Blind Quality Assessment of Multiply Distorted Images using Deep Neural Networks

Zhongling Wang[0000−0002−2129−1025], Shahrukh Athar[0000−0002−8871−669X], and Zhou Wang

Dept. of Electrical and Computer Engineering, University of Waterloo
Waterloo, ON, N2L 3G1, Canada
{zhongling.wang, shahrukh.athar, zhou.wang}@uwaterloo.ca

Abstract. In real-world visual content acquisition and distribution systems, a vast majority of visual content undergoes multiple distortions between the source and the end user. However, traditional image quality assessment (IQA) algorithms are usually validated and at times trained on image databases with a single distortion stage. Existing IQA methods for multiply distorted images remain limited in their scope and performance. In this work we design a first-of-its-kind blind IQA model for multiply distorted visual content based on a deep end-to-end convolutional neural network. The network is trained on a newly developed dataset which is composed of millions of multiply distorted images annotated with synthetic quality scores. Our tests on three publicly available subject-rated multiply distorted image databases show that the proposed model outperforms state-of-the-art blind IQA methods in terms of both accuracy and speed.

Keywords: Blind image quality assessment · Multiply distorted images · Convolutional Neural Networks (CNN) · Deep learning · Generalized Divisive Normalization (GDN) · Performance evaluation.

1 Introduction

The goal of objective Image Quality Assessment (IQA) is to predict the visual quality of images as perceived by humans. While simple error-based methods such as the Peak Signal-to-Noise-Ratio (PSNR) were the methods of choice in the past, it has been comprehensively shown to be poorly aligned with human perception of visual quality [1]. Significant strides have been made in the last two decades in designing perceptual quality methods and three major frameworks are now well-established in IQA research [2]: 1) Full-Reference (FR) IQA, 2) Reduced Reference (RR) IQA, and 3) No Reference (NR) or Blind IQA. To evaluate the quality of a distorted image, FR IQA methods require the complete availability of a pristine quality reference image, while RR IQA methods require access to certain features that have been extracted from the reference image. On the other hand, NR IQA methods evaluate the quality of the distorted image in the absence of the reference image. The performance of IQA methods is usually

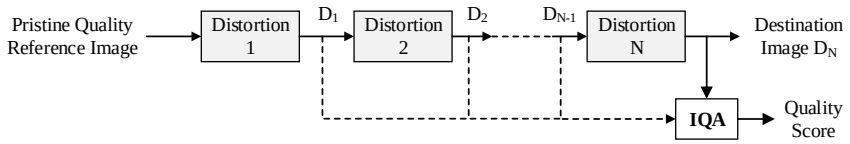

Fig. 1. The framework of a practical media distribution system.

tested on subject rated image databases such as LIVE Release 2 [3, 4], TID2008 [5], TID2013 [6], CSIQ [7], VCLFER [8], and CIDIQ [9]. Training for machine learning based IQA methods is also typically done on these databases.

Notwithstanding the significant progress made thus far in perceptual IQA research, the following challenges need to be addressed: 1) Although IQA methods are designed to handle different types of distortions, they are typically designed for images that have undergone a single stage of distortion, that is, they can handle one distortion at a time. This is in contrast to real world media delivery chains where the same visual content can undergo a number of distortions before reaching the end user, as depicted in Fig. 1. Designing IQA methods to deal with simultaneous distortions is quite challenging since the interactions of different distortions need to be accounted for. Thus, IQA for images with multiple simultaneous distortions has been a major challenge that future IQA research needs to address [10]. 2) In practical media delivery systems, access to pristine reference images is either extremely rare or altogether nonexistent, especially at the end user level. This, coupled with the multiple distortion nature of such systems, makes the use of FR and RR IQA infeasible. 3) One way to address the first two challenges is to use NR IQA. However, most NR methods are trained and tested on subject rated databases mentioned earlier, which have images with a single stage of distortion. Although there have been recent advances in the design of NR IQA methods to handle multiply distorted images using some new databases (as will be described later), such progress remains limited in scope. 4) The design of machine learning based IQA methods requires large-scale annotated image databases. However, subject-rated IQA databases, particularly for multiply distorted images remain quite limited, making it difficult to avoid model overfitting or to analyze interactions across distortions. These challenges motivate us to develop an end-to-end deep neural network (DNN) based NR or blind IQA model trained from synthetic scores.

2 Related Work

The first IQA database specifically designed for images with multiple simultaneous distortions (multiply distorted images), is the LIVE Multiply Distorted (MD) database [11]. Starting with 15 reference images, LIVE MD has 450 images divided into two parts, one each for the multiple distortion combinations

of 1) Gaussian blur followed by JPEG compression and 2) Gaussian blur followed by Gaussian noise. The MDID2013 database [12] has 12 reference images and one distortion combination of Gaussian blur followed by JPEG compression followed by white noise contamination. Overall, MDID2013 has 324 distorted images. The MDID database [13] has 20 reference images and 1600 multiply distorted images, where distortions are introduced first by adding Gaussian blur and contrast change, then JPEG or JPEG2000 compression, and finally Gaussian noise. The intensity of each distortion type is randomly selected and thus, MDID images may be distorted from 1 to 4 distortions. The MDIVL database [14] is composed of 10 reference images and 750 distorted images which are divided into two parts of 1) Gaussian blur followed by JPEG compresion and 2) Gaussian noise followed by JPEG compression. Two databases composed of authentically distorted images, where distortions are not artificially added but captured in real-world environments, have recently been published. The CID2013 database [15] is composed of 480 photographs captured by 79 different cameras of varying quality. The LIVE in the Wild Image Quality Challenge (WC) database [16] is composed of 1162 photographs taken by a diverse set of mobile device cameras.

Recently some blind IQA methods for multiply distorted images have been proposed. SISBLIM [12] is a training-free metric designed for singly and multiply distorted images through the fusion of estimates of noise, blur, JPEG compression, and joint effects. BoWSF [17] selects features sensitive to different distortion types, which are encoded through a Bag-of-Words model and mapped to a quality score. LQAF [18] uses Support Vector Regression (SVR) to map features such as phase congruency, gradient magnitude, gray level gradient co-occurrence matrix and the contrast sensitivity function to quality scores. An enhanced and multi-scale version of LQAF, called MS-LQAF is proposed in [19]. The training-based GWHGLBP [20] uses the gradient-weighted histogram of the local binary pattern (LBP) generated on the gradient map of the distorted image to capture the effects of multiple distortions. Jet-LBP [21] uses color Gaussian jets to generate feature maps from a distorted image. The LBP is applied to these feature maps to ascertain the effect of multiple distortions, leading to a weighted histogram which is mapped to quality scores through SVR. MUSIQUE [22] handles multiply distorted images and operates by performing distortion identification followed by distortion parameter estimation and score generation.

A major challenge in building Convolutional Neural Networks (CNN) for blind IQA is the lack of large-scale subject-rated data for training. Data augmentation by simple geometric transformations is widely used [23–25], though the content variation is still limited by the original training samples and the perceptual quality degradation due to these transformations is ignored. Cropping fixed size small patches from the original image is another popular way to increase training samples, but assigning the subjective quality scores of the entire image to all individual patches may introduce significant label noise. The CNN employed in [26] contains 10 convolutional layers and 2 fully connected layers to estimate the quality of 32×32 image patches. Patch weight estimation was incorporated before score pooling in order to reduce label noise. In [27] the

weights of patches were predicted based on Prewitt magnitude of segmentation of an image. In BIECON [24], an FR IQA model was used to derive local scores of 32×32 patches and then a CNN was pre-trained using these patches with corresponding FR scores. The model was then fine-tuned on a subject-rated dataset. Several models [23, 25, 28–30] alleviate the label noise problem by utilizing larger patches. As a consequence the number of training patches is reduced.

An early CNN-based blind IQA model IQA-CNN [31] was composed of a single convolutional layer with two fully connected layers. Quality scores of 32×32 image patches were predicted and then averaged to obtain the final result. This model was further extended to IQA-CNN++ [32] which has deeper networks and trained a multi-task CNN by predicting quality and distortion type simultaneously. Fu *et al.* [33] developed a CNN-based blind IQA model using a different pooling strategy along with a higher number of features compared to [31]. Distortion type information was incorporated in MEON [23], where the distortion classification sub-network was pre-trained without access to subject-rated data, for which a large amount of data is available for training. The quality prediction sub-network of MEON incorporated the distortion type information derived from the first sub-network. The two sub-networks were finally joint optimized on subject-rated datasets. Apart from end-to-end learning methods, some models [29, 30, 34] utilized SVR to predict the quality score based on CNN features. Instead of predicting a single quality score, some models [25, 28] predict the quality distribution of a given image using CNNs. Talebi *et al.* [25] approximated the score distribution through maximum entropy optimization. Zeng *et al.* [28] handle this problem by mapping the score to a series of vectorized probability quality representations defined by quality anchors.

A critical problem of the aforementioned learning-based methods including CNN-based models is that they are trained on datasets like LIVE Release 2 [3, 4], LIVE MD [11], LIVE WC [16], CSIQ [7], TID2008 [5] and TID2013 [6], which only have a limited number of subject-rated images. Multi-task learning (such as MEON [23]) and data augmentation only partially mitigate the problem. The lack of training data often leads to overfitting of these models and makes them hard to generalize to new data on which they are not trained. The ultimate goal of a *blind* IQA metric is to be robust to unseen data.

3 Proposed Model

We propose a CNN based approach to build a blind IQA model for multiply distorted images, namely End-to-end Optimized deep neural Network using Synthetic Scores (EONSS). EONSS addresses the following design issues: 1) The complex interactions between multiple distortions are learned through the deep CNN of perceptually motivated activation function; 2) a large-scale database of diverse content type and distortion variation is created. Such a database, together with a dedicated synthetic score generation approach, is employed to train the CNN, as opposed to earlier machine learning based methods that used

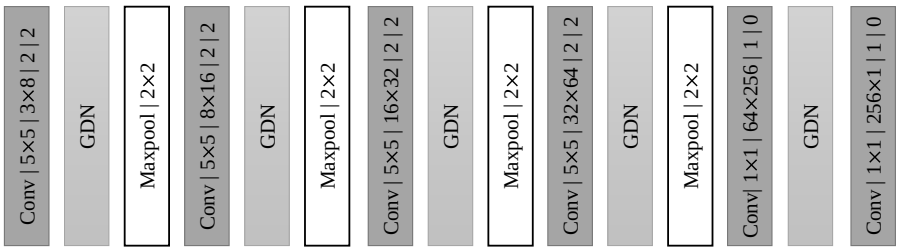

Fig. 2. The architecture of the proposed EONSS network. The format of the notation of the convolution layer is (*Conv* | *kernel width* × *kernel height* | *input channel* × *output channel* | *stride* | *padding*).

very limited data for training; 3) the trained model operates in a fast manner, suited for practical time-critical applications.

The network takes a 235×235 RGB image patch as input and predicts its quality. 235×235 patches contain more visually meaningful content than small patches and can represent the whole image better, therefore alleviating the issue of label noise. However, due to the limited training images in the classic IQA datasets (mentioned earlier), the number of training patches for many models decrease dramatically when larger patch size is utilized. Our model does not suffer from this issue since the number of training images in our dataset is sufficiently large (described later). We illustrate the architecture of the network in Fig. 2. The network consists of 6 stages of processing. Each of the first 4 stages contains a convolutional, generalized divisive normalization (GDN) [35] and max-pooling layer. These 4 stages aim at mapping the $235 \times 235 \times 3$ raw pixels from the image space to a lower-dimensional feature space where perceptually aware image distortions can be quantified more easily. The network reduces the spatial dimension through the use of convolution with stride 2×2 in the first 4 stages. 2×2 max-pooling is also used after each GDN layers in the first 4 stages to select the neurons that have the highest local response. Finally, the last 2 stages, which consist of 2 fully connected layers and a GDN transform layer in between, map the extracted feature to a single quality score. The spatial size of features are reduced to 1×1 before they are sent into the last 2 fully connected layers so that the number of weights in the fully connected layers are dramatically reduced. We apply GDN [35] instead of ReLU [36] after the convolution layers in the first 5 stages as the activation function to add non-linearity to the model. Although ReLU [36] is widely used as the activation function in CNNs, it suffers from strong higher-order dependencies, which is often compromised with a much larger network. Here we utilized a bio-inspired normalization transform, GDN, as the activation function. It helps decorrelate the high-dimensional features by using a joint nonlinear gain control mechanism. As a result a much smaller network is needed in order to achieve competitive performance.

The new Waterloo Multiply Distorted (Waterloo MD) IQA database has been used to learn the EONSS model. The construction of this database is beyond the focus of this work and will be covered in detail in our other publications. Suffice it to say that the Waterloo MD database is composed of 3570 pristine reference images and has distorted images which have been afflicted by up to two distortions. The database has three single distortion categories of Noise, Blur and JPEG compression, with 39270 images in each category. More importantly, the database has five multiple distortion categories of Noise-JPEG, Noise-JPEG2000, Blur-JPEG, Blur-Noise and JPEG-JPEG, with 667590 images in each category. Overall the database has around 3.45 million distorted images that have been annotated with synthetic quality scores in place of human subjective ratings. The Waterloo MD database is split randomly into training, validation and testing sets without overlapping image content of size 60%, 20% and 20%, respectively. Since the input dimension of the network is fixed to $235 \times 235 \times 3$, in the training phase, for the sake of time efficiency, we randomly sample one 235×235 patch from each image if its dimensions are larger. Since our dataset is large, this does not hinder us from creating a sufficiently large training set. In addition, by doing this, we obtain a batch of image patches of greater diversity and therefore significantly increase the time efficiency and prevent overfitting. The image quality of the sampled patch is considered to be the same as the original image during training since the 235×235 patch size can cover a relatively large area of the original image and therefore is able to contain perceptually meaningful content. In the validation and testing phases, for images with larger dimensions, we sample 235×235 patches from the original image with a stride of 128×128 in an overlapping manner. The average of the predicted quality scores of these patches is computed as the predicted quality of the original image.

We follow the approach in [37] to initialize the weights of the convolution layers. Adam [38] is used for optimization. The training batch size is chosen to be 50 and the image patches in each batch are randomly sampled from the training set only. The initial learning rate is set to 0.00001 and is decreased by a factor of 10 after every 2 epochs. Other parameters of Adam are set as default. We test the model performance (PLCC and SRCC) on the validation set after each epoch and stop training after 10 epochs when the performance on the validation set reaches a plateau. Finally, the model after 10 epochs of training is applied to the testing set.

4 Experimental Results

In addition to evaluating the performance of EONSS on the test set of the new Waterloo MD database, we tested its performance on three publicly available multiply distorted image databases: MDID [13], LIVE MD [11] and MDIVL [14]. It is pertinent to mention here that these three databases were not used in the training and validation process of EONSS, and their ground truths are the mean opinion scores (MOSs) obtained from subjective testing rather than syn-

Table 1. Performance comparison of IQA Algorithms. FR methods are in bold.

Part 1: SRCC					Part 2: PLCC			
Metric	MDID	LIVE MD	MDIVL	Weighted Average	MDID	LIVE MD	MDIVL	Weighted Average
IWSSIM	0.8911	0.8836	0.8588	0.8812	0.8983	0.9109	0.9056	0.9023
EONSS	0.8297	0.7260	0.8833	0.8274	0.8179	0.8437	0.8744	0.8372
CORNIA	0.7918	0.8340	0.8336	0.8098	0.7907	0.8679	0.8277	0.8130
GWHGLBP	0.7032	0.9698	0.5841	0.7141	0.7035	0.9663	0.5737	0.7110
ILNIQE	0.6900	0.8778	0.6238	0.7025	0.7053	0.8923	0.6303	0.7153
dipIQ	0.6612	0.6678	0.7131	0.6762	0.6738	0.7669	0.7627	0.7126
HOSA	0.6412	0.6393	0.7399	0.6673	0.6521	0.6768	0.7167	0.6734
SISBLIM	0.6554	0.8770	0.5375	0.6594	0.6700	0.8948	0.5724	0.6800
NIQE	0.6523	0.7738	0.5713	0.6501	0.6728	0.8387	0.5688	0.6716
PSNR	0.5784	0.6771	0.6136	0.6037	0.6091	0.7398	0.6806	0.6493
BIQI	0.6276	0.5556	0.5711	0.6009	0.6372	0.7389	0.6215	0.6493
NRSL	0.6458	0.4145	0.6047	0.5976	0.6502	0.4829	0.6794	0.6311
BRISQUE	0.4035	0.5018	0.6647	0.4893	0.4558	0.6045	0.6516	0.5321
MEON	0.4861	0.1917	0.5466	0.4550	0.5168	0.2339	0.5722	0.4862
DeepIQA[1]	0.4040	0.2379	0.5614	0.4195	0.4215	0.2897	0.5213	0.4271
QAC	0.3239	0.3579	0.5524	0.3906	0.6043	0.4145	0.5713	0.5650
LPSI	0.0306	0.2717	0.5736	0.2148	0.4336	0.5464	0.5715	0.4887
GMLOG	0.0546	0.1841	0.2656	0.1319	0.2626	0.3087	0.3830	0.3023

[1] Of the four NR models provided by the authors, the weighted model trained on LIVE Release 2 was used.

thetic scores. To compare how EONSS performs against other IQA methods, we also tested the performance of 15 publicly available blind IQA methods. These include two methods designed for multiply distorted images, GWHGLBP [20] and SISBLIM [12], and two methods designed by using CNNs, DeepIQA [26] and MEON [23]. The following state-of-the-art blind IQA methods were also tested for comparison: BIQI [39], BRISQUE [40], CORNIA [41], dipIQ [42], GMLOG [43], HOSA [44], ILNIQE [45], LPSI [46], NIQE [47], NRSL [48] and QAC [49]. The performance of two FR IQA methods, PSNR and the state-of-the-art IWSSIM [50] was also evaluated to provide a FR reference point. Two performance evaluation criteria were used: Spearman Rank Correlation Coefficient (SRCC) to assess prediction monotonicity and Pearson Linear Correlation Coefficient (PLCC) to assess prediction accuracy [51]. A five-parameter logistic function [4] was used to perform non-linear mapping of objective scores to MOS/DMOS of respective databases before the computation of PLCC. A better objective method should have higher SRCC and PLCC, ideally close to 1. Parts 1 and 2 of Table 1 respectively show the SRCC and PLCC of all tested methods for all three databases. To provide an overall comparison, weighted average SRCC and PLCC have been provided in Table 1 based on the number of images in the databases and methods in the table have been listed in descending order with respect to the weighted average SRCC numbers.

To draw statistically sound inferences about the performance of IQA methods, we carried out hypothesis testing on model prediction residuals (after non-linear mapping). First, a simple Kurtosis-based criterion was used to check for Gaussianity of residuals as in [4]. All model prediction residuals had a Kurtosis between 2 and 4, with the exception of GMLOG residuals on MDID database, and were assumed to be Gaussian. This allowed us to compare the model residuals through statistical significance testing by using the F-test [52]. The results

Table 2. Statistical Significance Testing results for competing IQA models [Database Order: MDID, LIVE MD, MDIVL]. Legend: IWSSIM (m1), EONSS (m2), CORNIA (m3), ILNIQE (m4), dipIQ (m5), GWHGLBP (m6), SISBLIM (m7), HOSA (m8), NIQE (m9), PSNR (m10), BIQI (m11), NRSL (m12), QAC (m13), BRISQUE (m14), LPSI (m15), MEON (m16), DeepIQA (m17), GMLOG (m18).

	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14	m15	m16	m17	m18
m1	___	111	111	111	111	101	111	111	111	111	111	111	111	111	111	111	111	111
m2	000	___	101	101	111	101	101	111	1_1	111	111	111	111	111	111	111	111	111
m3	000	010	___	101	111	101	101	111	111	111	111	111	111	111	111	111	111	111
m4	000	010	010	___	110	_0_	1__	110	11_	11_	11_	11_	11_	11_	11_	11_	111	111
m5	000	000	000	001	___	_01	_01	_11	_01	1_1	1_1	_11	111	111	111	111	111	111
m6	010	010	010	_1_	_10	___	11_	110	_1_	110	11_	110	11_	110	11_	11_	11_	111
m7	000	010	010	0__	_10	00_	___	_10	_1_	110	_1_	_10	11_	110	11_	11_	11_	111
m8	000	000	000	001	_00	001	_01	___	_01	10_	_01	_1_	111	111	111	111	111	111
m9	000	0_0	000	00_	_10	_0_	_0_	_10	___	110	_1_	_10	11_	110	11_	11_	11_	111
m10	000	000	000	00_	0_0	001	001	01_	001	___	__1	01_	_11	11_	111	111	111	111
m11	000	000	000	00_	0_0	00_	_0_	_10	_0_	__0	___	_10	_1_	11_	11_	11_	111	111
m12	000	000	000	00_	_00	001	_01	_0_	_01	10_	_01	___	1_1	10_	1_1	111	111	111
m13	000	000	000	00_	000	00_	00_	000	00_	_00	_0_	0_0	___	100	10_	1__	1__	1_1
m14	000	000	000	00_	000	001	001	000	001	00_	00_	01_	011	___	__1	_11	_11	111
m15	000	000	000	00_	000	00_	00_	000	00_	000	00_	0_0	01_	__0	___	01_	_1_	111
m16	000	000	000	00_	000	00_	00_	000	00_	000	00_	000	0__	_00	10_	___	1__	1_1
m17	000	000	000	000	000	00_	00_	000	00_	000	000	000	0__	_00	_0_	0__	___	1_1
m18	000	000	000	000	000	000	000	000	000	000	000	000	0_0	000	000	0-0	0-0	___

are shown in Table 2, where "1", "_", or "0" mean that the method in the row is statistically better, indistinguishable, or worse than the method in the column respectively, with 95% confidence. Each table entry has three digits, which represent testing on the MDID, LIVE MD and MDIVL databases, respectively.

From Tables 1 and 2 the following observations can be made: 1) EONSS performs better than all other blind IQA methods and the FR PSNR on the MDID and MDIVL databases. On LIVE MD, EONSS is statistically outperformed only by CORNIA, GWHGLBP, ILNIQE and SISBLIM, though it needs to be mentioned that GWHGLBP was trained on this very database. 2) Of all the blind IQA methods under test, EONSS and CORNIA are the only robust metrics since they perform consistently well on all three databases (PLCC greater than or close to 0.8). While methods such as SISBLIM, ILNIQE, NIQE and dipIQ perform well on one database, their performance drops on other databases. 3) Although CORNIA was originally designed by using images with a single distortion, its performance extends well for multiply distorted images. 4) The performance of all blind IQA methods is still a distance away from state-of-the-art FR IQA method IWSSIM, suggesting space for improvement. 5) The computational complexity of all IQA methods under test was evaluated in terms of their execution time to determine the quality of a 1024×1024 color image on a desktop computer with a 3.5GHz Intel Core i7-7800X processor, 16GB of RAM, NVIDIA GeForce GTX 1050Ti GPU and Ubuntu 18.04 operating system. The execution time relative to PSNR is given in Table 3, where metrics have been sorted in ascending order with respect to execution time. It appears that EONSS is around twenty or more times faster than competitive methods such as dipIQ, CORNIA, SISBLIM and ILNIQE, making it an excellent choice for practical applications.

Table 3. IQA method execution time.

Metric	Processing Unit	Execution Time (Seconds)	Execution Time Relative to PSNR
PSNR	CPU	0.0013	1.00
LPSI	CPU	0.0397	30.54
EONSS	GPU	0.0604	46.46
EONSS	CPU	0.0817	62.85
MEON	CPU	0.0819	63.00
MEON	GPU	0.0876	67.38
GMLOG	CPU	0.1044	80.31
HOSA	CPU	0.1309	100.69
QAC	CPU	0.1357	104.38
NRSL[1]	CPU	0.1421	109.31
GWHGLBP[1]	CPU	0.1469	113.00
DeepIQA	GPU	0.1549	119.15
BRISQUE	CPU	0.1823	140.23
NIQE	CPU	0.2941	226.23
BIQI	CPU	0.4634	356.46
IWSSIM	CPU	0.6067	466.69
dipIQ	CPU	1.6592	1276.31
CORNIA	CPU	2.0304	1561.85
SISBLIM	CPU	2.2005	1692.69
ILNIQE	CPU	2.5227	1940.54
DeepIQA	CPU	6.2818	4832.15

[1]Feature extraction time only.

5 Conclusion

We propose a new deep learning based blind IQA model called EONSS. Compared to other CNN-based models, such as DeepIQA [26] and MEON [23], EONSS uses a much simpler network architecture, but delivers superior performance with fairly low computational cost. The success of EONSS is partially due to its architecture design and its adoption of the perceptually motivated GDN as the activation function, and is also attributed to the Waterloo MD training database, which consists of millions of multiply distorted images whose quality has been synthetically annotated. The enormity and content-diversity of this database has provided sufficient data to the DNN to learn an adequate end-to-end blind IQA model. Since EONSS has been tested on publicly available subject-rated databases which were not part of the training and validation process, its superior performance has also validated the novel methodology of using large-scale synthetically annotated databases for learning new IQA models, providing a new perspective on how to resolve the longstanding problem of the lack of large-scale datasets in IQA research. Detailed account about the construction of the Waterloo MD database will be made available to the IQA community. The trained version of the EONSS model will also be made publicly available to facilitate reproducible research.

References

1. Z. Wang and A. C. Bovik, "Mean Squared Error: Love It or Leave It? A new look at Signal Fidelity Measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

2. Z. Wang and A. C. Bovik, "Modern Image Quality Assessment," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, no. 1, pp. 1–156, 2006, Morgan & Claypool Publishers.
3. H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2," Available: http://live.ece.utexas.edu/research/Quality/subjective.htm.
4. 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, Nov. 2006.
5. N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, J. Astola, M. Carli, and F. Battisti, "TID2008–A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics," *Adv. Modern Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009.
6. N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process.: Image Commun.*, vol. 30, pp. 57–77, Jan. 2015.
7. E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 011 006:1–011 006:21, 2010.
8. A. Zarić, N. Tatalović, N. Brajković, H. Hlevnjak, M. Lončarić, E. Dumić, and S. Grgić, "VCL@FER Image Quality Assessment Database," *AUTOMATIKA*, vol. 53, no. 4, pp. 344–354, 2012.
9. X. Liu, M. Pedersen, and J. Y. Hardeberg, "CID:IQ – A New Image Quality Database," in *Proc. Int. Conf. Image, Signal Process. (ICISP)*. Springer International Publishing, 2014, pp. 193–202.
10. D. M. Chandler, "Seven Challenges in Image Quality Assessment: Past, Present, and Future Research," *ISRN Signal Process.*, vol. 2013, Article ID 905685, pp. 1–53, 2013.
11. D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Conf. Rec. Asilomar Conf. Signals, Syst., Comput. (ASILOMAR)*, Pacific Grove, CA, USA, Nov. 2012, pp. 1693–1697.
12. K. Gu, G. Zhai, X. Yang, and W. Zhang, "Hybrid No-Reference Quality Metric for Singly and Multiply Distorted Images," *IEEE Trans. Broadcast.*, vol. 60, no. 3, pp. 555–567, Sept. 2014.
13. W. Sun, F. Zhou, and Q. Liao, "MDID: A multiply distorted image database for image quality assessment," *Pattern Recognit.*, vol. 61, pp. 153–168, 2017.
14. S. Corchs and F. Gasparini, "A Multidistortion Database for Image Quality," in *Proc. Int. Workshop Comput. Color Imag. (CCIW)*. Springer International Publishing, 2017, pp. 95–104.
15. T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen, and J. Häkkinen, "CID2013: A Database for Evaluating No-Reference Image Quality Assessment Algorithms," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 390–402, Jan. 2015.
16. D. Ghadiyaram and A. C. Bovik, "Massive Online Crowdsourced Study of Subjective and Objective Picture Quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, Jan. 2016.
17. Y. Lu, F. Xie, T. Liu, Z. Jiang, and D. Tao, "No Reference Quality Assessment for Multiply-Distorted Images Based on an Improved Bag-of-Words Model," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1811–1815, Oct. 2015.
18. C. Li, Y. Zhang, X. Wu, W. Fang, and L. Mao, "Blind multiply distorted image quality assessment using relevant perceptual features," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Quebec City, QC, Canada, Sept. 2015, pp. 4883–4886.

19. C. Li, Y. Zhang, X. Wu, and Y. Zheng, "A Multi-Scale Learning Local Phase and Amplitude Blind Image Quality Assessment for Multiply Distorted Images," *IEEE Access*, vol. 6, pp. 64 577–64 586, 2018.
20. Q. Li, W. Lin, and Y. Fang, "No-Reference Quality Assessment for Multiply-Distorted Images in Gradient Domain," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 541–545, Apr. 2016.
21. H. Hadizadeh and I. V. Bajić, "Color Gaussian Jet Features For No-Reference Quality Assessment of Multiply-Distorted Images," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1717–1721, Dec. 2016.
22. Y. Zhang and D. M. Chandler, "Opinion-Unaware Blind Quality Assessment of Multiply and Singly Distorted Images via Distortion Parameter Estimation," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5433–5448, Nov. 2018.
23. K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-End Blind Image Quality Assessment Using Deep Neural Networks," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018.
24. J. Kim and S. Lee, "Fully Deep Blind Image Quality Predictor," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 206–220, Feb. 2017.
25. H. Talebi and P. Milanfar, "NIMA: Neural Image Assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.
26. S. Bosse, D. Maniry, K. Müller, T. Wiegand, and W. Samek, "Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.
27. J. Li, L. Zou, J. Yan, D. Deng, T. Qu, and G. Xie, "No-reference image quality assessment using Prewitt magnitude based on convolutional neural networks," *Signal, Image, Video Process. (SIViP)*, vol. 10, no. 4, pp. 609–616, Apr. 2016.
28. H. Zeng, L. Zhang, and A. C. Bovik, "Blind Image Quality Assessment with a Probabilistic Quality Representation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Athens, Greece, Oct. 2018, pp. 609–613.
29. F. Gao, J. Yu, S. Zhu, Q. Huang, and Q. Tian, "Blind image quality prediction by exploiting multi-level deep representations," *Pattern Recognit.*, vol. 81, pp. 432 – 442, 2018.
30. S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image, Video Process. (SIViP)*, vol. 12, no. 2, pp. 355–362, Feb. 2018.
31. L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional Neural Networks for No-Reference Image Quality Assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, June 2014, pp. 1733–1740.
32. L. Kang, P. Ye, Y. Li, and D. Doermann, "Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Quebec City, QC, Canada, Sept. 2015, pp. 2791–2795.
33. J. Fu, H. Wang, and L. Zuo, "Blind image quality assessment for multiply distorted images via convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 1075–1079.
34. J. Li, J. Yan, D. Deng, W. Shi, and S. Deng, "No-reference image quality assessment based on hybrid model," *Signal, Image, Video Process. (SIViP)*, vol. 11, no. 6, pp. 985–992, Sept. 2017.
35. Q. Li and Z. Wang, "Reduced-Reference Image Quality Assessment Using Divisive Normalization-Based Image Representation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

36. V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Haifa, Israel, June 2010, pp. 807–814.
37. K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1026–1034.
38. D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. Int. Conf. Learn. Repr. (ICLR)*, San Diego, CA, USA, May 2015.
39. A. K. Moorthy and A. C. Bovik, "A Two-Step Framework for Constructing Blind Image Quality Indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.
40. A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
41. P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised Feature Learning Framework for No-Reference Image Quality Assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, June 2012, pp. 1098–1105.
42. K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: Blind Image Quality Assessment by Learning-to-Rank Discriminable Image Pairs," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3951–3964, Aug. 2017.
43. W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind Image Quality Assessment Using Joint Statistics of Gradient Magnitude and Laplacian Features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.
44. J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind Image Quality Assessment Based on High Order Statistics Aggregation," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4444–4457, Sept. 2016.
45. L. Zhang, L. Zhang, and A. C. Bovik, "A Feature-Enriched Completely Blind Image Quality Evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
46. Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Quebec City, QC, Canada, Sept. 2015, pp. 339–343.
47. A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
48. Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind Image Quality Assessment Using Statistical Structural and Luminance Features," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2457–2469, Dec. 2016.
49. W. Xue, L. Zhang, and X. Mou, "Learning without Human Scores for Blind Image Quality Assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Portland, OR, USA, June 2013, pp. 995–1002.
50. Z. Wang and Q. Li, "Information Content Weighting for Perceptual Image Quality Assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.
51. Video Quality Experts Group and others, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase II," 2003.
52. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman & Hall/CRC, 2011.