Q-DNN: A Quality-Aware Deep Neural Network for Blind Assessment of Enhanced Images

Qingbo Wu #1, Hongliang Li #2, Fanman Meng #3, King N. Ngan *,#4

School of Electronic Engineering, University of Electronic Science and Technology of China

* Department of Electronic Engineering, The Chinese University of Hong Kong

1 qbwu@uestc.edu.cn

2 hlli@uestc.edu.cn

3 fmmeng@uestc.edu.cn

4 knngan@ee.cuhk.edu.hk

***Abstract*—Image enhancement is widely popular due to its capability of producing "better" visual quality for specific applications. Although many enhancement algorithms have been developed in recent years, the studies towards blind assessment of enhanced images are still very lacking. In this paper, we propose a data-driven blind image quality assessment (BIQA) method based on the quality-aware deep neural network (Q-DNN). Unlike the conventional hand-crafted features designed for measuring the degradation level of specific distortion types, a supervised learning model is utilized in our Q-DNN, which is capable of adaptively updating the feature extractor and quality regressor for describing the visual artifacts caused by different image enhancement tasks. Experimental results on two challenging enhanced image databases show that the proposed method is significantly superior to the state-of-the-art BIQA metrics.**

***Index Terms*—Enhanced images, quality-aware deep neural network, blind image quality assessment, feature extractor, quality regressor.**

I. INTRODUCTION

Images captured in our daily life often present unsatisfying quality due to various external and artificial factors, such as, sensor noise, nonuniform illumination, undesirable action and so on. In this context, image enhancement becomes highly desirable for improving the perceptual image quality. Recent years have witnessed the significant progress of enhancement algorithms in many fields, including dehazing, deblurring and so on [1], [2]. However, due to lack of efficient quality metric, over- and under-enhancement are unavoidable in the practical applications of these enhancement algorithms. How to objectively evaluate and compare the perceptual qualities of different enhanced images has become the key to promote the advancement of image enhancement technologies.

Quality assessment of enhanced image encounters two major challenges in the real-world applications. Firstly, since there doesn't exist perfect enhanced image as a reference, only the blind metric is applicable in this task, which is more challenging than the full reference and reduced reference image quality assessment. Secondly, the evaluation criteria of image enhancement varies significantly across different applications, which brings great difficulty in developing general purpose BIQA metric. For example, in the character recognition task, the stronger edge and texture information are considered as "better" quality in evaluating the contrast enhancement algorithm. On the contrary, the blur effect is preferred for bokeh rendering to produce subtly soft feeling in many retouching applications.

Existing algorithms usually perform general purpose BIQA based on specific hand-crafted image features, which are followed by a quality regressor to map the image features to a quantitative quality score. A high prediction accuracy for the quality of degraded images is also achieved by these methods in many popular IQA databases [3], [4]. Unfortunately, most state-of-the-art BIQA methods are reported to perform poorly on the enhanced images [5], [6]. As discussed in [5], [6], the existing BIQA models are designed for measuring the obviousness of some typical synthetic artifacts, such as, blocking, blur and ringing. When coping with some new distortions caused by the enhancing process, these algorithms usually present poor generalization capability. Complicated changes across various application scenarios urgently require more flexible BIQA model, which is capable of adaptively updating the feature extractor and quality regressor for different visual artifacts.

In this paper, we propose a data-driven method for the blind assessment of enhanced images. Instead of using hand-crafted image features for specific distortion types, a quality-aware deep neural network (Q-DNN) is designed to learn the feature extractor and quality regressor simultaneously via a two-step architecture. Firstly, in order to capture the natural image prior across rich visual contents, we pre-train the Q-DNN in the large scale ImageNet dataset [7] which includes more than ten millions of images labeled by 1000 semantic categories. Secondly, to excavate the quality-aware features from previous category-relevant image prior, we fine-tune the pre-trained Q-

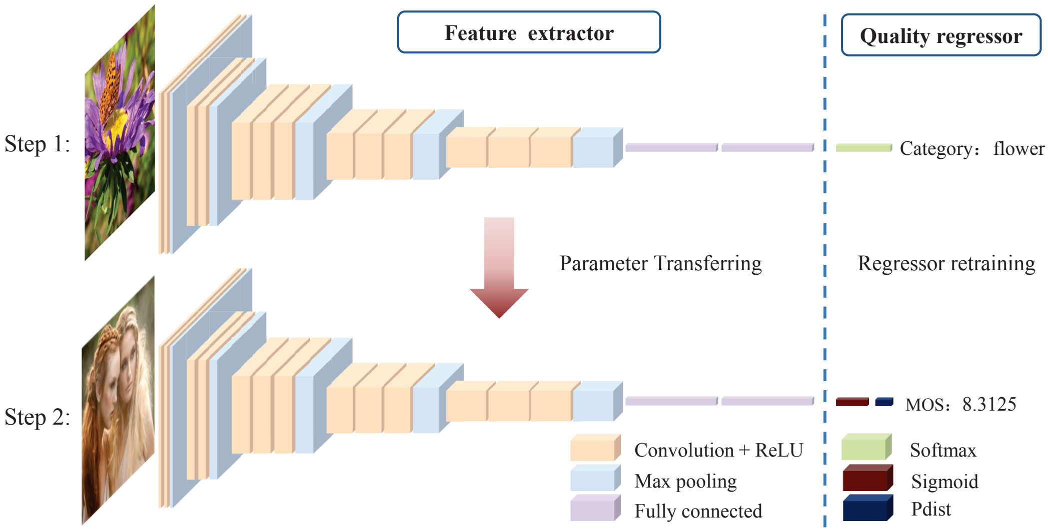

Fig. 1. Framework of the proposed two-step learning architecture.

DNN model by using human-rated images, which are labeled by the subjective quality scores. The proposed BIQA metric is easy to adapt to different enhancement applications by modifying the training samples in the fine-tuning step, which can update the feature extractor and quality regressor accordingly. Experiments on two publicly available enhanced image databases confirm the superiority of the proposed Q-DNN model in comparison to the traditional BIQA methods.

The remainder of this paper is organized as follows. Section II describes the details of proposed Q-DNN model. Section III shows the comparison results between our Q-DNN model and the state-of-the-art BIQA metrics on enhanced images. Finally, we draw the conclusion of this paper in Section IV.

II. Proposed Method

Inspired by the fact that the primate visual cortex shows deep hierarchies, we employ the deep neural network to simulate the image quality perception of human vision system, which is confirmed efficient for natural image description [8]. Specifically, the popular VGG network [9], which is composed of 16 convolution layers, is utilized as our feature extractor.

For clarity, the framework of the proposed method has been shown in Fig. 1, where the BIQA is decomposed into two steps. For short, we use F_1 and F_2 to the denote the feature extractors trained in steps 1 and 2, respectively. In our method, the step 1 aims to capture the natural image prior across significant variations of visual contents, which is crucial for enhancing the robustness of a BIQA metric [10]. To this end, we pre-train F_1 in a large scale database – ImageNet [7], where the category label is assigned to each image as the ground-truth. A Softmax layer is followed by the feature extractor to implement the supervised learning.

To convert the category-relevant feature extractor F_1 to a quality-aware descriptor F_2, in step 2, we try to fine tune the pre-trained feature extractor by using the application-specific training images, which are labeled by the subjective quality scores for different enhancement tasks. As shown in Fig. 1, the F_2 shares the same network structure with F_1, and the parameters pre-trained from F_1 are transferred to F_2 for initializing our quality-aware feature extractor. The Sigmoid and Pdist layers constitute the quality regressor, which aims for mapping the 4096-dimension feature produced from F_2 to a mean opinion score (MOS). In the following, we will describe the feature extractor and quality regressor in details.

A. Feature extractor model

As shown in Fig. 1, the feature vector is produced by recursively convoluting the sub-region of an image with a bank of linear filters. Let $w_i^k \in R^{m \times n}$ denote the parameters of kth $m \times n$ filter in the ith convolution layer, and $b_i^k \in R$ denote its corresponding bias. Then, the kth feature map produced in the ith convolution layer is represented by

$$h_i^k = g(w_i^k * \mathbf{h}_{i-1} + b_i^k) \tag{1}$$

where $\mathbf{h}_{i-1} = \{h_{i-1}^1, \cdots, h_{i-1}^{K_{i-1}}\}$ denotes K_{i-1} feature maps outputted from previous convolution layer and $\mathbf{h}_0$ corresponds to the input image. $g(\cdot)$ represents the nonlinear gating function in the Rectified Linear Unit (ReLU), i.e.,

$$g(x) = max\{0, x\} \tag{2}$$

Followed by the convolution + ReLU layer, a Max pooling operation is applied to each feature map and outputs the max value of the neighboring feature values in each non-overlapped patch. This non-linear down-sampling significantly reduces the computations in the following layer and tries to preserve the maximum filter responses for image representation.

After multiple convolution and pooling operations, two Fully connected layers are placed at the end of the feature extractor. The same convolution operation in (1) is implemented by the fully connected layers, except for enforcing the filter size w_i^k to 1×1. Finally, a $1 \times 1 \times 4096$ feature vector is produced from this deep convolutional neural network.

B. Quality regressor model

Given a feature vector y outputted from the feature extractor, a parameterized regression function $f(\cdot)$ is placed at the end of our Q-DNN model, which is represented by

$$f(y;\theta) = \omega * y + \beta \tag{3}$$

where the parameters of the regressor are $\theta = \{\omega, \beta\}$.

Let z_i denote the label of the ith training image, and $\mathcal{L}(\cdot)$ denote the empirical loss of $f(\cdot)$. The parameter optimization for $f(\cdot)$ can be achieved by minimizing the average loss across all training samples, i.e.,

$$\hat{\theta} = \arg\min \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\big(z_i, f(y_i;\theta)\big) \tag{4}$$

where N is the number of training samples.

As shown in Fig. 1, we employ different layers for training the regressors in two steps of our Q-DNN model. Specifically, the step 1 utilizes the category label in ImageNet, and the Softmax layer is used to obtain the probability of y_i belonging to the category z_i, i.e.,

$$P(z = z_i | y_i, \Theta) = \frac{\exp\big(f(y_i;\theta_i)\big)}{\sum_{j=1}^{C} \exp\big(f(y_i;\theta_j)\big)} \tag{5}$$

where the parameter set for C categories is $\Theta = \{\theta_1, \cdots, \theta_C\}$. Then, the loss function in step 1 is defined as

$$\mathcal{L}_1\big(z_i, f(y_i;\Theta)\big) = -\log\big(P(z = z_i | y_i, \Theta)\big) \tag{6}$$

In step 2, we want to build the correlation between y_i and the MOS value z_i. A Sigmoid layer is firstly applied to y_i, which regularizes the range of the feature value via a non-linear mapping, i.e.,

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{7}$$

Then, a Pdist based loss function is defined to measure the difference between z_i and its prediction, i.e.,

$$\mathcal{L}_2\big(z_i, f(y_i;\theta)\big) = \|z_i - f\big(\sigma(y_i);\theta\big)\|_2 \tag{8}$$

C. Parameter training

It is noted that the parameters of the feature extractor impact the quality prediction by feeding the regressor with y_i, which would change the losses in (6) and (8). Let $\mathbf{w}_i = \{w_i^1, \cdots, w_i^{K_i}, b_i^1, \cdots, b_i^{K_i}\}$ denote the parameters of ith convolution layer, and $\mathbf{W} = \{\mathbf{w}_i, \cdots, \mathbf{w}_L\}$ denote the parameter set collected from the L layers feature extractor. Then, we can rewrite the predicted image quality $\hat{z}_i$ to

$$\hat{z}_i = f(\mathbf{h}_0; \mathbf{W}, \theta) \tag{9}$$

Correspondingly, the optimization target of our Q-DNN model is changed from (4) to

$$\Omega = \arg\min \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\big(z_i, f(\mathbf{h}_0; \mathbf{W}, \theta)\big) \tag{10}$$

where the solving for both $\mathbf{W}$ and θ can be achieved via back-propagation algorithm [11].

III. Experimental Results

We evaluate the performance of the proposed Q-DNN model on two challenging enhanced image databases, including IVCDehazing [5] and DeblurStudy [6]. Each database focuses on different enhancement applications. More specifically, IVCDehazing is composed of 25 hazy images and 200 dazed results produced from 8 dehazing algorithms. DeblurStudy collects 100 real blurred images and 25 synthetic images which are then enhanced by 13 motion deblurring algorithms. For DeblurStudy database, only the deblurred results produced from the real blurred images are used in this experiment, which is consistent with the fact that the reference image is not available in the real-world application. Meanwhile, all images with invalid B-T scores (i.e., *NaN*) [6] are removed from the dataset.

For comparison, nine state-of-the-art BIQA algorithms are involved in our evaluation, which include six opinion-aware methods (i.e., BIQI [12], DIIVINE [13], BLIINDS-II [14], BRISQUE [15], NFERM [16] and TCLT [17]) and three opinion-free metrics (i.e., NIQE [18], IL-NIQE [19] and LPSI [20]). Since the opinion-aware methods require training process with human-rated images, we randomly split each database into non-overlapped subsets for training and testing. Following the criterion in [12], [14], [15], 80% of the unprocessed images and their enhanced versions are used for training, and the remainder images constitute the test set. The random splitting process is repeated 1000 times, and the median performance indexes are reported to eliminate the performance bias as much as possible. In this experiment, the Spearman's rank correlation coefficient (SRCC) and Kendall's rank correlation coefficient (KRCC) between the predicted qualities and human-rated scores are used to evaluate each BIQA metric.

The median SRCC and KRCC results are reported in Table I, where the best results in each column are highlighted by boldface. It is seen that the proposed Q-DNN algorithm, which adaptively updates the feature extractor and quality regressor for different enhancement tasks, significantly outperforms the state-of-the-art BIQA metrics, which are initially developed with hand-crafted features to measure the visual artifacts of distorted images. In the IVCDehazing database, a considerable

TABLE I
THE MEDIAN SRCC AND KRCC PERFORMANCE ACROSS 1000 TRAIN-TEST TRIALS

Method	IVCDehazing		DeblurStudy	
	SRCC	KRCC	SRCC	KRCC
BIQI	0.187	0.137	0.001	0.001
DIIVINE	-0.096	-0.060	-0.001	-0.001
BLIINDS-II	0.336	0.233	-0.079	-0.050
BRISQUE	0.274	0.188	-0.006	-0.005
NFERM	0.228	0.158	0.081	0.055
TCLT	0.204	0.121	-0.002	0.004
NIQE	0.209	0.131	-0.129	-0.087
IL-NIQE	0.134	0.097	-0.115	-0.074
LPSI	0.220	0.139	-0.005	0.004
Q-DNN	**0.548**	**0.401**	**0.148**	**0.100**

TABLE II
COMPARISON OF AVERAGE RUNNING TIME (SECONDS)

Method	IVCDehazing	DeblurStudy
BIQI	2.465	2.841
DIIVINE	30.550	136.813
BLIINDS-II	126.290	330.405
BRISQUE	0.810	1.318
NFERM	115.877	616.698
TCLT	8.155	47.385
NIQE	0.846	2.618
IL-NIQE	25.828	26.764
LPSI	0.288	**0.436**
Q-DNN	**0.225**	2.606

SRCC and KRCC improvements, which are up to 0.212 and 0.168, have been achieved by our Q-DNN model even in comparison with the second best BIQA metric – BLIINDS-II. For the challenging DeblurStudy database, the proposed method is one of the only three BIQA models delivering positive SRCC and KRCC results. Meanwhile, the performance of Q-DNN is also much better than the second and third best metrics including NFERM and BIQI.

In this evaluation, the running time of each BIQA metric is also investigated to roughly estimate their computational complexity. Specifically, the system platform is Intel Core 2 processor of speed 2.0GHz, 6GB RAM and Windows 7 64-bit version. All methods are implemented with the code released by authors and tested on MATLAB2015a software. The average running time for an image is test on both IVCDehazing and DeblurStudy databases and reported in Table II, where the fastest metrics are highlighted by boldface for clarity. It is seen that the proposed Q-DNN model costs the lowest running time in IVCDehazing database. In the DeblurStudy dataset, the proposed method is slower than the fastest LPSI metric, but shows comparable efficiency with respect to the second and fourth fastest metrics including BRISQUE and NIQE. The evaluations reported in Tables I and II confirm that the proposed Q-DNN model could make good trade-off in delivering high quality prediction accuracy and costing low computational complexity.

IV. CONCLUSION

This paper investigates the blind assessment problem on enhanced images. To adapt to the changes of evaluation criterion across different enhancement applications, a data-driven method is proposed based on the quality-aware deep neural network, which is capable of adaptively updating the feature extractor and quality regressor for different training samples. The proposed method is highly adaptable in comparison to the traditional BIQA models, which are dependent on specific hand-crafted features. Experimental results on two newly released enhanced image databases confirm the superiority of the proposed Q-DNN model against nine state-of-the-art BIQA algorithms.

ACKNOWLEDGMENT

This work was supported in part by National Natural Science Foundation of China (No. 61525102, No. 61502084 and No. 61601102).

REFERENCES

[1] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, “Efficient image dehazing with boundary constraint and contextual regularization,” in *IEEE International Conference on Computer Vision*, 2013, pp. 617–624.

[2] J. Pan, Z. Hu, Z. Su, and M. H. Yang, “Deblurring text images via l0-regularized intensity and gradient prior,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 2901–2908.

[3] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, *LIVE Image Quality Assessment Database Release 2*, [Online]. Available: http://live.ece.utexas.edu/research/quality.

[4] F. Meng, H. Li, S. Zhu, B. Luo, C. Huang, B. Zeng, and M. Gabbouj, “Constrained directed graph clustering and segmentation propagation for multiple foregrounds cosegmentation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 11, pp. 1735–1748, Nov 2015.

[5] K. Ma, W. Liu, and Z. Wang, “Perceptual evaluation of single image dehazing algorithms,” in *IEEE International Conference on Image Processing*, Sept 2015, pp. 3600–3604.

[6] W.-S. Lai, J.-B. Huang, Z. Hu, N. Ahuja, and M.-H. Yang, “A comparative study for single image blind deblurring,” in *IEEE Conferene on Computer Vision and Pattern Recognition*, 2016.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[8] L. Kang, P. Ye, Y. Li, and D. Doermann, “Convolutional neural networks for no-reference image quality assessment,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1733–1740.

[9] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2014.

[10] Z. Wang, “Objective image quality assessment: facing the real-world challenges,” in *IS&T Electronic Imaging: Image Quality and System Performance*, San Francisco, CA, Feb. 2016, invited keynote paper.

[11] R. Rojas, *Neural Networks: A Systematic Introduction*. New York, NY, USA: Springer-Verlag New York, Inc., 1996.

[12] A. Moorthy and A. Bovik, “A two-step framework for constructing blind image quality indices,” *IEEE Signal Processing Letter*, vol. 17, no. 5, pp. 513–516, 2010.

[13] A. K. Moorthy and A. C. Bovik, “Blind image quality assessment: From natural scene statistics to perceptual quality,” *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[14] M. Saad, A. C. Bovik, and C. Charrier, “Blind image quality assessment: A natural scene statistics approach in the DCT domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[15] A. Mittal, A. Moorthy, and A. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions Image Processing*, vol. 21, no. 12, pp. 4695–4708, Dec 2012.

[16] K. Gu, G. Zhai, X. Yang, and W. Zhang, “Using free energy principle for blind image quality assessment,” *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 50–63, Jan 2015.

[17] Q. Wu, H. Li, F. Meng, K. N. Ngan, B. Luo, C. Huang, and B. Zeng, “Blind image quality assessment based on multichannel feature fusion and label transfer,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 425–440, March 2016.

[18] A. Mittal, R. Soundararajan, and A. Bovik, “Making a “completely blind” image quality analyzer,” *IEEE Signal Processing Letter*, vol. 20, no. 3, pp. 209–212, March 2013.

[19] L. Zhang, L. Zhang, and A. C. Bovik, “A feature-enriched completely blind image quality evaluator,” *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2579–2591, Aug 2015.

[20] Q. Wu, Z. Wang, and H. Li, “A highly efficient method for blind image quality assessment,” in *IEEE International Conference on Image Processing*, Sept 2015, pp. 339–343.