

# Sampled Efficient Full-Reference Image Quality Assessment Models

Christos G. Bampis, Todd R. Goodall, Alan C. Bovik  
University of Texas at Austin, Austin, Texas 78712-0240, U.S.A.  
{bampis, tgoodall}@utexas.edu, bovik@ece.utexas.edu

**Abstract**—Existing full-reference image quality assessment models first compute a full image quality-predictive feature map followed by a spatial pooling scheme, thereby producing a single quality score. Here we study spatial sampling strategies that can be used to more efficiently compute reliable picture quality scores. We develop a random sampling scheme on single scale full-reference image quality assessment models. Based on a thorough analysis of how this random sampling strategy affects the correlations of the resulting pooled scores against human subjective quality judgements, a highly efficient grid sampling scheme is proposed which replaces the ubiquitous convolution operations with local block-based multiplications. Experiments on four different databases show that this block-based sampling strategy can yield results similar to methods that use a complete image feature map, even when the number of feature samples is reduced by 90%.

## I. INTRODUCTION

We study the effects of sub-sampling feature maps used in full-reference (FR) picture quality analyzers as a way of achieving efficiency without sacrificing performance. The work presented herein can be related to reduced-reference (RR) models, in the sense that the goal is to show, even for FR models, that the entire reference image is not necessarily required to produce accurate picture quality predictions. For example, one leading RR method, the RRED indices [1], uses subsampled wavelet subband information from both pristine and distorted images to predict image quality, thereby allowing a controlled, variable amount of information to be used. The RRED model ranges from a FR version, through graded levels of compared information, down to a single number derived from the reference. However, the RRED indices are computed using an overcomplete steerable wavelet transform, which requires all pixels from the reference and distorted images. Likewise, an approach based on block-based pooling strategies [2] demonstrates a similar limitation.

The approach proposed herein allows the final pooling of spatial quality features to be intertwined with the model computation. For example, most IQA models compute dense spatial feature maps which are generally redundant. Knowledge of such redundancies can be used to redesign the computation procedure. After appropriate redesign, most full-reference IQA models can yield statistically indistinguishable results.

Requiring the use of only a subset of the input data can have a large impact on future IQA algorithm design. The number of computational operations required for quality models is often a function of the image feature map size. By effectively

modifying the feature map size, existing IQA models can be modified for real-time applications, without a noticeable loss in predictive performance. Also, by such efficient sampling strategies, any FR method becomes a RR method, since only a portion of the source reference image is used for quality assessment. The rest of this paper is organized as follows. Section II describes a random sampling strategy for single scale FR methods. Section III studies the impact of sampling on quality prediction efficacy. Section IV describes our block-based sampling strategy, and Section V describes experimental results. Lastly, Section VI concludes the paper and discusses future work.

## II. RANDOM SAMPLING PROCESS

We first describe a simple yet instructive approach to FR random sampling. Assuming that  $I$  and  $J$  are pristine and distorted versions of the same  $m \times n$  image, a typical FR IQA model applies a  $w \times w$  filter to yield two filtered and/or non-linearly processed images  $I_f$  and  $J_f$ . To retain only valid portions (unaffected by image boundary effects) of the filtered images, the inner  $\tilde{m} \times \tilde{n}$  region is obtained, where  $\tilde{m} = m - w$  and  $\tilde{n} = n - w$ . Depending on the IQA model, this process may occur multiple times, possibly at different scales and/or orientations, either in parallel or in series, to obtain a number of filtered images.

In the case of the Structural Similarity Index Metric (SSIM) [3], several filtered and processed images are produced, including a local mean and variance map, for both  $I$  and  $J$ . These maps are combined to produce a dense SSIM map. Given the correlations produced by both filtering and by pre-existing relationships between pixel neighbors, the resulting dense SSIM map is highly redundant.

To test for redundancy in the dense SSIM map, a subset of it can be extracted using random sampling. Define  $N(p) = \left\lfloor \frac{p\tilde{m}\tilde{n}}{100} \right\rfloor$  as the number of pixels to select, where  $p \in (0, 100]$  is the percentage of the slightly cropped image size and  $\lfloor \cdot \rfloor$  is the floor function. Clearly,  $p = 100$  indicates that every available feature sample is used. Note that samples can be selected “with” and “without” replacement. The former allows for a particular pixel location to be selected more than once (every index is picked independently with probability  $\frac{1}{\tilde{m}\tilde{n}}$ ), while the latter results in a vector of unique samples. Using either random sampling scheme, a vector of computed points is gathered from the SSIM map. If these samples were adequate

to represent the overall picture quality, then SSIM would only need to be computed at the selected samples. Since the mean of the original SSIM map is used as the quality score, the mean of these sampled locations can be thought of as an estimate of the original “Mean SSIM”.

To formalize, let  $\mathcal{B}$  be the set of points  $\{\mathbf{x}_i\} = \{(x_i, y_i)\}$   $\forall i = 1 \dots |\mathcal{B}|$  in a given image. Given the SSIM index map  $M$ , the SSIM value  $s$ , is given by  $s = \sum_{i \in \mathcal{B}} M(\mathbf{x}_i)$ . Next consider  $T$  subsets of  $\mathcal{B}$ , defined as  $\mathcal{A}_j \subset \mathcal{B}$ , where  $|\mathcal{A}_j| = |\mathcal{B}| \frac{p}{100} = N(p)$  and  $j \in 1, 2, \dots, T$ . The estimate  $\hat{s}_j$  corresponding to the  $j$ th trial is  $\hat{s}_j = \sum_{i \in \mathcal{A}_j} M(\mathbf{x}_i)$ . Using

repeated independent trials, the distribution of  $\hat{s}_j$  across trials can be compared to  $s$ . Clearly,  $\hat{s}_j$  will generally differ from  $s$  to some degree for  $p < 100$ , and will be equal to  $s$  when  $p = 100$ . Therefore, we study the sampling distribution formed by  $\hat{s}_j$  when  $p < 100$ . From this point forward, the SSIM index map produced at all picture coordinates  $(x, y)$  minus any border cropping (the traditional SSIM index map) will be referred to as the parent distribution. Clearly, the parent distribution may be highly non-Gaussian. However, a histogram plotted using  $T$  realizations of  $\hat{s}_j$  yields an empirical probability density function (epdf) that appears Gaussian, as depicted in Fig. 1. Ignoring the obvious underlying image and model-related dependencies, the distribution of  $\hat{s}_j$  likely approaches a Gaussian shape as a consequence of the Central Limit Theorem; although conditions (such as independence) of sampled values from a SSIM map have not been established, it is likely nearly true when the between-sample distance is only moderately large. We have noted that the spread of the epdf decreases with larger  $p$ .

Fig. 1 depicts the sampling distributions both “with” and “without” replacement. In both cases, the epdfs can be well approximated by a Gaussian distribution. By examining these distributions, it can be seen that the “with” replacement scheme increases the spread of the sampling distribution, since for a given  $p$  it uses a smaller number of unique indices in the SSIM index map than the “without” replacement scheme. Hereafter, a “without” replacement scheme is used, because in practice reducing the number of samples is preferred. Next, by letting  $T = 10^4$  and  $p \in [10, 20, \dots, 100]$  for any given image, the sampling distributions can be studied as a function of  $p$ . As depicted in Fig. 1, decreases in  $p$  lead to wider sampling distributions. By contrast, if  $p = 100$ , the distribution degenerates into the mean of the parent distribution for each iteration.

There are two competing forces when  $p$  increases. On the one hand, the probability that a sample is picked close to some other sample already selected is increasing. Since samples that are closer to each other tend to have correlated SSIM values, local dependencies are introduced for these few neighbours of every sample. On the other hand, the number of samples that are independent for this sample are greatly increased, thereby supporting the CLT argument. This observation can be verified by examining the shape parameter of the generated sampling distributions. We adopt the moment matching method also

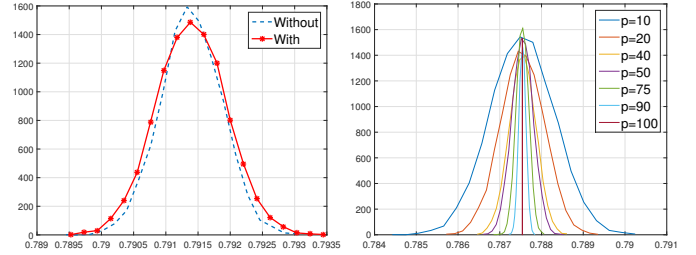


Fig. 1. Left: sampling distribution “with” and “without” replacement using  $p = 25$  and  $T = 10^4$  trials. Right: sampling distribution across different  $p$  values after  $T = 10^4$  trials. The vertical line for  $p = 100$  does not reach  $T = 10^4$  due to arithmetic precision errors. This is only a plotting artifact.

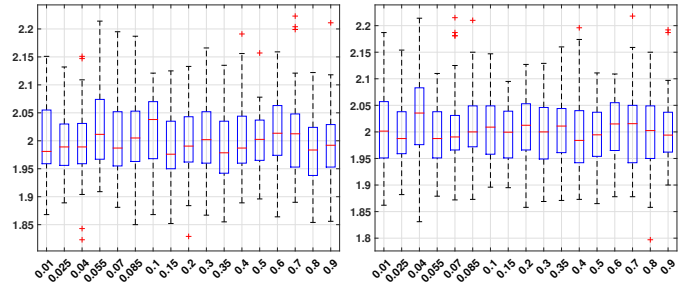


Fig. 2. Shape parameter for SSIM (left) and FSIM [5] (right) across different  $p$  values on 50 randomly selected image pairs on the LIVE [6] dataset. The sampling distribution is close to a gaussian since  $\alpha$  is close to 2. The number of trials was set to 5000.

used in [4] i.e. we examine whether the  $\alpha$  (shape) parameter is close to 2. The parameters  $\beta_l$  and  $\beta_r$  can also be used to check the symmetric nature of the sampling distribution. As shown in Figs. 2-3, the sampling distribution for SSIM and FSIM are very close to a symmetric gaussian distribution across different values of  $p$ .

In the simple random sampling strategy analyzed here, a full feature map is produced first, and then random indices from the SSIM map are selected to compute the final SSIM index. As an alternative, one could apply a block based computation around those indices, then compute the SSIM index. These two approaches yield the same result. However, we use the random sampling strategy only to analyze the properties of sampling FR IQA models. Later, we exploit these properties to propose more practical sampling schemes.

### III. IMPACT OF SAMPLING ON CORRELATIONS WITH HUMAN SCORES FOR FR IQA METHODS

Next, we analyze the relative quality prediction performance of original and sampled FR models. Let  $i \in [1, 2, \dots, K]$  index a collection of pristine/distorted image pairs  $\{I_i, J_i\}$ , where  $K$  is the total number of available image pairs. Suppose that the SSIM quality score produced from a full SSIM map with no sub-sampling is  $s_i$ , and let  $v_i \sim \mathcal{N}(0, \sigma_i^2)$  be a Gaussian distribution which will act as a noise term. Based on the empirical observations of the previous sections, the SSIM quality score of image pair  $i$  after sampling with percentage  $p$  can be modeled as  $z_i = s_i + v_i$ , where  $\sigma_i$  decreases with

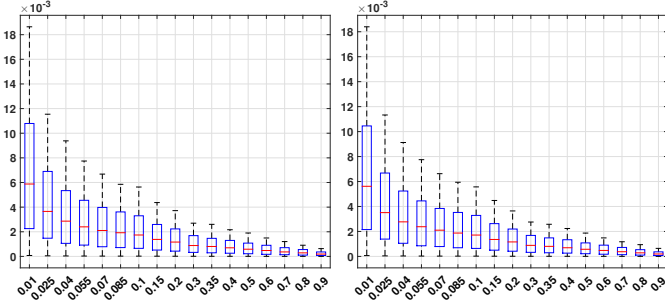


Fig. 3.  $\beta$  values for FSIM across different  $p$  values on 50 randomly selected image pairs on the LIVE dataset. Left:  $\beta_l$ ; Right:  $\beta_r$ . Both  $\beta$  values decrease when  $p$  increases. The sampling distribution for FSIM tends to be symmetric since, given  $p$ ,  $\beta_l$  is close to  $\beta_r$ . The number of trials was set to 5000.

increasing  $p$ . This model captures the observation that smaller  $p$  values increase the spread of the sampling distribution.

Using this model to describe the sampling effect, the samples  $s_i$  may be viewed as realizations of a random variable  $S$ , and the samples  $z_i$  of another random variable  $Z$ . Since  $\sigma_i$  will vary across images, consider the worst case, i.e. let  $\sigma_w = \max\{\sigma_1, \sigma_2, \dots, \sigma_K\} = \max_i\{\sigma_i\}$ , then let all  $\sigma_i$  be modeled as  $\sigma_w$ . In other words, let  $V \sim \mathcal{N}(0, \sigma_w^2)$  denote the worst case when choosing a particular  $p$  in a given sampling strategy. Therefore, the model can be simplified to  $Z = S + V$ .

The question, then, is how well the prediction from SSIM samples correlates with human subjective judgements of quality. The Pearson correlation coefficients (PCC) between the recorded human scores, denoted by  $Y$ , and the sampled ( $Z$ ) and non-sampled ( $S$ ) objective scores are given by:

$$P_{S,Y} = \frac{E[SY] - E[S]E[Y]}{\sqrt{E[S^2] - E[S]^2} \sqrt{E[Y^2] - E[Y]^2}} \quad (1)$$

$$\begin{aligned} P_{Z,Y} &= \frac{E[ZY] - E[Z]E[Y]}{\sqrt{E[Z^2] - E[Z]^2} \sqrt{E[Y^2] - E[Y]^2}} \\ &= \frac{E[XY] - E[X]E[Y]}{\sqrt{E[X^2] - E[X]^2 + \sigma_w^2} \sqrt{E[Y^2] - E[Y]^2}} \end{aligned} \quad (2)$$

From (1) and (2),  $P_{S,Y}$  refers to correlation without subsampling, and  $P_{Z,Y}$  refers to correlation after sampling. Naturally, it is to be expected that  $P_{Z,Y} \leq P_{S,Y}$  since  $Z$  is a noisy estimate of  $S$ . As shown experimentally, this may correspond to a very small drop in practice.

#### IV. BLOCK-BASED SSIM OPTIMIZATION

Naturally, to achieve efficient IQA, samples should be chosen such that they do not belong to the same block of size  $w$ . An obvious choice is to grid sample with a stride of at least  $w$  pixels in both the  $x$  and  $y$  direction. To do so,  $I$  and  $J$  can be directly multiplied by a grid of filters, then appropriately added and subtracted. This grid of filters is created by repeating the original  $w \times w$  filter  $m' = \lfloor \frac{m}{w} \rfloor w$  times in the  $x$  direction and  $n' = \lfloor \frac{n}{w} \rfloor w$  times in the  $y$  direction, thereby creating an image  $\hat{W}$  of size  $m' \times n'$  (see Fig. 4).

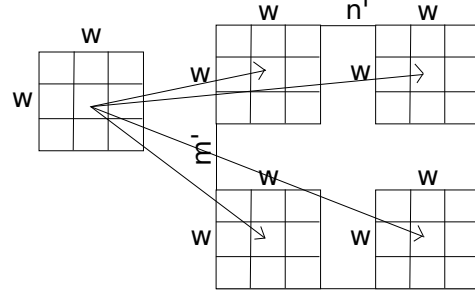


Fig. 4. Creating  $\hat{W}$  for BB-SSIM. Here,  $w = 3$  for simplicity.

By appropriately multiplying the original images with  $\hat{W}$ , a block-based efficient version of SSIM (BB-SSIM) can be derived. First,  $I$  and  $J$  are cropped and then multiplied element-wise by  $\hat{W}$ . Then, a block-based mean computation on  $I \odot \hat{W}$  and  $J \odot \hat{W}$  yields  $\mu_1$  and  $\mu_2$ . Further, compute  $\mu_{12} = \mu_1 \odot \mu_2$ ,  $\mu_1^2$ ,  $\mu_2^2$ ,  $I^2$  and  $J^2$  by element-wise multiplications e.g.  $J^2 = J \odot J$ . Multiplying  $I \odot J$ ,  $I^2$  and  $J^2$  by  $\hat{W}$  and subtracting  $\mu_{12}$ ,  $\mu_1^2$  and  $\mu_2^2$  yields  $\sigma_1^2$ ,  $\sigma_2^2$  and  $\sigma_{12}$ . Finally, using the SSIM formula, one can combine the block-sampled mean and  $\sigma$  images (for both images and their product), then take an average over the resulting block-based feature map to produce a single ‘‘Mean SSIM’’ score. Clearly, the efficiency of this method depends on whether the block-based summing step is optimized in hardware.

It can be shown that the original SSIM method needs  $3mn + \tilde{m}\tilde{n}(10w^2 + 17)$  Floating Point Operations (FLOPs). By contrast, BB-SSIM needs only  $m'n'(\frac{17}{w^2} + 13)$  FLOPs. Assume that  $m, n$  are large enough, i.e.  $\tilde{m} \approx m$ ,  $\tilde{n} \approx n$ ,  $m' \approx m$  and  $n' \approx n$ . Then, the original SSIM needs  $mn(10w^2 + 20)$  FLOPs whereas BB-SSIM needs  $mn(\frac{17}{w^2} + 13)$  FLOPs. Notice the gain term scales inversely with  $w^2$ . By setting  $w = 11$ , the speedup in terms of floating point operations (FLOPs) is approximately 93.6. In the experimental section, this computational gain is practically quantified.

#### V. EXPERIMENTAL SECTION

We conducted standard deviation measurements on the correlation scores for SSIM and PSNR and compared them for the same  $p$  in Fig. 5. Perception-based IQA models should be more compressible when compared to signal fidelity measures like PSNR, in part because the former uses spatial filters to compute neighborhood responses and also since the perceptual results are more consistent. Comparisons are made on the LIVE dataset, which consists of 779 artificially distorted images, each distorted with varying levels of jp2k, jpeg, white noise, fast fading or gaussian blur distortions. PSNR and SSIM predictions were computed using the sampling technique 100 times with  $p = 40$ . In Fig. 5, the sampling error is compared between SSIM and PSNR for all five distortions. Clearly, PSNR exhibits higher sampling errors across all five distortions. Standard deviation is observed to be influenced by distortion type. As an example, for the gaussian blur distortion the standard deviation is larger for both PSNR and SSIM.

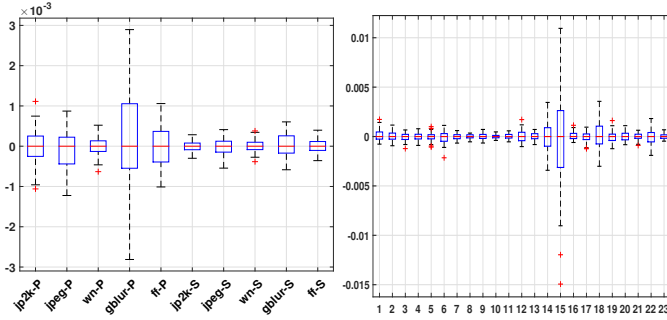


Fig. 5. Left: PSNR and SSIM correlation scores for  $p = 40$  across the five distortions in the LIVE database for 100 iterations. From left to right: distortions 1-5: PSNR, distortions 6-10: SSIM. Right: SSIM correlation scores for  $p = 40$  across the 23 distortions in the TID13 database for 100 iterations. The y-axis has standard deviation units.

To further understand the per-distortion behavior of the random sampling strategy, the same SSIM sampling strategy was followed on the 23 distortions described in the TID13 dataset. As shown in Fig. 5, distortions 14, 15 and 18 show higher sampling error than the other distortions. These indices correspond to “Non eccentricity pattern noise”, “Local block-wise distortions of different intensity” and “Change of color saturation”. The first two of these distortion categories are local distortions, leading to the conclusion that the random sampling strategy performs statistically worse on local distortions. This likely follows from the fact that few samples might be taken in the locally distorted region.

We now study the following three methods: the original FR method ( $p = 100$ ), the random sampling scheme using  $p = 1$  and a grid sampling scheme where we first precompute the entire feature map and then perform a 11-grid sampling. The parameter  $p = 0.01$  approximates the number of points used in the 11-grid sampling, where every non-overlapping block contains 121 pixels, of which only one value is kept.

In the experiments to follow, four popular image quality datasets are considered, namely CSIQ [7], TID2008 [8], LIVE [6] and TID2013 [9]. The FR IQA methods that are evaluated are PSNR, SSIM [3], IW-SSIM [10] and FSIM [5]. Note that for FSIM, only its grayscale version is used for comparability with the other methods. Table I lists the correlation scores for the four different FR methods. For IW-SSIM, only 10 iterations are performed due to computational reasons, but the results are still consistent. For random sampling, the average value obtained over all iterations is listed. It can be seen that, in many cases, results obtained using the 11-grid method are close to both the averaged random sampling quality predictions and the unsampled scores. However, a clear improvement is obtained on a more complex model such as IW-SSIM. The key advantage of the block-based sampling technique is that it always produces the same result, unlike the random sampling strategy. The advantage of using a deterministic sampling strategy is that the algorithm can be computationally optimized based on the keypoints. For the case of SSIM, when the block size is 11, the correlation score is 0.9111,

which is very close to the unsampled value of 0.9104 on the LIVE dataset. Similarly, the same block size used on the TID13 dataset achieves a correlation score of 0.6308, while the non-sampled correlation is 0.6313. Clearly, the grid sampling strategy produces nearly the same result.

As shown in Figs. 6-7, which depict boxplots per database for the random sampling strategy applied on SSIM, FSIM and IW-SSIM, using a small value of  $p$  increases the sampling uncertainty. In these boxplots, the horizontal axis shows the percentage  $\frac{p}{100}$  and the vertical axis shows the distribution of the correlation scores produced by different sub-sampling trials. Also, the average values of the SROCC correlation scores over all  $p$  are close to those obtained by using the entire feature map. Furthermore, using small values of  $p$  when sub-sampling usually produces an underestimate of the correlation score verifying the theoretical analysis laid out in Section III. As  $p$  increases, the CLT assumption likely still holds but the standard deviation of the sampling distribution reduces thus it is more difficult to determine whether the correlation score produced by larger  $p$  values is an underestimate of the actual value (denoted by green). Across all FR methods and datasets, the reduced information resulting from the random sampling strategy yields little loss in predictive power.

By implementing the block-based sampling method, clear performance gains are possible in terms of multiplies and adds, at least for SSIM. To test these gains, only the core operations in an existing Matlab SSIM implementation are modified. The execution times are provided in the columns of Table II. Note that BB-SSIM is approximately 1.5 to 2 times faster than SSIM. Also, note that the correlation scores for the 11-grid method, which uses a block size of  $11 \times 11$  reported in Table I are very close to the ones produced in Table II. The difference between the two lies in the following. Let  $B_1$  denote the first block from the full SSIM index feature map. The 11-grid method picks the center pixel in this block, while the BB-SSIM method will generate the top left pixel in this block instead.

## VI. CONCLUSION

We explored different sampling strategies to efficiently estimate several single scale FR IQA metrics. By studying random sampling strategies, it has been determined that systematically reducing the information used to compute a final IQA score generally has little effect on the correlation with human subjective scores. The underlying redundancies produced by many IQA metrics such as SSIM, along with image content dependencies and spatial uniformities of distributions of distortions, enables the grid sampling scheme, which can be implemented using block-based operations. These operations can be used as a building block for many other IQA (or VQA) models, since the sampling procedure is easily extensible. The experimental results obtained on four popular datasets demonstrate that the block-based sampling strategy can achieve results that are very close to traditional FR methods at a fraction of the cost. Although the sampling strategies proposed herein are appropriate for single scale FR methods, they may be extended

TABLE I

EXPERIMENTAL RESULTS OF RANDOM SAMPLING FOR FR IQA METRICS ON 4 DATASETS, ROWS 1-3: 100 ITERATIONS, ROW 4: 10 ITERATIONS.

Method	LIVE			TID13			TID08			CSIQ		
	$p = 100$	$p = 1$	11-grid	$p = 100$	$p = 1$	11-grid	$p = 100$	$p = 1$	11-grid	$p = 100$	$p = 1$	11-grid
PSNR	0.8756	0.8714	0.8742	0.6468	0.6215	0.6469	0.5559	0.5226	0.5560	0.8058	0.8066	0.8049
SSIM	0.9104	0.9103	0.9111	0.6313	0.6311	0.6308	0.6332	0.6326	0.6324	0.8369	0.8367	0.8369
FSIM	0.9634	0.9621	0.9607	0.8028	0.7970	0.8003	0.8825	0.8749	0.8764	0.9242	0.9230	0.9254
IW-SSIM	0.9568	0.9200	0.9572	0.7806	0.7467	0.7769	0.8566	0.7904	0.8534	0.9206	0.8611	0.9216

TABLE II

SROCC CORRELATION SCORES AND COMPUTATIONAL TIMES (IN SECONDS) ON DIFFERENT DATASETS. TIME CORRESPONDS TO THE ENTIRE COMPUTE TIME NEEDED ON EACH DATABASE.

Database	SSIM	BB-SSIM	t	$t_{BB-SSIM}$
LIVE	0.9104	0.9113	18.5313	11.2545
TID13	0.6313	0.6311	42.1429	23.1630
TID08	0.6332	0.6328	23.3832	12.7633
CSIQ	0.8369	0.8363	16.0295	10.0055

to multiscale FR, NR IQA methods [12], [13], [4] and to video quality assessment (VQA) [14], [15], [16], which are interesting future research directions.

## REFERENCES

- [1] R. Soundararajan and A. C. Bovik, "R-RED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517–526, 2012.
- [2] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal processing: Image communication*, vol. 19, no. 2, pp. 121–132, 2004.
- [3] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.
- [4] A. Mittal, G. S. Muralidhar, J. Ghosh, and A. C. Bovik, "Blind image quality assessment without human training using latent quality factors," *Signal Processing Letters, IEEE*, vol. 19, no. 2, pp. 75–78, 2012.
- [5] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [6] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, 2006.
- [7] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electronic Imaging*, vol. 19, no. 1, pp. 011 006–011 006, 2010.
- [8] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.
- [9] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Visual Information Processing (EUVIP), 2013 4th European Workshop on*. IEEE, 2013, pp. 106–111.
- [10] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.
- [11] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, Nov 2003, pp. 1398–1402.
- [12] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *Signal Processing Letters, IEEE*, vol. 20, no. 3, pp. 209–212, 2013.
- [13] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *Image Processing, IEEE Transactions on*, vol. 21, no. 12, pp. 4695–4708, 2012.

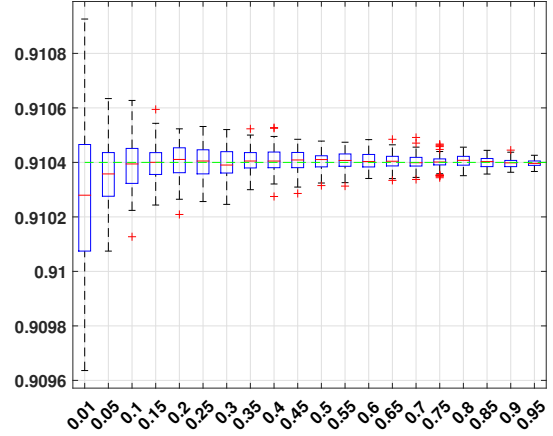


Fig. 6. SSIM correlation scores against percentage  $\frac{p}{100}$  on LIVE over 100 iterations. The actual correlation score is denoted by green. This figure can be better seen in color.

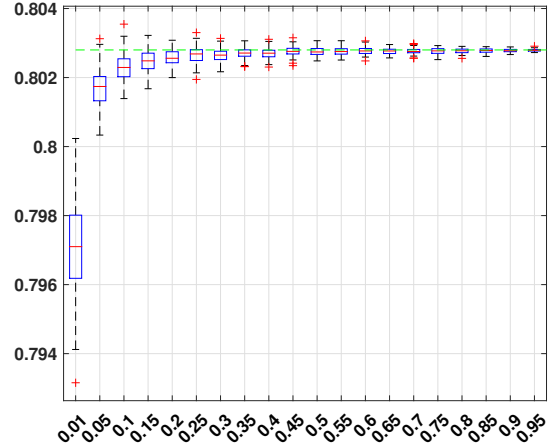


Fig. 7. FSIM correlation scores against percentage  $\frac{p}{100}$  on TID13 over 100 iterations. The actual correlation score is denoted by green. This figure can be better seen in color.

- [14] K. Seshadrinathan and A. C. Bovik, "A structural similarity metric for video based on motion models," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 1. IEEE, 2007, pp. 1–869.
- [15] —, "Motion tuned spatio-temporal quality assessment of natural videos," *Image Processing, IEEE Transactions on*, vol. 19, no. 2, pp. 335–350, 2010.
- [16] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *Image Processing, IEEE Transactions on*, vol. 21, no. 8, pp. 3339–3352, 2012.