https://doi.org/10.2352/EI.2022.34.9.IQSP-317

Image enhancement dataset for evaluation of image quality metrics

Altynay Kadyrova, Marius Pedersen, Bilal Ahmad, Dipendra J. Mandal, Mathieu Nguyen, Pauline Hardeberg Zimmermann; Department of Computer Science, Norwegian University of Science and Technology, Gjøvik, Norway

Abstract

Image enhancement is important in different application areas such as medical imaging, computer graphics, and military applications. In this paper, we introduce a dataset with enhanced images. The images have been enhanced by five end users, and these have been evaluated by observers in an online image quality experiment. The enhancement steps by the end users and subjective results are analysed in detail. Furthermore, 38 image quality metrics have been evaluated on the introduced dataset to reveal their suitability to measure image enhancement. The results show that the image quality metrics have low to average performance on the new dataset.

Introduction

Image enhancement plays a vital role in a variety of imaging fields. Image capture enhancement is often performed to create visually pleasing images and extensive research has been carried out to introduce image enhancement techniques. Image enhancement is nowadays not only performed by the imaging device, but also by the end user with post-processing. This post-processing can, for example, be sharpening, adjusting brightness, saturation or contrast. Most cameras or mobile phones allow for easy post-processing. With the ease of enhancing the quality of the captured image, it is also valuable to check if the image has been improved. The quality evaluation can be subjective as well as objective. It is assumed that the subjective evaluation is a gold standard. However, Image Quality Metrics (IQMs) that provide objective evaluation can be more advantageous than subjective evaluation as they are easy to compute and does not require observer involvement.

In this paper, we analyse the post-processing by the end users to understand how they enhance images as well as analysing the judgement made by the observers to these images. At last, we evaluate the performance of the IQMs on the enhanced images.

Obtaining a better understanding of how images are enhanced and how observers judge enhanced images can help in image enhancement, but also contribute to improving the IQMs for enhanced images. Also, there are a few datasets for evaluating the performance of the IQMs on enhanced images [1]. As a result, the goal is to introduce a new dataset with enhanced images together with an in-depth analysis of the dataset, but also to evaluate a set of IQMs on the new dataset.

This paper is organised as follows: works about quality evaluation of enhanced colour images are described first, followed by our dataset description which includes the subjective experiment. Afterwards, selected state-of-the-art IQMs are presented. Finally, we present our results and discussion followed by conclusions and future works.

Related works

Image enhancement is one of the steps of image processing with the aim to improve the image quality [2]. For example, noise can be removed as much as possible or brightness can be adjusted. However, there is still much work to do on quality evaluation of enhanced images. According to Vu et al. [3], the artistic impression of an image might influence the quality evaluation of enhanced images. Also, the image enhancement level within the image might have low level of enhancement which makes quality evaluation of enhanced images challenging. Nevertheless, there are studies which try to tackle this challenge. Fairchild and Johnson [4] proposed usage of colour appearance models to evaluate the quality of enhanced images. The majority of studies focus on specific quality attribute enhancement or combination of two or more quality attributes. For instance, contrast is one the most important quality attribute [5] and Wang et al. [6] created a framework for guided image contrast enhancement which can output visually appealing enhanced images automatically. The enhancement of colour, contrast, sharpness, and brightness were combined in Vu et al. work [3] and the authors created a dataset (named digitally retouched image quality) based on these enhancements. An evaluation measure on enhancement of contrast was proposed by Qureshi et al. [7] where the authors created a dataset of enhanced images. There are a limited number of datasets on image enhancement [1]. Contrast Enhancement Evaluation Database (CEED) [8] was created for testing different contrast enhancement techniques. Similar to the CEED, Colourlab Contrast Enhanced Image Dataset (CCEID) [1] focused on contrast enhanced images. However, in these datasets an in-depth analysis of the subjective results was not performed which makes it difficult to understand why certain enhancements are preferred over others.

In order to evaluate the quality of enhanced images objectively, various IQMs can be applied. According to Amirshahi et al. [1], IQMs are mostly applied on degraded images rather than on enhanced images. They evaluated more than twenty IQMs (Full Reference (FR) for colour images, No Reference (NR) for colour images, FR for greyscale images, NR for greyscale images) on contrast enhanced images. Cheng et al. [9] found that the performance of the FR metrics was lower in comparison with the NR metrics on sharpness enhanced images.

Methodology
Dataset of enhanced images

The dataset contains 16 natural colour images (i.e., original) where 2 images were reproduced from the CCEID dataset [1] (Figure 1). The sRGB images were resized to 600x500 pixels. The original images have an average quality and a wide range of spatial information and colourfulness [10]. Moreover, they contain large areas of the same colour, neutral grey, different levels of hues, contrast, brightness, and memory colours. They were enhanced using the edit function of Instagram by five end users in an uncontrolled environment, who were asked to enhance the quality of the images according to their perception. The following five attributes were chosen from Instagram: brightness, contrast, saturation, sharpness, and warmth. All have a range of [-100 100] values, except the sharpness which has a range of [0 100]. The resulting dataset of original and enhanced images consists of 96 images.

The subjective experiment was done using the web platform QuickEval [11]. In other words, we performed an online experiment with undefined information about illumination, viewing distance, and display details of the observers. We chose to do the online experiment even though there is a debate on online versus offline experiment conditions. The offline experiments are controlled which is good for reproducability purposes while there can be variations in screen resolution, lighting and viewing conditions during the online experiments. Performing online experiment allows for more and diverse observers which is important for statistical significance of data. The experiment was done as a forced choice pair comparison due to its simplicity in use. Given a pair of images, the observers were asked to select the image which they perceived as having the highest quality. We also included the original 16 images into the experiment. However, the observers were not informed about it, giving the observers the original and five enhanced versions to evaluate.

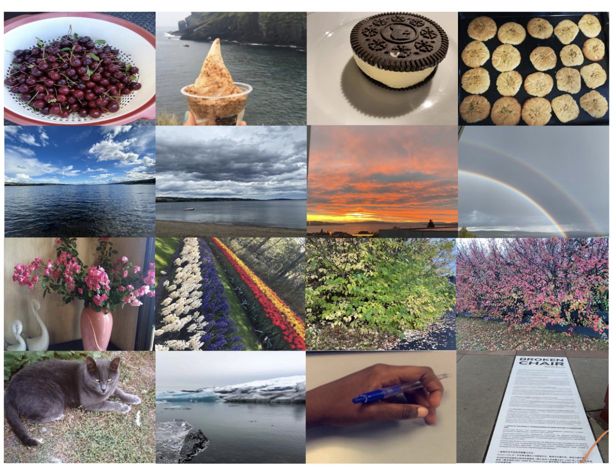

Figure 1. *Original images in the dataset.*

The experiment was divided into eight sessions with the aim of reducing the duration of the experiment for the observers. The observers were free to choose to do just one session or more, where one session lasted between 8 and 12 minutes on average. Some images were repeated in different sessions for the purpose of analysis. The images were randomised for each observer in each session. In total, we had 45 different observers. Among them, only one observer completed all eight sessions. There were 15 observers per session on average following the CIE guideline [12]. The judgements from the observers were processed into z-scores [13].

Image quality metrics

We evaluate a wide range of FR and NR IQMs on the dataset. These have been selected to represent different categories of IQMs, from those based on structural similarity to those measuring aesthetic quality. These are also selected based on their performance in different studies [14, 15], where they have been compared against subjective data.

The FR IQMs used in our study are as follows:

- Structural Similarity Index Metric (SSIM), which is based on the Human Visual System's (HVS) ability to extract scene structural information [16] and Pedersen et al. [17] stated that it is able to evaluate attributes such as sharpness, contrast, lightness, and artefacts.
- Visual Saliency based Index (VSI), which is based on assumption that there is a relation between image's perceptual quality and its saliency map [18].
- Edge Strength Similarity (ESSIM), which is used for distorted images and based on assumption that pixels' edge strength can fully represent image's semantic information in order to derive perceptual fidelity between reference and distorted images' semantic information [19]. It is chosen because it considers semantic information which is important aspect for enhanced images.
- Feature Similarity (FSIM) and the FSIMc (colour), which are based on the low-level features of the HVS [20].
- Visual Information Fidelity (VIF), which quantifies the image information loss happened due to a distortion process [21].
- Color Image Difference (CID), which is a colour extension of SSIM [22] and has shown to provide good performance compared to other IQMs [14].
- Convolutional Neural Networks Quality (CNNQ), which is based on comparing convolutional neural network features between reference and distorted images [23].
- Blur, which uses a local neighbourhood to estimate the amount of blur between original and distorted images [24].
- Spatial Hue Angle Metric (SHAME-II), which calculates the colour difference with consideration of the HVS's spatial features [25].
- Total Variation of Difference (TVD-TV-NORM), which works with a proposed normalization step to be robust to scale differences between images of different content [26].

The NR IQMs are as follows:

- Cumulative Probability of Blur Detection (CPBD), which is a perceptual-based no-reference image sharpness metric [27].
- Anistropic Quality Index (AQI), which is based on measuring the variance of entropy in different directions [28].
- Blind Image Quality Assessment through Anisotropy (BIQAA), which measures blur and Gaussian noise through anisotropy [28].
- Autoregressive-based Image Sharpness Metric luminance (ARISMCL) and the ARISMCc (colour), which measure image sharpness by considering luminance and chrominance information in the autoregressive parameter space, respectively [29].

- Spectral and Spatial Sharpness (S3), which is a sharpness measure based on image's spectral and spatial properties [30].
- JPEGSCORE, which analyses features that capture artefacts introduced by JPEG compression [31].
- Just Noticeable Distortion Discrete Cosine Transform (JNDDCT), which operates on proposed new formula for luminance adaptation adjustment [32].
- Blind Image Integrity Notator using Discrete cosine transform Statistics (BLIINDS2), which uses a Bayesian inference model to predict quality scores for specific features [33].
- Blur Metric, which predicts the amount of blur by investigating variations of neighbouring pixels [34].
- CPCQI, which measures contrast-distorted images based on information maximisation [35].
- Entropy-based No-reference Image Quality Assessment (ENIQA), which is based on entropy where image features are extracted from the spatial and frequency domains [36].
- Fog Aware Density Evaluator (FADE), which predicts the visibility of a foggy scene [37].
- Feature Maps–Based Referenceless Image Quality Evaluation Engine (FRIQUEE), which predicts authentically distorted images' perceptual quality [38].
- Gradient Information Filter (GIF), which is based on gradient information and a filter of the HVS [39].
- High Order Statistics Aggregation (HOSA), which uses K-means clustering, diagonal covariance and coskewness, and support vector regression to find a mapping between the subjective scores and image's perceptual features [40].
- JPEG2000, which is based on natural scene statistics for measuring the quality of JPEG2000 compression [41].
- FFTJPEG and FFTJPEGm, which measure JPEG compression in images [42].
- Natural Factor, which measures the naturalness factor of an image [43].
- Natural Image Quality Evaluator (NIQE), which is based on the natural scene statistics and requires many training data of human opinions on images which are distorted [44].
- No Reference JPEG2000 (NRJPEG2000), which is based on spatial features to measure JPEG2000 compression [45].
- Perceptual Sharpness Index (PSI), which is a sharpness metric based on local edge gradient analysis [46].
- Neural Image Assessment [1] (NIMA), which measures both technical and aesthetic qualities of images [47].
- Multi-level Spatially Pooled activation blocks (MLSP), which is a deep learning approach to aesthetics quality assessment [48].
- Just Noticeable Blur Metric (JNBM), which can estimate image blurriness across different content [49].
- Iterative Estimation in Discrete cosine transform Domain (IEDD), which works with highly textured images to estimate noise variance [50].

For more information on the IQMs, we refer the reader to the respective publications.

[1] We use the PyTorch implementation from `https://github.com/kentsyx/Neural-IMage-Assessment`

Results

We analyse the enhancement to the images by the end users first, then the judgements by the observers, and at last the performance of the IQMs.

Analysis of enhancement

The contrast was the attribute with the most and largest changes by the end users, followed by brightness, saturation, warmth, and sharpness. These were changed in 67, 44, 37, 27, and 18 images, respectively. In 27 images only one attribute was modified, in 28 images two attributes were modified, in 17 images three attributes, in 5 images four attributes, and one image with changes in all five attributes. This indicates that with modification in one or two attributes the end users were able to produce a result they were pleased with. One end user modified only two attributes for the enhancement while the remaining four end users did enhancement using all attributes. From our observations the type of enhancement can be linked with the content of the images.

The end users changed the images in a range from [-100,100] in all but the sharpness attribute. On average contrast, warmth, and saturation have been positively increased while brightness has been slightly decreased. It is also worth noting that brightness had the most spread when it came to changes. In Figure 2, we show a boxplot of the distribution of changes done by the end users to the images.

We also investigated the pairwise correlation between the changed attributes and there was a little or no correlation between the attributes, indicating that the end users did not modify them in combination. In a similar way, we investigated the changes between the end users who modified the images, and the correlation is weak (maximum 0.39).

Analysis of subjective scores

The original was not judged to have the highest quality by the observers (i.e., it did not have the highest z-score in any image) and in three images it had the lowest z-score. In 6 images the original had a positive z-score (higher than 0) and in 10 images it had a negative z-score (lower than 0). This shows that an enhanced version was judged to have the highest quality in each image. The five end users have created images that have both low and high z-scores, indicating that the observers did not prefer or dislike the enhancements by a single end user.

There was no clear correlation between the changes of each attribute and the subjective scores. There is also no clear indication that smaller or larger changes are linked with the responses from the observers.

Investigation of the correlation between the subjective scores and the enhancements for each image has been done. Some observations can be made. In certain images where the modification is large, the quality of the image goes down. This is apparent in images where the saturation has been changed making the images too saturated and they start to look unnatural. There are other images, especially those with memory colours such as grass, fruits, and sky, where an increase in saturation is linked with higher quality as long as it is limited to avoid making the image unnatural. This is aligned with the findings of Siple and Springer [51] where people preferred objects which were saturated. In specific images sharpening is correlated with the subjective scores, being in images where details are important. Based on the observations from

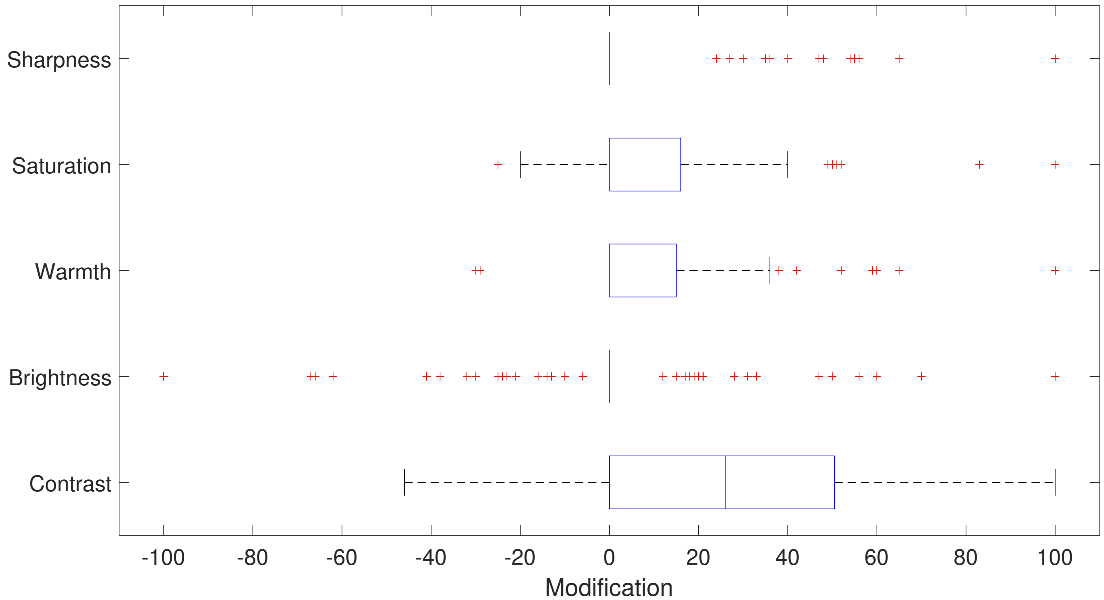

Figure 2. *Boxplot of distribution of changes in the sharpness, saturation, warmth, brightness, and contrast attributes.*

the experiment, the quality is linked with content of the images where certain modifications are preferred by the observers.

Performance evaluation of image quality metrics

Figure 3 shows the non-linear Pearson correlation with a 95% confidence interval for the IQMs based on a 5-parameter logistic function [52]. As can be seen, all IQMs have medium or low correlation with the subjective scores. The highest correlation values are obtained by CID, TVD-TV-NORM, CPCQI, and SHAME-II, but their confidence intervals overlap with most of the other IQMs. It is interesting to note that these are FR IQMs except CPCQI. Also other FR IQMs (SSIM and VSI) have a higher correlation than most of the NR IQMs. The reason for the low correlation for the FR IQMs is the inclusion of the original image in the experiment, as this always is estimated to have the highest quality by the IQMs, but as indicated in the previous subsection it has not been judged to have the highest quality by the observers. Analysis of the performance of CID, having the highest correlation, reveals that images with more changes by the end users are problematic, as this naturally leads to higher difference from the original, resulting in a low correlation. In many of the IQMs, we also observe problems with scale differences between the image sets which has also been reported as a problem for IQMs [26]. TVD-TV-NORM has the second highest correlation and was made to be more robust to this.

NIMA and MLSP are designed to evaluate aesthetic image quality, but do not have a higher correlation compared to those IQMs dealing with technical quality. We have also analysed the results for each of the nine scenes in the dataset and there are scenes where some IQMs are able to correlate well with the subjective scores. However, none of the IQMs are able to do so for all scenes. These results indicate that the dataset is very challenging for the IQMs.

Conclusions and future works

Image enhancement is a necessary step in many applications. In this work, we have analysed the enhancements by the end users and the judgements by the observers on enhanced images. In addition, we have evaluated the IQMs on a new dataset of enhanced images. The results indicate that enhanced images are challenging task for the IQMs. Our dataset is released at `https://www.ntnu.edu/web/colourlab/software`.

Further work can include creating an image quality metric specifically designed for enhanced images and performing experiment in a controlled environment.

Acknowledgments

This research was funded by the ApPEARS project from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 814158.

We thank observers for their participation in the experiment and Seyed Ali Amirshahi for fruitful discussions.

References

[1] S. A. Amirshahi, A. Kadyrova, and M. Pedersen. How do image quality metrics perform on contrast enhanced images? In *2019 8th European Workshop on Visual Information Processing*, pages 232–237. IEEE, 2019.

[2] D. M. Chandler, M. M. Alam, and T. D. Phan. Seven challenges for image quality research. In *Human Vision and Electronic Imaging XIX*, volume 9014, page 901402. International Society for Optics and Photonics, 2014.

[3] C. T. Vu, T. D. Phan, P. S. Banga, and D. M. Chandler. On the quality assessment of enhanced images: A database, analysis, and strategies for augmenting existing methods. In *2012 IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 181–184, 2012.

[4] M. D. Fairchild and G. M. Johnson. icam framework for image

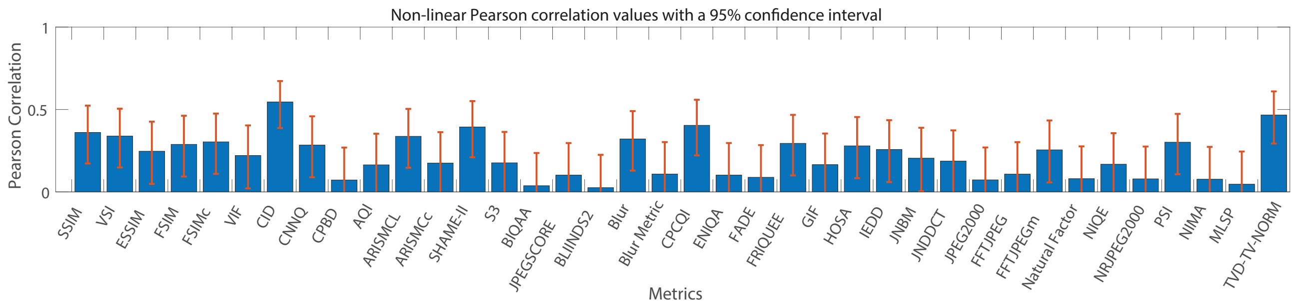

Figure 3. *Non-linear Pearson correlation for the evaluated IQMs.*

appearance, differences, and quality. *Journal of Electronic Imaging*, 13(1):126–139, 2004.

[5] M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albregtsen. Attributes of image quality for color prints. *Journal of Electronic Imaging*, 19(1):011016, 2010.

[6] S. Wang, K. Gu, S. Ma, W. Lin, X. Liu, and W. Gao. Guided image contrast enhancement based on retrieved images in cloud. *IEEE Transactions on Multimedia*, 18(2):219–232, 2015.

[7] M. A. Qureshi, A. Beghdadi, and M. Deriche. Towards the design of a consistent image contrast enhancement evaluation measure. *Signal Processing: Image Communication*, 58:212–227, 2017.

[8] A. Beghdadi, M. A. Qureshi, B. Sdiri, M. Deriche, and F. Alaya-Cheikh. Ceed-a database for image contrast enhancement evaluation. In *2018 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2018.

[9] Y. Cheng, M. Pedersen, and G. Chen. Evaluation of image quality metrics for sharpness enhancement. In *10th International Symposium on Image and Signal Processing and Analysis*, pages 115–120. IEEE, 2017.

[10] S. Winkler. Analysis of public image and video databases for quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 6(6):616–625, 2012.

[11] K. V. Ngo, J. J. Storvik, C. A. Dokkeberg, I. Farup, and M. Pedersen. Quickeval: a web application for psychometric scaling experiments. In *Image Quality and System Performance XII*, volume 9396, page 93960O, 2015.

[12] J. Morovic. Guidelines for the evaluation of gamut mapping algorithms. *CIE*, 153:D8–6, 2003.

[13] P. G. Engeldrum. *Psychometric Scaling: A Toolkit for Imaging Systems Development*. Imcotek press, 2000.

[14] M. Pedersen. Evaluation of 60 full-reference image quality metrics on the cid: Iq. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 1588–1592. IEEE, 2015.

[15] M. Pedersen and J. Y. Hardeberg. Full-reference image quality metrics: Classification and evaluation. *Foundations and Trends® in Computer Graphics and Vision*, 7(1):1–80, 2012.

[16] Z. 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, 2004.

[17] M. Pedersen, Y. Zheng, and J. Y. Hardeberg. Evaluation of image quality metrics for color prints. In *Scandinavian Conference on Image Analysis*, pages 317–326. Springer, Ystad, Sweden, May, 2011.

[18] L. Zhang, Y. Shen, and H. Li. Vsi: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image processing*, 23(10):4270–4281, 2014.

[19] X. Zhang, X. Feng, W. Wang, and W. Xue. Edge strength similarity for image quality assessment. *IEEE Signal processing letters*, 20(4):319–322, 2013.

[20] 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, 2011.

[21] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on image processing*, 15(2):430–444, 2006.

[22] I. Lissner, J. Preiss, P. Urban, M. S. Lichtenauer, and P. Zolliker. Image-difference prediction: From grayscale to color. *IEEE Transactions on Image Processing*, 22(2):435–446, 2012.

[23] S. A. Amirshahi, M. Pedersen, and S. X. Yu. Image quality assessment by comparing cnn features between images. *Electronic Imaging*, (12):42–51, 2017.

[24] M. Elsayed, F. Sammani, A. Hamdi, A. Albaser, and H. Babalghoom. A new method for full reference image blur measure. *International Journal of Simulation: Systems, Science Technology*, 19:4, 2018.

[25] M. Pedersen and J. Y. Hardeberg. A new spatial filtering based image difference metric based on hue angle weighting. *Journal of Imaging Science and Technology*, 56(5):50501–1, 2012.

[26] M. Pedersen and I. Farup. Improving the robustness to image scale of the total variation of difference metric. In *3rd International Conference on Signal Processing and Integrated Networks*, pages 116–121. IEEE, 2016.

[27] N. D. Narvekar and L. J. Karam. A no-reference image blur metric based on the cumulative probability of blur detection (cpbd). *IEEE Transactions on Image Processing*, 20(9):2678–2683, 2011.

[28] S. Gabarda and G. Cristóbal. Blind image quality assessment through anisotropy. *JOSA A*, 24(12):B42–B51, 2007.

[29] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang. No-reference image sharpness assessment in autoregressive parameter space. *IEEE Transactions on Image Processing*, 24(10):3218–3231, 2015.

[30] C. T. Vu, T. D. Phan, and D. M. Chandler. S3: A spectral and spatial measure of local perceived sharpness in natural images. *IEEE transactions on image processing*, 21(3):934–945, 2011.

[31] Z. Wang, H. R. Sheikh, and A. C. Bovik. No-reference perceptual quality assessment of jpeg compressed images. In *International Conference on Image Processing*, volume 1, pages I–I. IEEE, 2002.

[32] X. H. Zhang, W. S. Lin, and P. Xue. Improved estimation for just-noticeable visual distortion. *Signal Processing*, 85(4):795–808, 2005.

[33] M. A. Saad, A. C. Bovik, and C. Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE transactions on Image Processing*, 21(8):3339–3352, 2012.

[34] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas. The blur effect: perception and estimation with a new no-reference perceptual blur metric. In *Human vision and electronic imaging XII*, volume 6492, page 64920I, 2007.

[35] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen. No-reference quality metric of contrast-distorted images based on information maximization. *IEEE transactions on cybernetics*, 47(12):4559–4565, 2016.

[36] X. Chen, Q. Zhang, M. Lin, G. Yang, and C. He. No-reference color image quality assessment: from entropy to perceptual quality. *EURASIP Journal on Image and Video Processing*, 2019(1):77, 2019.

[37] L. K. Choi, J. You, and A. C. Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing*, 24(11):3888–3901, 2015.

[38] D. Ghadiyaram and A. C. Bovik. Perceptual quality prediction on authentically distorted images using a bag of features approach. *Journal of vision*, 17(1):32–32, 2017.

[39] L. Ying, Z. Li, and C. Zhang. No-reference sharpness assessment with fusion of gradient information hvs filter. *Journal of Image and Graphics*, 20(11):1446–1452, 2015.

[40] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann. Blind image quality assessment based on high order statistics aggregation. *IEEE Transactions on Image Processing*, PP(99):1–1, 2016.

[41] H. R. Sheikh, A. C. Bovik, and L. Cormack. No-reference quality assessment using natural scene statistics: Jpeg2000. *IEEE Transactions on image processing*, 14(11):1918–1927, 2005.

[42] FFTJPEG, `github.com/scienstanford/iqmetrics/blob/master/noRef`, accessed on 04.01.21.

[43] Y. Gong and I. F. Sbalzarini. Image enhancement by gradient distribution specification. In *Asian Conference on Computer Vision*, pages 47–62. Springer, 2014.

[44] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

[45] Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita. No reference image quality assessment for jpeg2000 based on spatial features. *Signal Processing: Image Communication*, 23(4):257–268, 2008.

[46] C. Feichtenhofer, H. Fassold, and P. Schallauer. A perceptual image sharpness metric based on local edge gradient analysis. *IEEE Signal Processing Letters*, 20(4):379–382, 2013.

[47] H. Talebi and P. Milanfar. Nima: Neural image assessment. *IEEE Transactions on Image Processing*, 27(8):3998–4011, 2018.

[48] V. Hosu, B. Goldlucke, and D. Saupe. Effective aesthetics prediction with multi-level spatially pooled features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 9375–9383, 2019.

[49] R. Ferzli and L. J. Karam. A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb). *IEEE transactions on image processing*, 18(4):717–728, 2009.

[50] M. Ponomarenko, N. Gapon, V. Voronin, and K. Egiazarian. Blind estimation of white gaussian noise variance in highly textured images. *Electronic Imaging*, (13):382–1, 2018.

[51] P. Siple and R. M. Springer. Memory and preference for the colors of objects. *Perception & psychophysics*, 34(4):363–370, 1983.

[52] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451, 2006.