Predicting the Quality of Images Compressed After Distortion in Two Steps

Xiangxu Yu, Christos G. Bampis, Praful Gupta, and Alan C. Bovik

Laboratory for Image and Video Engineering (LIVE), Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA

ABSTRACT

Full-reference and reduced-reference image quality assessment (IQA) models assume a high quality reference against which to measure perceptual quality. However, this assumption may be violated when the source image is upscaled, poorly exposed, or otherwise distorted before being compressed. Reference IQA models on a compressed but previously distorted "reference" may produce unpredictable results. Hence we propose 2stepQA, which integrates no-reference (NR) and reference (R) measurements into the quality prediction process. The NR module accounts for imperfect quality of the reference image, while the R component measures further quality from compression. A simple, efficient multiplication step fuses these into a single score. We deploy MS-SSIM as the R component and NIQE as the NR component and combine them using multiplication. We chose MS-SSIM, since it is efficient and correlates well with subjective scores. Likewise, NIQE is simple, efficient, and generic, and does not require training on subjective data. The 2stepQA approach can be generalized by combining other R and NR models. We also built a new data resource: LIVE Wild Compressed Picture Database, where authentically distorted reference images were JPEG compressed at four levels. 2stepQA is shown to achieve standout performance compared to other IQA models. The proposed approach is made publicly available at https://github.com/xiangxuyu/2stepQA .

Keywords: Image quality assessment, two-step, reference-no-reference, low quality reference image

1. INTRODUCTION

The past few years have witnessed tremendous growth in the viewing and sharing of digital images and videos.[1] Numerous consumer-driven and social media applications, such as Snapchat, Facebook, and Twitter allow users to rapidly post and share images that they capture with their handheld devices. These images are prone to a number of very common in-capture distortions, such as camera noise, poor exposure, motion blur, and compression artifacts. Streaming companies such as Netflix and Youtube offer a large amount of legacy content including old movies or television programs.[2] These video streams, which are obtained from diverse sources, are often subjected to source-related artifacts such as interlacing, film grain or upscaling. These source videos of low quality are then encoded for transmission. Being able to evaluate the quality of these twice-distorted images and videos on the receiver side is of great interest. In this paper, we only consider still pictures that undergo two stages of distortions (source impairments followed by compression), but the principle is extensible to general distortion types and videos.

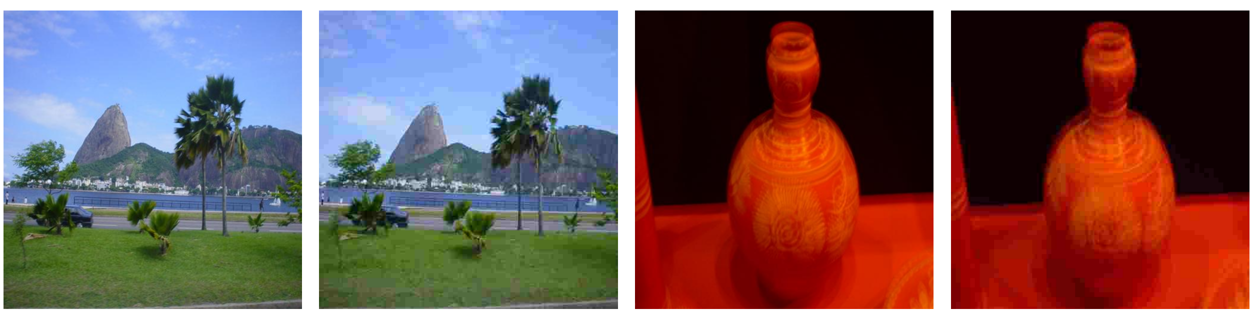
In the majority of cases, human viewers are the end users, and picture quality evaluation is crucial when storing, processing and transmitting visual data. Since subjective quality rating is impractical at scale, numerous perceptually-designed objective image quality assessment (IQA) algorithms have been developed for image quality prediction. These are broadly classified either as reference (R) models (including both full-reference and

Further author information: (Send correspondence to X.Y.)
X.Y.: E-mail: yuxiangxu@utexas.edu
C.G.B.: E-mail: bampis@utexas.edu
P.G.: E-mail: praful_gupta@utexas.edu
A.C.B.: Email: bovik@ece.utexas.edu
Proc. SPIE 10752, Applications of Digital Image Processing XLI, 107520K (17 September 2018); doi: 10.1117/12.2318995 Event: SPIE Optical Engineering + Applications, 2018, San Diego, California, United States

reduced-reference), if both the distorted image and ostensibly pristine version of it are available to be compared. Full-reference models such as SSIM,[3] MS-SSIM,[4] MAD,[5] FSIM,[6] SR-SIM,[7] VSI[8] and HaarPSI,[9] deliver excellent quality prediction performance on subjective quality databases such as LIVE IQA,[10] TID2013[11] and CSIQ.[5] Reduced-reference models such as RRED[12] and SpEED-QA[13] extract perceptually relevant features representing only a subset of the avaliable image information when estimating image quality. No-reference (NR) IQA algorithms are applied when only a distorted image is available. These models are based on either natural scene statistics (NSS) or learning process, or both. Models like NIQE[14] are "completely blind" approaches in that they are only driven by NSS of good quality images, while models such as BRISQUE[15] and BIQA[16] exploit features combined with support vector regression. Purely data-driven NR IQA methods, such as CORNIA,[17] and other deep learning-based approaches,[18–21] have also been proposed in the recent years.

A common assumption of R IQA models that are used to quality-assess compressed images is that the source image to be coded is of high quality, hence can be regarded as a "pristine" reference. However, when source artifacts are present, this assumption is violated. Therefore, a reference computation, if applied on a low-quality source image and a compressed version of it, will likely produce an incorrect quality score. Here we have shown that reference quality prediction of imperfect images that are subsequently compressed can be greatly improved by accounting for the source image quality. Towards this end we have devised a simple but effective two-step R-NR IQA concept which is applied on a pair of images: a possibly distorted reference image and a compressed (hence further distorted) version of it. An initial NR step captures the prior quality of the reference image before compression, while the R step determines the further quality degradation between the reference and the distorted images. Here we focus on processing a possibly distorted reference image to predict the quality of a JPEG compressed version of it. This development is organized as follows. Section 2 details our two-step IQA approach, while Section 3 describes a recently developed subjective image database that we created to develop and evaluate the two-step method. Section 4 discusses the experimental results and Section 5 concludes with avenues for future work.

2. TWO-STEP REFERENCE-NO-REFERENCE IQA MODEL



(a) MOS: 79

(b) MOS: 51
DMOS: 52
MS-SSIM: 0.9346

(c) MOS: 45

(d) MOS: 36
DMOS: 32
MS-SSIM: 0.9674

Figure 1. (a) A high quality reference image. (b) A JPEG compressed version of (a). (c) A low quality reference image. (d) A JPEG compressed version of (c).

Reference image quality models have generally assumed that a source reference image is of high quality, e.g., like the image shown in Fig. 1(a). When this is true, it is straightforward to apply an R IQA model (e.g. MS-SSIM) to compare the quality of the original source image against that of its JPEG compressed version (e.g., the image in Fig. 1(b)). However, if the source is of reduced quality (Fig. 1(c)), then the reference quality prediction may be unreliable. To illustrate this, consider the following. The source images in Fig. 1(a) and Fig. 1(c) were drawn from the LIVE In the Wild Challenge IQA database. Each has an associated mean opinion score (MOS). In a separate study, the compressed versions in Fig. 1(b) and 1(d) were assigned both MOS and differential mean opinion scores (DMOS). MOS may be regarded as the absolute subjective quality of a distorted image, while DMOS reduces the effect of content by differencing the MOS of the source and distorted versions of each

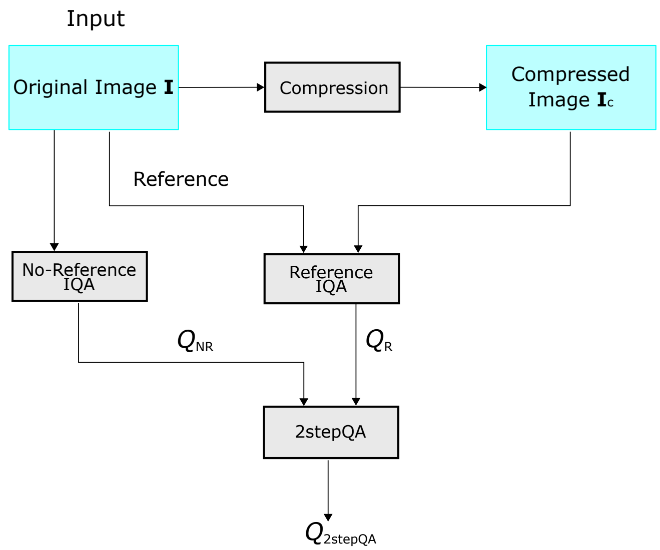


Figure 2. Overview of 2stepQA.

image. In practice, DMOS calculations are intended to remove content biases (e.g. due to image aesthetics), but in our case, this only reflects quality degradations arising from compression. In Fig. 1, the MS-SSIM values are in monotonic agreement with DMOS (increasing MS-SSIM corresponds to decreasing DMOS), while this is not the case for MOS (increasing MS-SSIM corresponds to decreasing in MOS). In other words, both MS-SSIM and DMOS conclude that 1(d) is of higher quality than 1(b), which contradicts the subjective opinion. This example shows that given a low quality reference image, a reference quality prediction may fail to correctly predict the quality of a compressed version of it.

To deal with this highly plausible situation, we introduce what we call the 2stepQA, which is a model and algorithm that combines R and NR quality measurements in a unified manner. The concept of 2stepQA is illustrated in Fig. 2. Given a reference source image $\mathbf{I}$ and a compressed version $\mathbf{I_c}$, the R module compares $\mathbf{I_c}$ with $\mathbf{I}$, while the NR module determines the quality of the source image $\mathbf{I}$. The NR prediction may be thought of as "prior" knowledge, that is used when measuring the reference perceptual quality. The scores produced by the R and NR modules are then combined, e.g. by computing a suitable product of the R and NR prediction scores, yielding a final 2stepQA score.

We now visually illustrate the potential advantages of 2stepQA. Consider a hypothetical image quality axis in Fig. 3 that spans the entire quality range, e.g. from "Bad" to "Excellent". The NR block of 2stepQA evaluates the quality of $\mathbf{I}$ by, for example, computing its perceptual distance from the space of high-quality natural images. By contrast, the R model measures the distance between $\mathbf{I}$ and $\mathbf{I_c}$: When $\mathbf{I}$ is of high quality, it will lie closely to the natural image space, and hence the quality of a compressed version of it may be accurately predicted as a reference algorithm score. However, if the quality of the reference image is poor, then the result of the NR will serve to correct the reference score. While there are many ways to combine the no-reference and reference scores into a single prediction, we have found that defining 2stepQA as a simple product of (functions of) the two scores to be quite effective.

2.1 Reference IQA Module

A high-performance R IQA module is an essential ingredient of 2stepQA. An excellent choice is the multi-scale structural similarity (MS-SSIM) index,[4] which has found considerable commercial success. MS-SSIM extracts luminance, contrast and structural information in a multi-scale fashion, delivering quality scores ranging from 0 to 1, where larger values correspond to better quality. In our framework, MS-SSIM is applied to compare $\mathbf{I}$ with $\mathbf{I_c}$, the latter being degraded compression.

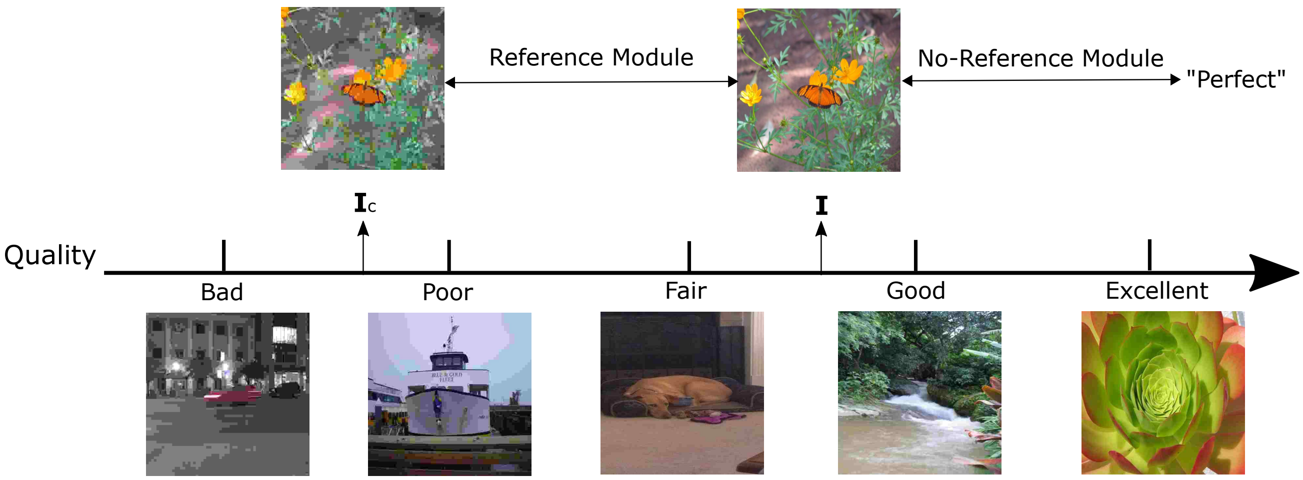


Figure 3. Visual illustration of 2stepQA concept. The represented quality condition ranges from low quality to high quality on the image quality axis with five typical levels and example images. Image $\mathbf{I}$ represents a possible pristine image with fair quality level and $\mathbf{I_c}$ is its compressed version. Reference module measures the deviation of $\mathbf{I_c}$ from $\mathbf{I}$, while no-reference module evaluates the distance between $\mathbf{I}$ and "perfect" natural image space.

2.2 No-Reference IQA Module

The end goal of 2stepQA is to accurately predict the perceptual quality of a compressed source image. Towards this aim, an efficient NR IQA model is also required and enough. In our prototype 2stepQA model, we use the NIQE index,[14] which is a blind NSS-based[22] IQA model that requires no training on distorted image or human opinions of them. The main idea behind NIQE is that the empirical distributions of mean-subtracted and divisively normalized luminance coefficients of high-quality images follow a gaussian distribution. However, in the presence of distortions, the empirical distributions tend to stray from gaussianity. By measuring these statistical deviations, a robust, blind image quality assessment engine is arrived at. Unlike data-driven approaches such as BRISQUE,[15] NIQE is very general, and serves to simplify 2stepQA while delivering good prediction performance.

2.3 2stepQA Model

Combining reference and no-reference models is the main idea behind 2stepQA. As such, 2stepQA should satisfy two properties: when the reference image is of low quality, the overall 2stepQA score should reflect this, by being strongly affected by the NR score, rather than relying on the reference comparison alone. Conversely, when the reference image is of high quality, the 2stepQA score should align closely with the R IQA score. A direct and simple way to satisfy these properties is to define 2stepQA as a product:

$$Q_{\mathsf{2step}} = Q_{\mathsf{R}} \cdot Q_{\mathsf{NR}} \tag{1}$$

where Q_{R} corresponds to a suitably mapped score (here we use MS-SSIM) computed on a reference image and a distorted version of it, while Q_{NR} is a suitably mapped NR score (here, NIQE) of the reference image. A key aspect of 2stepQA is that other combinations of reference and no-reference models may be used. This lends a high degree of versatility to 2stepQA as it provides a general framework for addressing this problem. Because of popularity and efficacy of the MS-SSIM and NIQE models, we define the exemplar 2stepQA model to be

$$Q_{\mathsf{2step}} = \text{MS-SSIM} \cdot (1 - \frac{\text{NIQE}}{\alpha}) \tag{2}$$

i.e. $Q_{\mathsf{R}} = \text{MS-SSIM}$, $Q_{\mathsf{NR}} = 1 - \frac{\text{NIQE}}{\alpha}$ and $\alpha = 100$. The form of Q_{NR} is arrived at since NIQE increases with worsening picture quality on a scale of about [0,100] on known databases. MS-SSIM rises with better picture quality over the range [0,1]. While 2stepQA is very simple, we will show that it is a very effective formulation.

2.4 Generalized Exponential Model

The range of quality scores realized by an IQA algorithm can vary significantly depending on the algorithm's design. For instance, a number of R IQA models, such as SSIM, MS-SSIM, FSIM, and so on, are designed to output quality scores ranging from 0 to 1, while NR IQA models, which are data-dependent or trained on human opinions of image quality, usually have the same range as MOS/DMOS. In order to generalize 2stepQA to account for other combinations of R-NR components, it is necessary to ensure they reside on a common scale, and therefore, it is important to scale them appropriately before combining them into a single piece. To rescale R and NR module scores into the same range, a simple logistic rescaling process is applied to match the ranges and trends of the reference and no-reference scores with MOS. A four-parameter, monotonic logistic function is utilized to fit the predicted quality score Q to the subjective MOS as given by

$$Q' = \beta_2 + \frac{\beta_1 - \beta_2}{1 + e^{-(Q - \beta_3/|\beta_4|)}} \tag{3}$$

where Q' is the rescaled score after finding the least-squares best-fitting logistic function over the four parameters $\{\beta_i; i = 1, .., 4\}$.

Any IQA database, in principle, can be used as the training data source for a given IQA model to determine β parameters of Eq. (3). In our experiments, the optimal βs are computed by minimizing the least squares error between the rescaled objective scores and the subjective MOS from the entire LIVE IQA Database. Since a reference image may possibly be afflicted by an unrestrained set of complex distortions, images of all distortion types from the LIVE corpus were utilized for fitting the logistic function to obtain parameters β_{NR1} to β_{NR4}. On the other hand, since in our simulations distorted images are corrupted by only compression artifacts, the JPEG subset of the LIVE IQA database was used to determine the parameters β_{R1} to β_{R4}.

Given a distorted image and its reference image, the NR quality score Q_{NR} is obtained from the NR module applied on the reference image, while the R quality score Q_{R} is obtained from the R component applied on both distorted and its reference. The rescaled scores Q'_{NR} and Q'_{R} are computed as

$$Q'_{\mathsf{NR}} = \beta_{\mathsf{NR2}} + \frac{\beta_{\mathsf{NR1}} - \beta_{\mathsf{NR2}}}{1 + e^{-(Q_{\mathsf{NR}} - \beta_{\mathsf{NR3}}/|\beta_{\mathsf{NR4}}|)}} \tag{4}$$

$$Q'_{\mathsf{R}} = \beta_{\mathsf{R2}} + \frac{\beta_{\mathsf{R1}} - \beta_{\mathsf{R2}}}{1 + e^{-(Q_{\mathsf{R}} - \beta_{\mathsf{R3}}/|\beta_{\mathsf{R4}}|)}}. \tag{5}$$

The rescaled scores Q'_{NR} and Q'_{R}, which belong to the same range as MOS are combined as follows:

$$Q_{\mathsf{G}} = (Q'_{\mathsf{NR}})^{\gamma} \cdot (Q'_{\mathsf{R}})^{1-\gamma} \tag{6}$$

where $\gamma \in [0, 1]$ is the exponential parameter used to adjust the relative importance of R and NR components of the generalized exponential model. As discussed in Section 4, γ assumes value close to 0.5 for most of the databases indicating that both R and NR components when scaled appropriately contribute equally towards quality prediction. Hence, depending on the application, γ can be fixed to $\gamma = 0.5$ or optimized appropriately using the training set.

3. A NEW DISTORTED-THEN-COMPRESSED IMAGE DATABASE

The majority of publicly available large-scale benchmark IQA databases, such as LIVE IQA,[10] TID2013[11] and CSIQ,[5] use high quality images as reference that correspond to distorted images that were created by introducing synthetic distortions on the high quality source images. While these popular IQA databases have been quite valuable in the development of classical IQA algorithms, they do not address the degraded source quality in the context of subsequent compression distortion. We have created such a resource, which we call the LIVE Wild Compressed Picture Database. We started with a set of 80 distorted images randomly drawn from the existing LIVE In the Wild Challenge Image Quality Database,[23] which contains a large number of images with widely diverse authentic image distortions, that were captured using a representative variety of mobile devices.

Since our goal is to predict visual quality when images of varying quality are subjected to compression, these 80 source images were then JPEG compressed into four different and perceptually distinguishable levels, yielding four compressed images per distorted content, and 320 distorted-then-compressed images in total. We then conducted a subjective study to collect human opinion scores on the 400 images (including the 80 distorted source images). The raw data was processed based on[24] and we followed the ITU-R BT 500.13 recommendation[25] for subject rejection.

4. PERFORMANCE EVALUATION

Table 1. Performance of different reference and no-reference IQA models on the live wild compressed picture database. The best performance algorithm is highlighted in bold font. Italics indicate no-reference algorithms.

	PSNR	MS-SSIM	FSIM	VSI	*NIQE*	*BRISQUE*	*CORNIA*	2stepQA
SROCC	0.4227	0.8930	0.9101	0.7953	*0.8457*	*0.9091*	*0.9005*	**0.9311**
LCC	0.4299	0.8923	0.9134	0.8153	*0.8407*	*0.8966*	*0.8955*	**0.9305**

Table 2. SROCC performance of different combinations of reference and no-reference IQA models using general two-step exponential model on the Live Wild Compressed Picture Database. The corresponding exponential parameter value γ is in bracket. Results using MOS as NR module score are posted for comparison.

	PSNR	MS-SSIM	FSIM	VSI
NIQE	0.6609(0.63)	0.9283(0.47)	0.9263(0.37)	0.8775(0.54)
BRISQUE	0.6833(0.62)	0.9333(0.46)	0.9357(0.41)	0.8980(0.53)
CORNIA	0.6807(0.51)	0.9356(0.39)	0.9375(0.36)	0.8992(0.46)
MOS	0.6156(0.71)	0.9401(0.61)	0.9474(0.56)	0.8946(0.64)

Table 3. LCC performance of different combinations of reference and no-reference IQA models using general two-step exponential model on the Live Wild Compressed Picture Database. The corresponding exponential parameter value γ is in bracket. Results using MOS as NR module score are posted for comparison.

	PSNR	MS-SSIM	FSIM	VSI
NIQE	0.6830(0.63)	0.9268(0.47)	0.9278(0.37)	0.8839(0.54)
BRISQUE	0.6743(0.62)	0.9309(0.46)	0.9355(0.41)	0.8988(0.53)
CORNIA	0.6747(0.51)	0.9353(0.39)	0.9394(0.36)	0.9026(0.46)
MOS	0.6064(0.71)	0.9403(0.61)	0.9499(0.56)	0.8981(0.64)

We now describe the experimental analysis of 2stepQA and its comparison against models that was carried out on the LIVE Wild Compressed Picture Database. The performances of the compared objective prediction models was measured using the Spearman Rank Order Correlation Coefficient (SROCC) and the Pearson's (linear) Correlation Coefficient (LCC). The former measures the monotonicity between the subjective and objective scores, while the latter measures the degree of linear agreement between them. The subjective scores were processed by the subject rejection protocols in[25] and[26] which yield similar results. For both evaluation metrics, a larger value denotes better model performance. The reported correlation values were calculated by performing 1000 random 80%-20% splits without content overlap, then taking the median value of the results. Before computing LCC, a logistic non-linearity[24] was used to map the quality scores to the subjective scores.

The rescaled indices (1) and (2) in each case were computed and these parameters were held constant in all the experiments. We compared 2stepQA against the state-of-the-art reference models PSNR, MS-SSIM,[4] FSIM[6] and VSI,[8] with results given in Table 1. Among the reference models, PSNR was the worst-performing, but it does not utilize any perceptual properties. MS-SSIM and FSIM both delivered good performances. However, 2stepQA

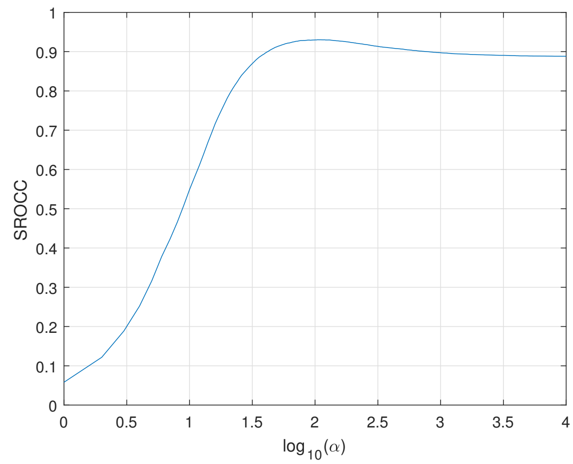


Figure 4. The performance of 2stepQA index under varied α values.

outperformed all of the other approaches showing that accounting for the quality of the reference source image is an effective strategy. We also report the performance of the NR models NIQE[14] and BRISQUE[15](applied on the distorted images only), and found that both performed well, but not as well as 2stepQA. Fig. 4 plots the performance of (2) against α. The 2stepQA index clearly attains its best performance at about $\alpha = 100$.

One of the greatest advantages of the 2stepQA framework is that it can incorporate reference and/or no-reference metrics depending on the application. To demonstrate this, we also report the results of other R-NR combinations in Table 2 and 3 using the generalized exponential model. The values of the logistic function parameter β_1 to β_4 of different R or NR models were determined using the LIVE Image Database. The value of the exponential parameter γ was determined to yield the best performance on each training set, and the predictor was then applied on the testing set. We also included the MOS as an NR component to demonstrate how well the NR IQA algorithms perform when used in combinations with popular R IQA algorithms. Across all possible R-NR combinations, we observed significant improvements when using Eq. (6).

5. FUTURE WORK

We proposed a novel two-step reference-no-reference IQA framework that integrates reference and no-reference information using a simple and efficient multiplication step. We believe that this approach will be useful to better perceptually optimize picture compression systems.[27] We evaluated the proposed approach on a recently developed subjective database and found that it produces standout performance compared to state-of-the-art reference models. We envision developing more sophisticated bayesian ways to integrate reference and no-reference information[2] e.g., by conditioning reference quality predictions on prior no-reference predictions of the source picture. We believe that the ideas presented here relate fully to the video quality assessment (VQA) problem as well, in cases where the original source video is not of high quality and can be subjected to source inspection prior to compression.

REFERENCES

[1] Bovik, A. C. and Wang, Z., [*Modern Image Quality Assessment*], New York: Morgan and Claypool (2006).

[2] Bovik, A. C., Bampis, C., and Goodall, T. R., "75-1: Invited paper: Perceptual issues of streaming video," in [*SID Int. Symp. Dig. Tech. Papers.*], **48**(1), 1097–1100, Wiley Online Library (2017).

[3] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (Apr. 2004).

[4] Wang, Z., Simoncelli, E. P., and Bovik, A. C., "Multiscale structural similarity for image quality assessment," in [*Proc. 37th Asilomar Conf. Signals Syst. Comput.*], **2**, 1398–1402 (Nov. 2003).

[5] Larson, E. C. and Chandler, D. M., "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.* **19**(1), 011006:1–011006:21 (Mar. 2010).

[6] Zhang, L., Zhang, L., Mou, X., and Zhang, D., "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.* **20**(8), 2378–2386 (2011).

[7] Zhang, L. and Li, H., "SR-SIM: A fast and high performance IQA index based on spectral residual," in [*Proc. IEEE Intl. Conf. Image Process.*], 1473–1476 (2012).

[8] Zhang, L., Shen, Y., and Li, H., "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.* **23**(10), 4270–4281 (2014).

[9] Reisenhofer, R., Bosse, S., Kutyniok, G., and Wiegand, T., "A Haar wavelet-based perceptual similarity index for image quality assessment," *Signal Process., Image Commun.* **61**, 33–43 (2018).

[10] Sheikh, H. R., Sabir, M. F., and Bovik, A. C., "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.* **15**(11), 3440–3451 (Nov. 2006).

[11] Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., et al., "Image database TID2013: Peculiarities, results and perspectives," *Signal Process., Image Commun.* **30**, 57–77 (2015).

[12] Soundararajan, R. and Bovik, A. C., "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.* **21**(2), 517–526 (Feb. 2011).

[13] Bampis, C. G., Gupta, P., Soundararajan, R., and Bovik, A. C., "SpEED-QA: Spatial efficient entropic differencing for image and video quality," *IEEE Signal Process. Lett.* (2017).

[14] Mittal, A., Soundararajan, R., and Bovik, A. C., "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.* **20**(3), 209–212 (2013).

[15] Mittal, A., Moorthy, A. K., and Bovik, A. C., "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.* **21**(12), 4695–4708 (2012).

[16] Li, Q., Lin, W., and Fang, Y., "BSD: Blind image quality assessment based on structural degradation," *Neurocomputing* **236**, 93–103 (2017).

[17] Ye, P., Kumar, J., Kang, L., and Doermann, D., "Unsupervised feature learning framework for no-reference image quality assessment," in [*Proc. IEEE Conf. Comp. Vis. Pattern Recog.*], 1098–1105, IEEE (2012).

[18] Kim, J., Zeng, H., Ghadiyaram, D., Lee, S., Zhang, L., and Bovik, A. C., "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 (2017).

[19] Hou, W., Gao, X., Tao, D., and Li, X., "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learn. Syst.* **26**(6), 1275–1286 (2015).

[20] Yang, J., Jiang, B., Zhu, Y., Ji, C., and Lu, W., "An image quality evaluation method based on joint deep learning," in [*Neural Inf. Process.*], 658–665, Springer (2017).

[21] Bosse, S., Maniry, D., Müller, K.-R., Wiegand, T., and Samek, W., "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.* **27**(1), 206–219 (2018).

[22] Ruderman, D. L., "The statistics of natural images," *Netw. Comput. Neural Syst.* **5**(4), 517–548 (1994).

[23] Ghadiyaram, D. and Bovik, A. C., "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.* **25**(1), 372–387 (2016).

[24] Seshadrinathan, K., Soundararajan, R., Bovik, A. C., and Cormack, L. K., "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.* **19**(6), 1427–1441 (2010).

[25] Union, I. T., "Methodology for the subjective assessment of the quality of television pictures itu-r recommendation bt.500-13," *Tech. Rep.* (2012).

[26] Li, Z. and Bampis, C. G., "Recover subjective quality scores from noisy measurements," in [*Proc. Data Compress. Conf.*], 52–61, IEEE (2017).

[27] Channappayya, S. S., Bovik, A. C., and Heath Jr, R. W., "Rate bounds on ssim index of quantized images," *IEEE Trans. Image Process.* **17**(9), 1624–1639 (2008).