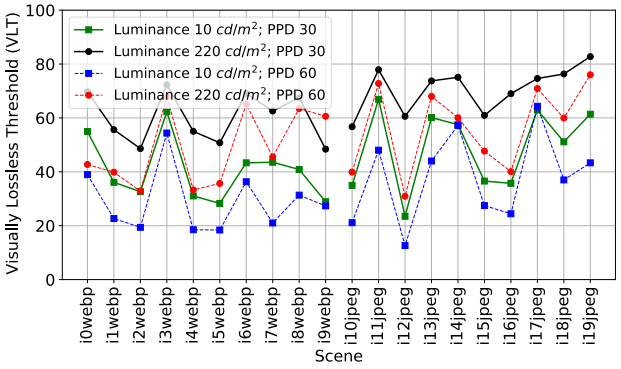


Figure 3: Visually lossless threshold (VLT) for all four viewing conditions shown in different colors. Note that the color used are consistent with Figures 2 and 4.

shows that for this image, the distortions were spotted in different parts of the image for different observers (sky, trees, grass, people), which could explain the variability. Therefore, we opt to use a lower value of P_{det} to find a VLT. This way the VLT reflects the results for the most attentive observers, who could spot the most critical part of an image. Another interesting case is image *i17jpeg*, for which the curves are close together and the slopes are steep (low inter-observer variance). As seen in Figure 4, the distortions for that image were consistently detected by most observers in a large smooth area of the sky.

For each condition the VLT is found by selecting the quality factor l for which $P_{det} = 0.25$ (less than 25% of the population can see the difference). Such *population* VLT values are shown in Figure 3. The display brightness (circles vs. squares) seems to have a higher impact on the VLT than viewing distance (continuous vs. dashed lines). Another interesting observation is that some images were less affected by the display brightness (*i3webp*, *i11jpeg*, *i17jpeg*) than the others (*i12jpeg*, *i16jpeg*). As shown in Figure 4, the distortions in less affected images were typically spotted in bright and smooth regions, for which Weber's law can compensate for the loss of display brightness. It is also important to note that the changes in VLTs between the viewing conditions are different for each image, suggesting a strong interaction between image content and viewing conditions.

Evaluation of image metrics

In this section, we evaluate how accurately image quality and visibility metrics can predict VLTs. The predicted VLTs then can be used to automatically adjust compression parameters to achieve a trade-off between image visual quality and bit-rate. We evaluated both hand-crafted image quality metrics (PSNR, SSIM [30], MS-SSIM [29], FSIM [36], HDRVQM [22]) and one metric based on machine learning (deep photometric image quality metric: PU-PieApp) [20]. Additionally, we evaluate two visibility metrics: a high dynamic range visible difference predictor (HDRVDP3) [19,21], and a CNN-based deep photometric visibility metric (DPVM) [35]. Note that for the HDRVDP3, we report the results of both the visibility predictor (labelled HDRVDP3V) and the quality predictor (labelled HDRVDP3Q). We could not include metrics intended for VLT prediction in [15] and [3], as we could not access the trained models. Furthermore, correspondence with the authors revealed that the models were trained on

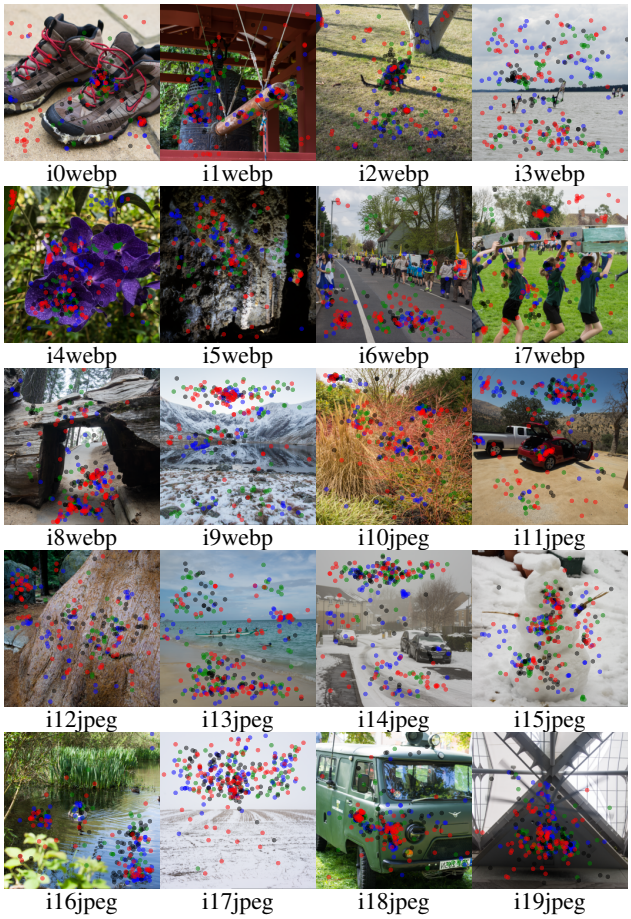


Figure 4: The images used in the experiment and the distributions of observer clicks (blue for 10 cd/m^2, 60 ppd, green for 10 cd/m^2, 30 ppd, black for 220 cd/m^2, 30 ppd and red for 220 cd/m^2, 60 ppd). Colors are consistent with Figures 2 and 3.

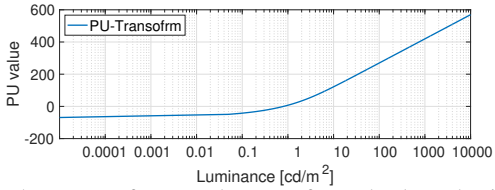


Figure 5: PU-transform used to transform absolute physical values (in cd/m^2) into approximately perceptually uniform units that can be used with existing quality metrics. Note, some values are negative – to ensure that the luminance range of a typical (sRGB) monitor of 1 cd/m^2 to 80 cd/m^2 is mapped to the range 0–255.

very small datasets and thus could not generalize well beyond the training data.

Luminance-aware metrics

Hand-crafted image quality metrics, PSNR, SSIM, MS-SSIM and FSIM were not designed to account for display luminance. Aydın et al. proposed a simple method for adapting existing quality metrics to operate on different display luminance levels [1]. The image is first transformed to luminance emitted from a display assuming a model of that display. The luminance values are then converted into the Perceptually Uniform (PU) or logarithmic values. In our work, we use the PU transform, as it has proven to achieve better results than the logarithmic coding [12]. The shape of the PU-transform is shown in Figure 5.

To transform standard-dynamic-range (SDR) images from gamma-encoded sRGB colors to linear RGB values shown on the high-dynamic-range (HDR) display, gain-gamma-offset display model is often used [5]:

$$C_{\text{lin}} = (L_{\text{peak}} - L_{\text{black}}) C_{\text{sRGB}}^{\gamma} + L_{\text{black}}, \quad (2)$$

where C_{lin} is the linear color value, C_{sRGB} is the gamma-encoded color value for one of the channels (R, G or B), and $\gamma = 2.2$. In our experiments we set the peak luminance of the display to $L_{\text{peak}} = 10$ cd/m^2, or $L_{\text{peak}} = 220$ cd/m^2, and the black level L_{black} was set to $0.001\,L_{\text{peak}}$ (assuming that the contrast of the display was 1000:1 and there were no ambient light reflections) for both cases. We used PU encoding to adapt PSNR, SSIM and FSIM to different luminance conditions. Other metrics (HDRVDP-3, DPVM, HDRVQM and PU-PieApp,) are photometric by design, and thus do not require the application of PU-transform.

Viewing distance-aware metrics

To account for viewing distance in VLT prediction, we followed the procedure in [35]. We re-sampled images with angular resolution of 30 ppd, to the angular resolution of 60 ppd with Lanczos filter [27] before passing them to the metrics. 60 ppd is the highest resolution in our dataset and also a reasonable limit for most visual tasks, since the sensitivity of visual system drops rapidly below 30 cpd [2]. We did not perform the re-sampling for HDRVDP3, as it automatically accounts for the viewing distance.

Metrics validation

To validate the metrics, we performed 5-fold cross-validation. Our goal was to find a mapping between the quality score predicted by the metrics and the VLTs for $P_{det} = 0.25$. We also experimented with different threshold P_{det} values and obtained similar results (not reported here).

We assume that a metric prediction can be mapped to the predicted probability of detection $\tilde{P}_{det}$ by a monotonic function f:

$$\tilde{P}_{det} = f(M(T_l, R)), \quad (3)$$

where M is the quality metric, R is the reference image and T_l is the test image encoded at the quality level l. Therefore, the predicted VLT $\tilde{l}$ can be found as:

$$\tilde{l} = \underset{l}{\text{argmin}} \, ||f(M(T_l, R)) - 0.25||_2 \quad \wedge \quad l \in 2, 4, .., 100. \quad (4)$$

Since we are interested in the single value of VLT rather than finding the function f, we instead estimate the VLT using a single (per metric) value q_{VLT}:

$$\tilde{l} = \underset{l}{\text{argmin}} \, ||M(T_l, R) - q_{VLT}||_2 \quad \wedge \quad l \in 2, 4, .., 100, \quad (5)$$

which we optimize per metric so that $||\tilde{l} - l||_2$ is minimized (where l is the measured VLT). The individual q_{VLT} was found for each fold so that the results are reported for the testing set.

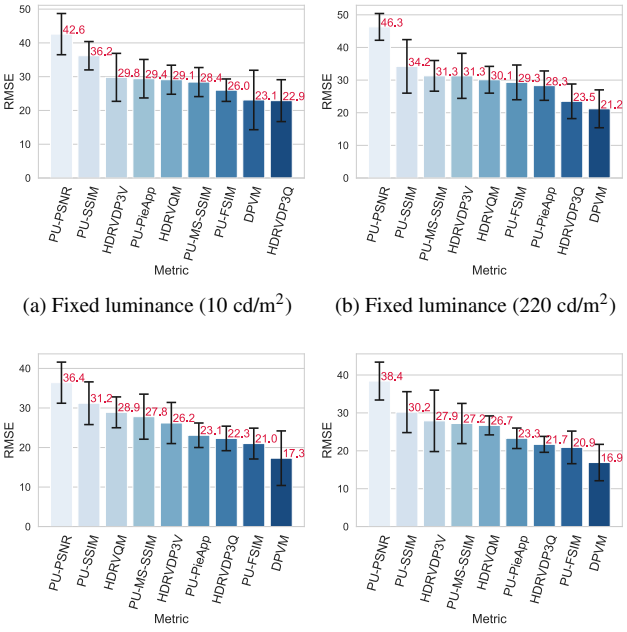
Since HDRVDP3V and DPVM produce a map of detection probabilities, rather than a single quality value, we consider a percentile value from the probability map to be a prediction. For DPVM, similar to [33], we search for the optimal percentile that minimizes root-mean-squared-error (RMSE) between the predicted and measured VLT. The best percentile for DPVM and HDRVDP3V were 86 and 97 respectively.

Results and discussion

We first explore how well each of the metrics can account for the viewing distance and the display brightness. The metrics that use PU transform are prefixed with PU-. For each experiment we fix one of the viewing conditions and let the metric predict the VLT, while accounting for another viewing condition (e.g., how well the metric can account for the viewing distance at a fixed luminance level of 10 cd/m^2). The results of the 5-fold cross-validation are shown in Figure 6. Metrics, which were designed to account for the viewing distance (HDRVDP3Q and DPVM) match the experimental VLT better across different viewing distances compared to the metrics that used a simple resampling. When viewing distance is fixed and metrics are trained to account for the display brightness, DPVM performs the best among the tested metrics. It is followed by three metrics with comparable performance: PU-FSIM, HDRVDP3Q and PU-PieApp. In general, all metrics better account for the display luminance changes, compared to changes in viewing distance. In all experiments the visibility predictor HDRVDP3V performed unexpectedly worse than HDRVDP3Q. This implies that HDRVDP3V requires fine-tuning for more accurate perfomance.

Figure 7a shows the results of the 5-fold cross-validation for all viewing conditions, i.e., each metric needed to account for both display brightness and viewing distance. The results show that DPVM is the best performing metric, followed by five metrics of comparable performance (verified in the paired ttest): HDRVDP3Q, PU-PieApp, PU-FSIM, HDRVQM and PU-MS-SSIM. The visibility predictor HDRVDP3V performed worse than its quality counterpart. It is also worth noting a relatively good performance of PU-FSIM, as this is a hand-crafted metric that has not been trained on this task. The RMSE of the DPVM is 21.9, which is slightly larger, than the average variation of the VLT across the population – 15.9 (refer to Figure 2). This difference can be acceptable, and the metric should be robust enough to adaptively encode images displayed at different luminance levels and at a different viewing distance in a visually lossless manner.

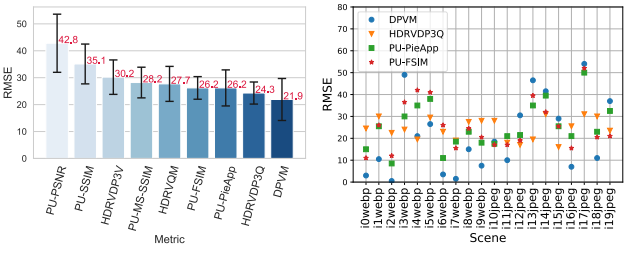
Figure 7b shows the RMSE per image for the four best performing metrics. It is worth noting that while DPVM resulted in substantially smaller RMSE for some images, it was the worst performing metric for others. The performance of the DPVM did not correlate with the folds. The hand-crafted metrics with a few trainable parameters, HDRVDP3Q and PU-FSIM, tend to be more consistent and vary less in RMSE than machine-leaning based metrics, PI-PieApp and DPVM. Most metrics (except HDRVDP3Q) showed a very large error for images with large smooth gradient areas (e.g., sky), *i3webp,i4webp,i5webp,i13jpeg, i14jpeg and i17jpeg*, suggesting that those metrics could be worse at modeling contrast masking.



(a) Fixed luminance (10 cd/m^2) (b) Fixed luminance (220 cd/m^2)

(c) Fixed distance (30 ppd) (d) Fixed distance (60 ppd)

Figure 6: Five-fold cross-validation results for the compared metrics under varied viewing conditions. We fix one the viewing conditions and make the metric account for the remaining condition. In the first row we isolate distance as the component impacting the VLT and test whether metrics can take it into account under fixed display luminance. Similarly, in the second row, we fix the distance, and let the metrics account for the display luminance. Error bars show the standard deviation.



(a) (b)

Figure 7: (a) Five-fold cross-validation results for the compared metrics. Error bars show the standard deviation. (b) RMSE per image.

Conclusions

We carefully collected a novel dataset for visually lossless image compression under varying display brightness and viewing distance conditions and compared the performance of the state-of-the-art image quality and visibility metrics on this dataset. The results indicate that the display brightness and the viewing distance can significantly affect the compression level required for visually lossless coding. We found that, recently proposed deep photometric visibility metric (DPVM) is the best at matching the experimental thresholds for the collected dataset.

Acknowledgments

We would like to thank the volunteers who participated in our experiments and Maryam Azimi for the insightful comments on the manuscript. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme — grant agreement N° 725253 (EyeCode).

References

[1] Tunç Ozan Aydın, Rafał Mantiuk, and Hans-Peter Seidel. Extending quality metrics to full dynamic range images. In *Human Vision and Electronic Imaging XIII*, Proc. of SPIE, pages 6806–10, San Jose, USA, January 2008.

[2] Peter Barten. *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*. SPIE Press, 1999.

[3] C. Fan, H. Lin, V. Hosu, Y. Zhang, Q. Jiang, R. Hamzaoui, and D. Saupe. 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)*, pages 1–6, June 2019.

[4] Fei Gao, Yi Wang, Panpeng Li, Min Tan, Jun Yu, and Yani Zhu. Deepsim: Deep similarity for image quality assessment. *Neurocomputing*, 257:104 – 114, 2017. Machine Learning and Signal Processing for Big Multimedia Analysis.

[5] Rolf R. Hainich. *Perceptual display calibration*. CRC Press, 2016.

[6] A.J. Hussain, Ali Al-Fayadh, and Naeem Radi. Image compression techniques: A survey in lossless and lossy algorithms. *Neurocomputing*, 300:44 – 69, 2018.

[7] ITU-R. Parameter values for the hdtv standards for production and international programme exchange. ITU-R Recommendation BT.709-6, Mar 2015.

[8] Frank Jakel and Felix A. Wichmann. Spatial four-alternative forced-choice method is the preferred psychophysical method for naive observers. *Journal of Vision*, 6(11):13–13, 11 2006.

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

[10] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, and A. C. Bovik. Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment. *IEEE Signal Processing Magazine*, 34(6):130–141, Nov 2017.

[11] Jangyoo Kim and Sanghoon Lee. Deep learning of human visual sensitivity in image quality assessment framework. *Computer Vision and Pattern Recognition (CVPR)*, pages 1969–1977, 2017.

[12] P. Korshunov, P. Hanhart, T. Richter, A. Artusi, R. Mantiuk, and T. Ebrahimi. Subjective quality assessment database of HDR images compressed with jpeg xt. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 1–6, May 2015.

[13] Pavel Korshunov and Touradj Ebrahimi. Context-dependent JPEG backward-compatible high-dynamic range image compression. *Optical Engineering*, 52(10):102006, aug 2013.

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

[15] H. Liu, Y. Zhang, H. Zhang, C. Fan, S. Kwong, C. . J. Kuo, and X. Fan. Deep learning-based picture-wise just noticeable distortion prediction model for image compression. *IEEE Transactions on Image Processing*, 29:641–656, 2020.

[16] Xiaohua Liu, Zihao Chen, Xu Wang, Jianmin Jiang, and Sam Kowng. JND-Pano: database for just noticeable difference of JPEG compressed panoramic images. In Richang Hong, Wen-Huang Cheng, Toshihiko Yamasaki, Meng Wang, and Chong-Wah Ngo, editors, *Advances in Multimedia Information Processing – PCM 2018*, pages 458–468. Springer International Publishing, 2018.

[17] Vladimir Lukin. Performance analysis of visually lossless image compression. *Conference on Video Processing and Quality Metrics for Consumer Electronics*, 01 2012.

[18] Siwei Ma, Xinfeng Zhang, Chuanmin Jia, Zhenghui Zhao, Shiqi Wang, and Shanshe Wanga. Image and video compression with neural networks: A review. *IEEE Transactions on Circuits and Systems for Video Technology*, PP:1–1, 04 2019.

[19] Rafał Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.*, 30(4), July 2011.

[20] Aliaksei Mikhailiuk, Maria Perez-Ortiz, Dingcheng Yue, Wilson Suen, and Rafal Mantiuk. Consolidated dataset and metrics for high-dynamic-range image quality assessment. In *Under review for IEEE Transactions on Multimedia (TMM)*, 2020.

[21] Manish Narwaria, Rafal K. Mantiuk, Mattheiu Perreira Da Silva, and Patrick Le Callet. 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, jan 2015.

[22] Manish Narwaria, Matthieu Perreira Da Silva, and Patrick Le Callet. Hdr-vqm: An objective quality measure for high dynamic range video. *Signal Processing: Image Communication*, 35:46 – 60, 2015.

[23] 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. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, jan 2015.

[24] Ekta Prashnani, Hong Cai, Yasamin Mostofi, and Pradeep Sen. Pieapp: Perceptual image-error assessment through pairwise preference. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[25] A. Raid, Wael Khedr, Mohamed El-dosuky, and Wesam Ahmed. Jpeg image compression using discrete cosine transform - a survey. *International Journal of Computer Science and Engineering Survey*, 5, 05 2014.

[26] Zhanjun Si and Ke Shen. *Research on the WebP Image Format*, pages 271–277. Springer, 12 2016.

[27] Ken Turkowski. *Filters for Common Resampling Tasks*, page 147–165. Academic Press Professional, Inc., USA, 1990.

[28] Gregory K. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, pages 30–44, 1991.

[29] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, 2003.

[30] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[31] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simon-

celli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[32] Andrew B. Watson and Denis G. Pelli. Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2):113–120, Mar 1983.

[33] K. Wolski, D. Giunchi, S. Kinuwaki, P. Didyk, K. Myszkowski, A. Steed, and R. K. Mantiuk. Selecting texture resolution using a task-specific visibility metric. *Computer Graphics Forum*, 38(7):685–696, 2019.

[34] Krzysztof Wolski, Daniele Giunchi, Nanyang Ye, Piotr Didyk, Karol Myszkowski, Radosław Mantiuk, Hans-Peter Seidel, Anthony Steed, and Rafał K. Mantiuk. Dataset and metrics for predicting local visible differences. *ACM Trans. Graph.*, 37(5), November 2018.

[35] Nanyang Ye, Krzysztof Wolski, and Rafał K. Mantiuk. Predicting visible image differences under varying display brightness and viewing distance. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019 (CVPR 2019)*, 2019.

[36] L. Zhang, L. Zhang, X. Mou, and D. Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, Aug 2011.

[37] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, Aug 2011.

[38] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep networks as a perceptual metric. In *CVPR*, 2018.

Author Biography

Aliaksei Mikhailiuk has recently received his PhD in Computer Science under the supervision of Dr. Rafal Mantiuk. Nanyang Ye is currently an assistant professor at Shanghai Jiao Tong University. He received his PhD at the University of Cambridge, UK, having previously received his MA degree from Tsinghua University, China.

Rafał K. Mantiuk is Reader (Associate Professor) at the Department of Computer Science and Technology, University of Cambridge (UK). He received PhD from the Max-Planck-Institute for Computer Science (Germany). His recent interests focus on computational displays, novel display technologies, rendering and imaging algorithms that adapt to human visual performance and viewing conditions in order to deliver the best images given limited resources, such as computation time, bandwidth or dynamic range. He contributed to early work on high dynamic range imaging, including quality metrics (HDR-VDP), video compression and tone-mapping. Further details: `http://www.cl.cam.ac.uk/~rkm38/`.

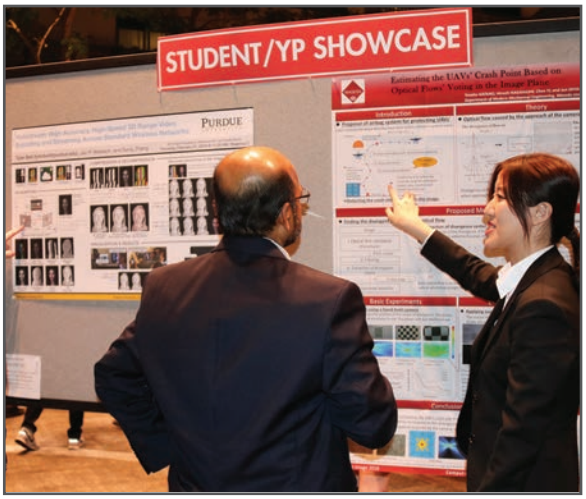
JOIN US AT THE NEXT EI!

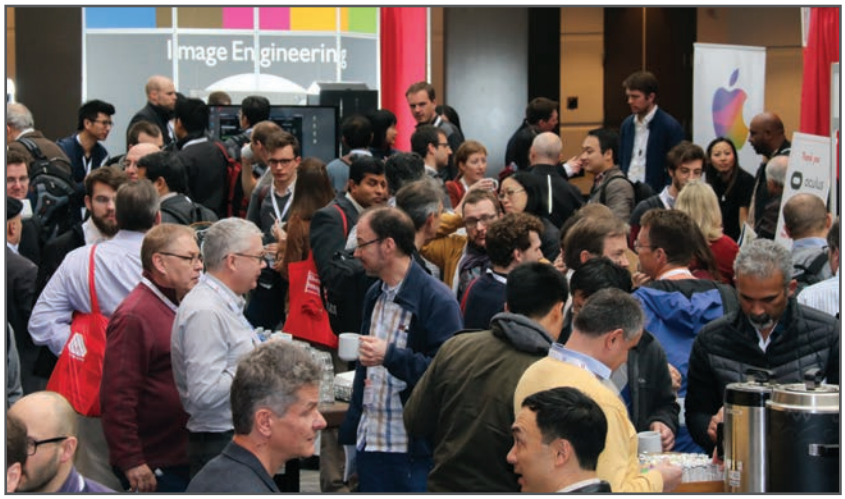
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org