

Fig. 1: (a)-(c) Images from the TID2013 [16] ordered by their predicted Lateral Chromatic Aberration level (0, 2, 4 pixels accordingly). (d) a plot of the expected versus the predicted image order, with accuracy correlation coefficient  $\rho = 0.9$ .

3 **Abstract.** In the imaging era, we acquire and assess image quality utilizing Human Visual System (HSV). Along the years,  
 4 various algorithms were developed in order to evaluate the perceptual image quality, by imitating HSV. However, certain image  
 5 distortions cannot be easily measured and quantified since they are ill-defined. In this paper, we assess the distortion level by  
 6 inducing image quality order - on image pair. Our method takes advantage of the fact that simulating image distortion and  
 7 evaluating the relative image quality, is easier than assessing its absolute quality. Accordingly, we define image order based on their  
 8 quality. Later, we seek an optimal mapping of image pair to an ordered pair of scalars, which will maintain the image quality order.  
 9 This task is addressed by constructing a Deep Neural Network which minimizes the violation of image quality order. Subsequently,  
 10 we extend the method to order a set of images by utilizing the level of a chosen distortion. We demonstrate the validity of our  
 11 method on Latent Chromatic Aberration and Moire distortions.

12 **Key words.** Image Quality Assessment, Deep Neural Network, Image Processing, Dimensional Reduction, Resnet.

13  
 14 **1. Introduction**

15 Evaluating the quality of an image is a vital task in the domain of image processing. This is crucial for  
 16 measuring the performance of image processing algorithms, which unintentionally may affect negatively on  
 17 image quality (e.g. denoising algorithm can affect the sharpness of the edges). Image quality is tightly related  
 18 to the questions *What is a good image?* and *Why is image quality assessment so difficult?* [22]. In order to  
 19 avoid repeated quality evaluation by image quality experts, an automatic image quality procedure is required.  
 20 Image quality (IQ) assessment can be addressed either by absolute or relative measure. While in the former  
 21 method, a number representing the IQ of a single image is computed, in the latter approach, given two images,  
 22 we indicate which image looks better. It should be noted that providing a relative measure is easier, as opposed  
 23 to evaluating the quality of a single image and scoring it based on its defects. The reason for this is that there  
 24 is no need to specify what image characteristics influence the image quality scoring (e.g. whether an image is  
 25 sharp enough, or if the resolution is satisfactory).

26 Along the years there were various algorithms suggested cracking the enigma of Human Visual System

---

\*Submitted to the editors 10/4/2019.  
<sup>†</sup>Intel Corporation, Israel ([orx.shimshi@intel.com](mailto:orx.shimshi@intel.com))  
<sup>‡</sup>School of Mathematical Sciences, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel ([alecsan1@post.tau.ac.il](mailto:alecsan1@post.tau.ac.il), <http://www.tau.ac.il/~alecsan1/>)

(HVS), which constantly aid humans in this task [8, 13]. Those studies paved the way towards designing an objective procedure for image quality evaluation [5, 2, 13, 25, 3]. However, measuring the perceptual image quality still remains a challenge. In latest years, the interest in image quality metrics was renewed, with the dissemination of the Deep Neural Networks (DNN). The main idea of those works is that an image quality metric can be defined and evaluated on images. Subsequently, a DNN is constructed to associate between the image and its calculated quality. Recently published papers [6, 7, 10, 12, 23, 24, 20], demonstrate the benefits of using DNN for evaluating the IQ. The key idea of these methods is that they require defining an absolute image quality metric. This obligates the designer to develop a metric and evaluate the distortion as a pre-processing step of DNN. However, the main challenge is that certain image distortions are ill-defined (though they are easy to acquire or simulate). Therefore, there exist distortions with no satisfactory metric solution.

In this paper, we address the question of image quality assessment via relative measure. We introduce relative order-preserving image quality, in order to bypass the challenge of formulating and defining the desired distortion. The method does not require formulating IQ measure. Instead, it learns to rank the image by its quality, with respect to a chosen distortion. In this context, we look on image distortion as the degradation introduced to the ideal image or the deviation from the "perfect" image. Our work was inspired by the work of [21], which learns the semantic hierarchy over words, sentences, and images. In [21] the order is defined on words based on their hypernymys, whereas we aim at maintaining the order of images based on their quality. Our procedure is later extended to rank a set of images with respect to chosen image quality. In what follows, we introduce our relative IQ method (subsection 2.1); and its use when ranking a set of images (subsection 2.2). Then, a description of the creation of datasets (which serves for training and testing) is given (section 3). Later, the validity of our method is demonstrated by learning the order-preserving dimension reduction of two distortions: Chromatic Aberration and Moire (section 4). The paper concludes with a discussion of future directions for methodology enhancements (section 5).

50

## 51 2. Proposed Method

### 52 2.1. Image Quality Assessment of Image Pair

53 The relative measure is formulated as *Is image B more distorted than image A*. We learn to answer the  
 54 question by introducing image quality order, with respect to a selected distortion,  $d$ . Subsequently, we look for  
 55 dimension reduction from the image space to the real numbers, which will maintain the IQ order. This mapping  
 56 will be later utilized to make order in an unseen image pair. Laid in mathematical terms, we define *IQ-order*  
 57 of two given *Region of Interest* (ROI) of an image as:

58 DEFINITION 2.1. *Let  $R_A$  and  $R_B$  be two registered ROIs. Then, ROI-IQ-order is defined as  $R_A <_d R_B$ , meaning*  
 59 *that ROI  $R_A$  is less distorted then ROI  $R_B$ , with respect to  $d$ .*

60 DEFINITION 2.2. *Let  $A$  and  $B$  be two images, as well as the set of their ROIs  $\{R_{A,i}\}, \{R_{B,i}\}$ . If the following*  
 61 *condition stands  $\#\{i \mid R_{A,i} <_d R_{B,i}\} > \#\{i \mid R_{A,i} >_d R_{B,i}\}$  then we define Image-IQ-order as  $A <_d B$ . I.e.*  
 62 *image  $A$  is less distorted then image  $B$  with respect to the tested ROIs.*

63 DEFINITION 2.3. *Let  $A$  and  $B$  be two images  $A <_d B$  along with  $S = \{\langle R_{A,i}, R_{B,i} \rangle \mid R_{A,i} <_d R_{B,i}\}_{i=1..N}$*   
 64 *be their set of ordered registered ROI pairs, and  $f: \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$  a dimensional reduction function. We say*  
 65 *that the mapping  $f$  is IQ-order-preserving if for any ROI pair in  $S$ ,  $f$  is order-preserving. That is if  $\forall i$*   
 66  *$R_{A,i} <_d R_{B,i} \implies f(R_{A,i}) < f(R_{B,i})$ .*

67 Image quality question can now be formulated with respect to the order-preserving mapping.

68 **Problem definition:** Let  $A, B$  be two images  $A <_d B$  and  $S = \{\langle R_{A,i}, R_{B,i} \rangle \mid R_{A,i} <_d R_{B,i}\}_{i=1..N}$  be set of  
 69 ordered registered ROI pair set. We wish to find a mapping  $f$ , which is IQ-order-preserving. In particular, we  
 70 look for a function, which will minimize the image quality order violation

$$71 \quad (2.1) \quad f = \operatorname{argmin}_{f: \mathbb{R}^{n \times m} \rightarrow \mathbb{R}} \frac{1}{N} \sum_{\langle R_{A,i}, R_{B,i} \rangle \in S} E(R_{A,i}, R_{B,i})$$

72 where the loss function for an ordered pair  $\langle R_A, R_B \rangle$  is defined as

## 2.2. Ranking Image Set

The definition of "order" on image pair can be extended to rank set of images, where the images will be ordered based on the amount of their distortion. Once the mapping  $f$  which minimizes (2.1) is found, it can be used for image ranking. Thus, given an order-preserving mapping  $f$ , low/high values of  $f$  correspond to the level of distortion. As a result, one can order a given image set with respect to the values of  $f$ . Specifically, for a set of images  $\{A_j\}_{j=1..J}$ , with the corresponding ROIs  $R_{j,i}$ , the values of the learned order-preserving mapping  $f(R_{j,i})$  are evaluated. Subsequently, the ranking of this set with respect to a given distortion is achieved by ordering the values  $f(R_{j,i})$  for each specific ROI index,  $i$ , and then calculate the median or the ranking across all the image patches,  $j$ .

For example, we would like to rank four images using three patches with given predicted IQ values. Table 1 illustrates all the steps of the processes starting with acquiring the IQ values, ranking their ROI's and later ranking the images. We demonstrate this concept in real case scenario by ranking images with respect to Chromatic Aberration distortion in Figure 1. Images (a)-(c) are ordered according to their rank, and graph in (d) is a plot of the expected versus the predicted ranking.

Table 1: Example of ranking four image, utilizing their three patches

IQ values	Patches Ranks	Images rank
[1 2 3 4; 4 8 9 12; 2 3 5 4]	[1 2 3 4; 1 2 3 4; 1 2 4 3]	[1 2 3 4]

## 2.3. Order Accuracy Evaluation

Measuring the accuracy of image pair predicted order can be done by evaluating the True Positive (TP) percentage. Furthermore, measuring the accuracy of the predicted order of image set is performed using the Spearman's correlation coefficient [14], which is widely utilized to detect trends in data - given reference data. Thus, given two measurements vectors  $x, y$ , and their corresponding ranks,  $r_x$  and  $r_y$  the Spearman's correlation coefficient is calculated as the  $\rho = \frac{cov(r_x, r_y)}{\sigma_{r_x} \sigma_{r_y}}$ . Subsequently, for a patch  $i$ , we apply the Spearman's correlation of the ranks of the predicted values  $f(R_{j,i})$ , and a monotonically increasing sequence with equal length (result in a correlation coefficient  $\rho_i$ ). The accuracy of the predicted image set ranking is the median of  $\rho_i$  across all image patched. In the example above, the correlation coefficients of the patches ranking are [1 1 0.8], with  $median(\rho_i) = 1$ . Therefore, we conclude that there is a monotonicity trend in the data, and the rank prediction is perfect.

## 2.4. Network Architecture

We design the order-preserving mapping as a Deep Neural Network (DNN). The network architecture comprises a Siamese network of a pair of ResNet architecture [9], each performing a dimension reduction. Later, we calculate the loss function which maximizes the distance between mismatches of the dimension reduction via equation 2.1 (which is also called squared negative smoothed hinge loss (SNSHL) [17]. See Table 2 for detailed network architecture details.

Table 2: Architecture of order-preserving network. Building blocks are shown in brackets (and consists of three consequent ReLU’s), with the numbers of blocks stacked. The input to the network is two concatenated color patches of  $32 \times 32$ .

Layer name	Output size	Order-preserving Net104-layer	
slice		slice point 3	
conv1	$16 \times 16 \times 2$	$7 \times 7, 64, \text{stride } 2$	$7 \times 7, 64, \text{stride } 2$
conv2.x	$8 \times 8 \times 2$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3.x	$4 \times 4 \times 2$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
conv4.x	$2 \times 2 \times 2$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
conv5.x	$1 \times 1 \times 2$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
conv1	$1 \times 1 \times 2$	$1 \times 1, 1, \text{stride } 2$	$1 \times 1, 1, \text{stride } 2$
squared negative smoothed hinge loss	$1 \times 1 \times 1$	$1 \times 1, 1, \text{stride } 2$	$1 \times 1, 1, \text{stride } 2$

104

105 **3. Experiments**

106 Our experimental flow consisted of the following steps: (a) acquire a dataset of images for training and  
 107 validation. (b) pair/simulate images of the same scene, each corresponding to different levels of distortion, and  
 108 accompanied with an IQ-order. (c) extract ROIs of the image pair, with the desired distortion. (d) learn the  
 109 order-preserving dimensional reduction function,  $f$ . (e) test the method on various datasets. This general flow  
 110 is illustrated in Figure 2.

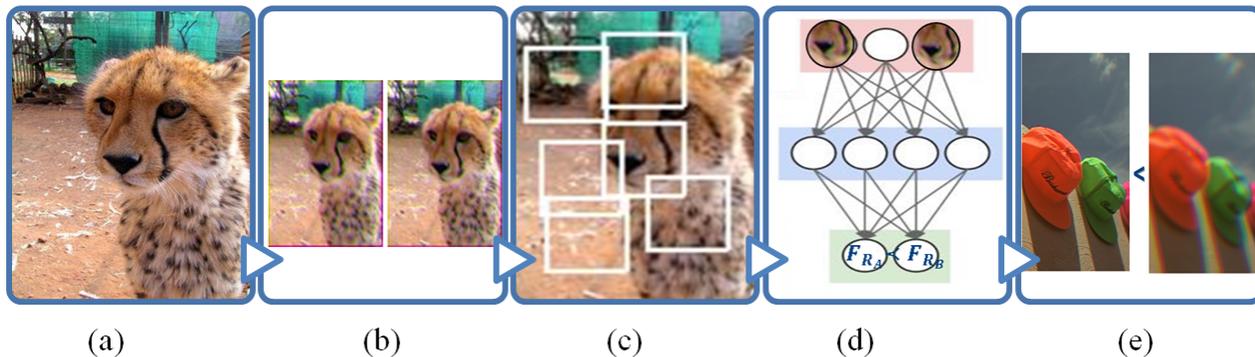


Fig. 2: (a) Acquire a database of images with a chosen distortion (subsection 3.1) (b) create pairs of images each corresponding to different levels of the chosen distortion (subsection 3.2) (c) find areas which correspond to high values of the distortion, and cropping ROIs of size  $32 \times 32$  (subsection 3.3) (d) learn the order-preserving dimension reduction and (e) test it on new datasets (subsection 3.4)

### 111 3.1. Database

112 Since no dataset of distorted images pairs accompanied by the level of distortion was available for training-  
 113 testing purposes, we simulated a distortion on existing images or created synthetic images. Each image in the  
 114 dataset was utilized to create a pair of images, each with a random level of distortion. More details on how each  
 115 distortion was simulated can be found below. After the model was trained, we tested it one real images. We  
 116 acquired and performed mean opinion score on TE42.v2 chart (designed and produced by Image Engineering  
 117 [1]).

### 118 3.2. Distortion Simulation

119 We tested the proposed order-preserving method on Lateral Chromatic Aberration (LCA) and Moire dis-  
 120 tortions (Figure 3). The LCA distortion appears when the colors convergence point is not unique (which stems  
 121 from a failure of a lens to focus). This effect is especially seen as a blur and “rainbow” edge in areas of contrast.  
 122 The LCA dataset was constructed by distorting the ImageNet dataset [4], where the RGB channels of each image  
 123 of the dataset were shifted with a random shift of size  $\sim U(1, 5)$  pixel, in one of the square diagonal directions.  
 124 The Moire distortion (or aliasing) is an effect that causes different signals to become indistinguishable when  
 125 sampled [18]. It occurs upon the existence of repetitive patterns of high spatial frequencies, which are sampled  
 126 with different frequency. Since natural images usually do not depict constant frequency which could serve for  
 127 training Moire distortion, we had to create synthetic image dataset with constant high-frequency patterns. Our  
 128 dataset contained the following simulated repetitive patterns (i.e. resolution bars, resolution net, Siemens-star,  
 129 resolution wedges, concentric rings). In order to simulate the Moire effect we used image resize with bicubic  
 130 interpolation without antialiasing option, with randomly sampled resize factor  $\sim U(1.5, 10)$ .

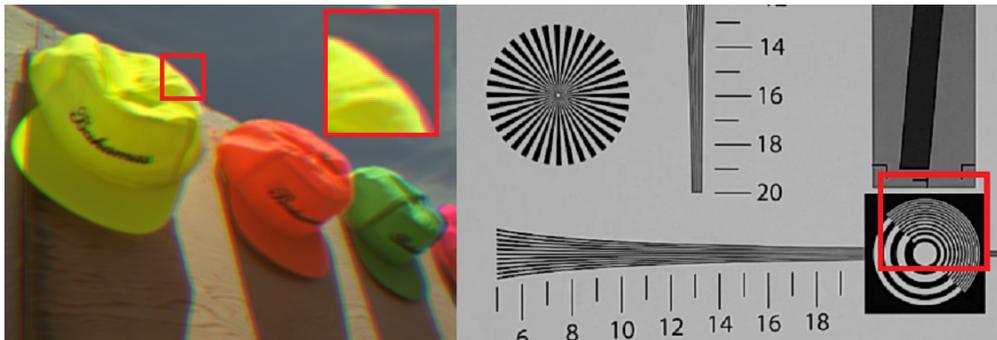


Fig. 3: Left: image with Chromatic Aberration distortion form the TID2013 dataset [16]. Right: resolution chart with Moire effect. Marked in red are the areas with the desired distortion.

### 131 3.3. ROIs Extraction

132 Once IQ-Order-preserving image pair  $\langle A, B \rangle$  is created (by applying two random levels of the chosen  
 133 distortion) the IQ-Order-preserving ROIs of size 32x32 are extracted. The ROIs pairs  $\langle R_{A,i}, R_{B,i} \rangle$ , are chosen  
 134 as the ones corresponding to the maximal values of the error map:  $ErrMap = |A - B|$ . The amount of ROI is  
 135 approximately 0.025% of the image area.

### 136 3.4. IQ Order Dimension Reduction Training

137 In our studies, we formulated the order-preserving function utilizing the Caffe framework [11]. First, a  
 138 dataset for each distortion (LCA and Moire) was constructed, with the entire amount of image patches was  
 139 about 6 Million. Later, we trained the IQ order-preserving dimensional reduction network to minimize the loss  
 140 function  $E(R_A, R_B)$ .

141

142 **4. Experimental Results**

143 Our tests were performed both by predicting the IQ order of image pairs (accuracy was measured by the  
 144 percentile True Positive (TP)), as well as ranking image sets (the accuracy was evaluated with the Spearman  
 145 correlation coefficient ( $\rho$ )). First, we tested our method on synthetic data (prepared in [subsection 3.4](#), or images  
 146 from TID2013 [16]). Then, we performed a wide scale experiment by comparing the predicted LCA and Moire  
 147 ranks to consistency with Human Opinion score on acquired images of TE42.v2 [1]. The results were summarized  
 148 in [Table 3](#).

149 **4.1. Image Quality Assessment of Image Pair**

150 Our trained models were evaluated on test data prepared in [subsection 3.4](#) and the TP percentile was  
 151 calculated, resulting in 97%, 94% for LCA and Moire distortions accordingly. Thus, tests performed on synthetic  
 152 data resulted in satisfactory results. However, the question is whether a model which was trained on synthetic  
 153 data can perform well in real life conditions? For this task, we aimed at performing an experiment using real  
 154 images, with various levels of LCA and Moire distortions, accompanied by image ranks. Since such databases  
 155 were not available, we had to acquire two tests sets of images of TE42 version2 chart [1], one LCA, and other  
 156 containing Moire effect (about 13 images in each). Subsequently, the images were ordered based on human  
 157 opinion score. Later, we simulated 150 different Monte Carlo experiments - by sampling **pairs** of images from  
 158 the test set, which contain the required distortion. In order to simulate “different” rank test, each pair was  
 159 randomly cropped to a size of 150x150 pixels, while depicting the desired distortion. For each cropped pair, we  
 160 predicted the IQ metric and ranked the cropped patches. The experiment resulted in TP percentile of 80% and  
 161 85% for LCA and Moire.

162 **4.2. Image Quality Assessment of Image Set**

163 Ranking image set text was performed on the TID2013 dataset[16], on images created with chromatic  
 164 aberrations (distortion marked as No. 23 in the dataset). The dataset contained images with chromatic  
 165 aberration levels 1-5, unfortunately, it is not specified what levels of LCA were used for TID2013 creation.  
 166 Based on our examination Level 5 of LCA in TID2013 was more than 5 pixels (the value used in our training  
 167 set), therefore we didn’t use it in the ranking procedure. It should be noted that the ROIs were selected by  
 168 calculating the error map as described in C. Fortunately, the images were labeled according to the distortion  
 169 level, and we could check our prediction accuracy against the expected one. The median ranking for the 25  
 170 image sets resulted in accuracy correlation of  $\rho = 1$ . Examples of order-preserving ranking can be seen in  
 171 [Figure 4](#).

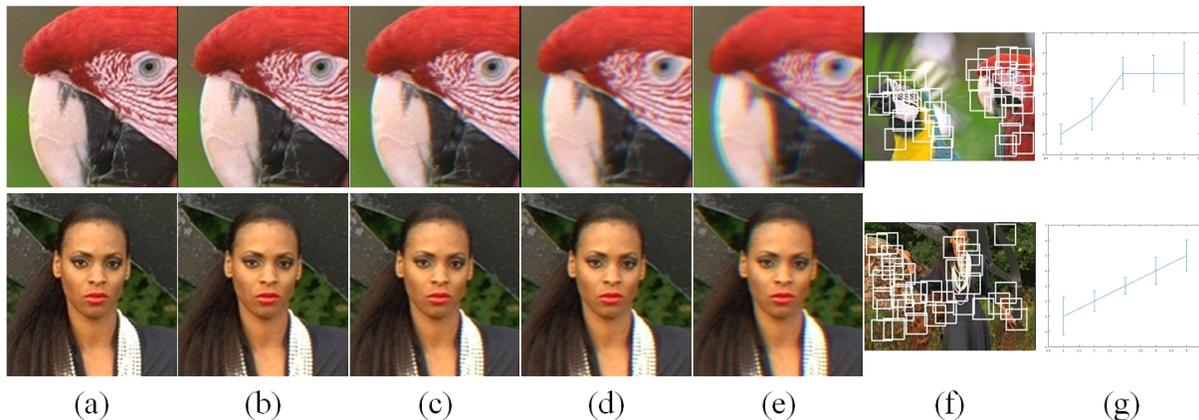


Fig. 4: Zoomed in Images ordered by the rank of the predicted IQ measure from the TID2013 dataset (a)-(e). Inspected ROIs, chosen via error map from [subsection 3.3](#) (f), a graph with the expect vs prediction rank values (g). The accuracy correlation of those sets is  $\rho = 0.8, 1$  (from top to bottom row).

172 Subsequently, we utilized the ordered images set acquired for the image pairs test of TE42.v2 chart (dis-  
 173 cussed above). We simulated 150 image sets by we sampling quadruplets of images from the original dataset  
 174 and cropped them randomly to a size of 150x150 pixels while depicting the desired distortion. For each cropped  
 175 quadruplet, we predicted the IQ metric and ranked the cropped patches. The experiment resulted in a median  
 176 accuracy of  $\rho = 0.7$  and  $\rho = 0.8$  for LAB and Moire (see examples of prediction in Figure 5).

Table 3: Summary of order-preserving image quality assessment tests for LAC and Moire distortions added artificially or acquired in real images

Test	LCA	Moire
Synthetic data, test set (%TP)	97%	94%
TE42 chart (%TP)	80%	85%
TID2013 ( $\rho$ )	1	N/A
TE42 chart ( $\rho$ )	0.7	0.8

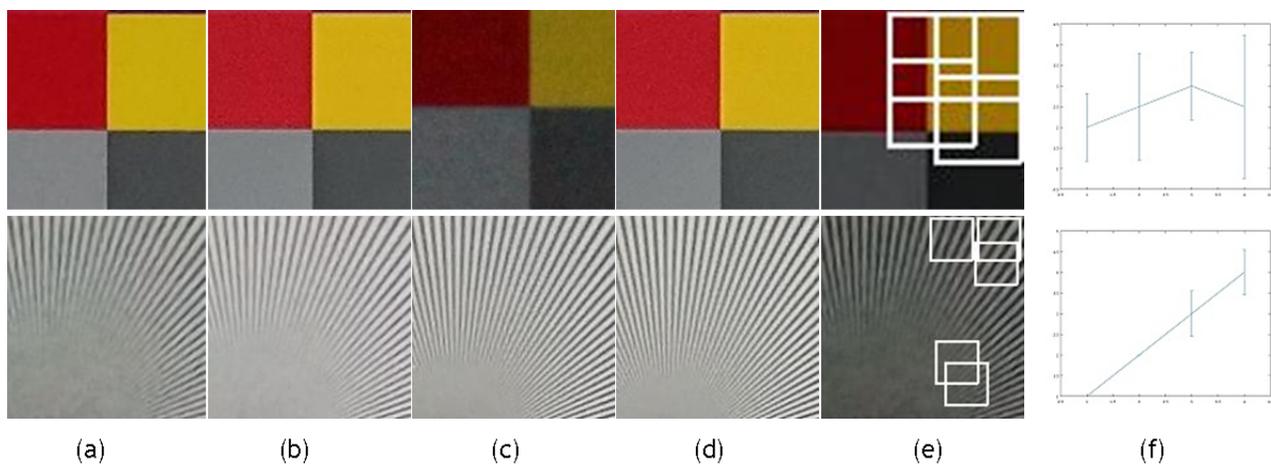


Fig. 5: Example of quadruplets, ordered by the rank (a-d), ROIs used for ranking (e), the median predicted rank (f). In the first row, LCA distortion was measured - resulted with accuracy of 0.9. In the second row Moire was assessed, and resulted with accuracy of 1.

177

## 178 5. Summary and Future Directions

179 Assessing image quality is a key problem when evaluating image processing algorithms. The challenge can  
 180 be addressed either by absolute or relative measurements. As the absolute metric is sometimes ill-defined, and  
 181 can be harder to implement, a relative metric can solve the problem. In this paper we suggest a relative method,  
 182 which look for an optimal mapping that maintain the order of pair of images. Namely, given a pair of images the  
 183 mapping returns pair of scalars, that are ordered based on the IQ order. Subsequently, we propose extending  
 184 the mechanism for ranking set of almost registered images. The ranking is performed by ordering the images  
 185 by the found function values.

186 We demonstrated the validity of our method by constructing a Deep Neural Network and testing it on two  
 187 distortions: Chromatic Aberration and the Moire. The test's accuracy on synthetic data as well as real data,  
 188 showed satisfactory results. This demonstrated that even though that the training was performed on synthetic  
 189 data, the results achieved on real data were good. Our method paves the way towards measuring other image  
 190 distortions. Initial experiments showed the potential of utilizing the infrastructure for edge roughness and  
 191 sharpness assessment.

192 In addition, in image processing there are many challenges which require relative image quality assessment.  
 193 Among them are algorithms comparison and calibration. In case there are several algorithms that need to be  
 194 surveyed (and no reliable Ground Truth is available [19]), using relative image quality can evaluate the quality  
 195 of the algorithms. On the other hand, a single algorithm can produce several images with different qualities,  
 196 just by changing the input parameters. Finding the optimal parameters which will produce an image with good  
 197 quality, is often a challenge [15]. However, one of the main problems in algorithm calibration is *what measure*  
 198 *should be used for the calibration task*. Utilizing a tailored relative measure can solve the issue.

## 199 6. Acknowledgments

201 We thank Shay Maymon and Dmitry Paus for their insightful and valuable comments and suggestions. We  
 202 would also like to thank Dmitry Grilikhes and Evgeny Bespechansky for assisting with obtaining the suitable  
 203 charts.

## 204 REFERENCES

- 205 [1] TE42. <https://www.image-engineering.de/products/charts/all/425-te42>. Accessed: 2018-03-25
- 206 [2] Barney Smith, E.H., Maggard, E., Line, S., Shaw, M.: Quantifying print quality for practice. In: NIP & Digital Fabrication  
 207 Conference, vol. 2015, pp. 157–162. Society for Imaging Science and Technology (2015)
- 208 [3] van den Branden Lambrecht, C.J.: A working spatio-temporal model of the human visual system for image restoration and  
 209 quality assessment applications. In: Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings.,  
 210 1996 IEEE International Conference on, vol. 4, pp. 2291–2294. IEEE (1996)
- 211 [4] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Computer  
 212 Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 248–255. IEEE (2009)
- 213 [5] Faigenbaum, S., Shaus, A., Sober, B., Turkel, E., Piasetzky, E.: Evaluating glyph binarizations based on their properties. In:  
 214 Proceedings of the 2013 ACM symposium on Document engineering, pp. 127–130. ACM (2013)
- 215 [6] Gao, F., Tao, D., Gao, X., Li, X.: Learning to rank for blind image quality assessment. IEEE transactions on neural networks  
 216 and learning systems **26**(10), 2275–2290 (2015)
- 217 [7] Gao, F., Wang, Y., Li, P., Tan, M., Yu, J., Zhu, Y.: Deepsim: Deep similarity for image quality assessment. Neurocomputing  
 218 **257**, 104–114 (2017)
- 219 [8] Granrath, D.J.: The role of human visual models in image processing. Proceedings of the IEEE **69**(5), 552–561 (1981)
- 220 [9] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on  
 221 computer vision and pattern recognition, pp. 770–778 (2016)
- 222 [10] Hou, W., Gao, X., Tao, D., Li, X.: Blind image quality assessment via deep learning. IEEE transactions on neural networks  
 223 and learning systems **26**(6), 1275–1286 (2015)
- 224 [11] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional  
 225 architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia, pp.  
 226 675–678. ACM (2014)
- 227 [12] Jiang, Q., Shao, F., Jiang, G., Yu, M., Peng, Z.: Supervised dictionary learning for blind image quality assessment using  
 228 quality-constraint sparse coding. Journal of Visual Communication and Image Representation **33**, 123–133 (2015)
- 229 [13] Kleinmann, J., Wueller, D.: Investigation of two methods to quantify noise in digital images based on the perception of the  
 230 human eye. In: Image Quality and System Performance IV, vol. 6494, p. 64940N. International Society for Optics and  
 231 Photonics (2007)
- 232 [14] Lehmann, E., D’abrera, H.: Nonparametrics: Statistical methods based on ranks, holden-day inc. San Francisco pp. 300–315  
 233 (1975)
- 234 [15] Nishimura, J., Gerasimow, T., Sushma, R., Sutic, A., Wu, C.T., Michael, G.: Automatic isp image quality tuning using  
 235 nonlinear optimization. In: 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 2471–2475. IEEE  
 236 (2018)
- 237 [16] Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti,  
 238 F., et al.: Image database tid2013: Peculiarities, results and perspectives. Signal Processing: Image Communication **30**,  
 239 57–77 (2015)
- 240 [17] Rennie, J.D., Srebro, N.: Loss functions for preference levels: Regression with discrete ordered labels. In: Proceedings of the  
 241 IJCAI multidisciplinary workshop on advances in preference handling, pp. 180–186. Kluwer Norwell, MA (2005)
- 242 [18] Sidorov, D.N., Kokaram, A.C.: Suppression of moiré patterns via spectral analysis. In: Visual Communications and Image  
 243 Processing 2002, vol. 4671, pp. 895–907. International Society for Optics and Photonics (2002)
- 244 [19] Smith, E.H.B.: An analysis of binarization ground truthing. In: Proceedings of the 9th IAPR International Workshop on  
 245 Document Analysis Systems, pp. 27–34. ACM (2010)
- 246 [20] Talebi, H., Milanfar, P.: Nima: Neural image assessment. IEEE Transactions on Image Processing **27**(8), 3998–4011 (2018)
- 247 [21] Vendrov, I., Kiros, R., Fidler, S., Urtasun, R.: Order-embeddings of images and language. arXiv preprint arXiv:1511.06361  
 248 (2015)
- 249 [22] Wang, Z., Bovik, A.C., Lu, L.: Why is image quality assessment so difficult? In: Acoustics, Speech, and Signal Processing  
 250 (ICASSP), 2002 IEEE International Conference on, vol. 4, pp. IV–3313. IEEE (2002)

- 251 [23] Xu, L., Li, J., Lin, W., Zhang, Y., Zhang, Y., Yan, Y.: Pairwise comparison and rank learning for image quality assessment.  
252 *Displays* **44**, 21–26 (2016)
- 253 [24] Yu, S., Wu, S., Wang, L., Jiang, F., Xie, Y., Li, L.: A shallow convolutional neural network for blind image sharpness  
254 assessment. *PloS one* **12**(5), e0176632 (2017)
- 255 [25] Zhang, Y., Chandler, D.M.: Opinion-unaware blind quality assessment of multiply and singly distorted images via distortion  
256 parameter estimation. *IEEE Transactions on Image Processing* **27**(11), 5433–5448 (2018)