

Blind Quality Assessment of Camera Images Based on Low-Level and High-Level Statistical Features

Yutao Liu , Ke Gu , Shiqi Wang , Member, IEEE, Debin Zhao , Member, IEEE, and Wen Gao, Fellow, IEEE

Abstract—Camera images in reality are easily affected by various distortions, such as blur, noise, blockiness, and the like, which damage the quality of images. The complexity of distortions in camera images raises significant challenge for precisely predicting their perceptual quality. In this paper, we present an image quality assessment (IQA) approach that aims to solve this challenging problem to some extent. In the proposed method, we first extract the low-level and high-level statistical features, which can capture the quality degradations effectively. On the one hand, the first kind of statistical features are extracted from the locally mean subtracted and contrast normalized coefficients, which denote the low-level features in the early human vision. On the other hand, the recently proposed brain theory and neuroscience, especially the free-energy principle, reveal that the human brain tries to explain its encountered visual scenes through an inner creative model, with which the brain can produce the projection for the image. Then, the quality of perceptions can be reflected by the divergence between the image and its brain projection. Based on this, we extract the second type of features from the brain perception mechanism, which represent the high-level features. The low-level and high-level statistical features can play a complementary role in quality prediction. After feature extraction, we design a neural network to integrate all the features and convert them to the final quality score. Extensive tests performed on two real camera image datasets prove the validity of our method and its advantageous predicting ability over the competitive IQA models.

Index Terms—Image quality assessment (IQA), no-reference (NR)/blind, neural network, free-energy principle, natural image statistics.

Manuscript received December 4, 2017; revised May 5, 2018; accepted June 4, 2018. Date of publication June 25, 2018; date of current version December 20, 2018. This work was supported in part by the Major State Basic Research Development Program of China (973 Program) under Grant 2015CB351804, in part by the National Science Foundation of China under Grant 61703009, and in part by the Young Elite Scientist Sponsorship Program by China Association for Science and Technology under Grant 2017QNRC001, and in part by Young Top-Notch Talents Team Program of Beijing Excellent Talents Funding (2017000026833ZK40). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hantao Liu. (*Corresponding author: Ke Gu*)

Y. Liu and D. Zhao are with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: yt.liu@hit.edu.cn; dbzhao@hit.edu.cn).

K. Gu is with the Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke.doctor@gmail.com).

S. Wang is with the Department of Computer Science, City University of Hong Kong, Kowloon Tong, Hong Kong (e-mail: shiqwang@cityu.edu.hk).

W. Gao is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China, and also with the National Engineering Laboratory for Video Technology and Key Laboratory of Machine Perception, School of Electrical Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: wgao@pku.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMM.2018.2849602

I. INTRODUCTION

WITH the development of image processing applications, there has been an increasing need in developing objective image quality assessment (IQA) approaches which can accurately evaluate the perceptual quality of the image. According to the accessibility of an original or distortion-free image, existing objective IQA methods can be categorized into full-reference (FR) [1]–[3], reduced-reference (RR) [4]–[6] and no-reference (NR) [7]–[9] approaches. In brief, FR methods evaluate the image quality by fully referring to the reference image, RR methods referring to the reference image partially for IQA and NR methods perform quality evaluation without any information of the reference image. NR IQA is also called blind IQA (BIQA).

Till now, the most influential FR method is PSNR, which is pervasively adopted in the image processing systems due to its simplicity and effectiveness. However, PSNR is also found invalid under some conditions, such as the distorted images with similar PSNRs can exhibit different subjective quality [1]. In this regard, Wang *et al.* presented the famous SSIM from a new philosophy which says the human visual system (HVS) tends to extract the structures from the vision so that gauging the structural variation can well predict the image quality degradation. Following this philosophy, some SSIM variants methods were proposed successively and further improved the prediction accuracy of SSIM, e.g., information content weighting-SSIM (IW-SSIM) [10], multi-scale SSIM (MS-SSIM) [11] and structural similarity weighting-SSIM (SW-SSIM) [12], etc. From different perspectives, some other FR approaches were also proposed and reveal state-of-the-art performance for quality evaluation. The visual information fidelity (VIF) method was proposed from the perspective of information theory [2]. Zhang *et al.* employed two low-level features in developing a feature similarity index (FSIM) [13]. A fast reliable image quality predictor by fusing micro-and macro-structures was introduced in [14]. Sparse representation was employed to extract features for quality evaluation in [15]. Visual saliency was also introduced into the IQA algorithm design [3], [16]–[18].

Although FR methods produce the best prediction performance, the condition of FR methods is also rigorous as the original image of the distorted image should be known, which is unpractical in real applications. Therefore, RR IQA was proposed that eases the FR IQA conditions. In RR IQA, partial information is utilized to infer the distorted image quality, which indicates the reference image itself is not required such that the

data rate can be reduced greatly. Some classical RR IQA approaches have been proposed in the past decade. In [19], a RR IQA approach was proposed by modeling the wavelet transform coefficients with the generalized Gaussian distribution (GGD) and the distribution parameters were utilized to measure the image quality. Rehman *et al.* in [20] designed the RR algorithm by differentiating the structural and nonstructural feature variations in the DNT domain. Zhai *et al.* introduced the free-energy principle into IQA design and proposed the RR metric called FEDM [4]. In [21], [22], Liu *et al.* employed sparse representation and further improved the predicting ability of FEDM.

Compared with FR and RR IQA metrics, NR IQA methods are expected to predict the image quality without referring to the original image, which makes NR IQA the most challenging task for IQA researchers. Existing NR IQA methods can be classified into two types, which are distortion-specific and general-purpose methods. Distortion-specific methods mainly deal with one or several particular kinds of distortions. In [23], Li *et al.* proposed a blind image blur evaluation approach based on discrete Tchebichef moments, in which the shape changes of the blurred images are well captured by discrete orthogonal moments for blurriness estimation. Zoran *et al.* proposed an image noise level estimation method as the noisy image violates the scale invariance principle [24]. In [25], Li *et al.* proposed a new quality model for DIBR-synthesized view images, in which a disoccluded region detection method is first proposed using SIFT-flow-based warping and the global sharpness estimation is conducted through a reblurring-based strategy. At last, the overall quality score is derived by pooling the scores of disoccluded regions and global sharpness. A no-reference quality measure was proposed for the compressive sensing recovered images, which extracts the local and global features to characterize the quality degradations. The local features are derived to describe the sharpness and texture variations. The global features are extracted from the statistics of the MSCN and SVD coefficients. Support Vector Regression (SVR) is utilized to regress the features to the final quality [26]. Other distortion-specific works also cover methods that handle blur [27], [28], [29], [30], ringing artifacts [31], JPEG compression [32], etc. Compared with distortion-specific methods, general-purpose NR methods are developed to evaluate the image quality without assuming the distortion types in advance. In [33], Moorthy *et al.* presented BIQI, which has a two-step framework. In the first step, Natural scene statistics (NSS) features were extracted and applied to classify the distorted image to one of some predefined distortions. In the second step, the same set of features were utilized again to evaluate the image quality. Later, BIQI was extended to DIIVINE by introducing richer NSS features [9]. While BRISQUE [8] and BLIINDS-II [34] operated in only one successive step, namely extracted features and then utilized them to predict the image quality. The difference between BRISQUE and BLIINDS-II lies in that BRISQUE extracted the NSS features in the spatial domain and the NSS features of BLIINDS-II come from the DCT domain of the distorted image. In [7], Gu *et al.* proposed NFERM, whose features were designed from three aspects, which are free-energy principle, structural degradation model and image naturalness

measurement. In summary, all the above models conduct in a similar way which contains a training stage and a testing stage. In the training stage, quality-aware features are extracted from a set of impaired images. In the next, a quality module that maps the features onto the assigned subjective scores is learned. In the testing stage, given a distorted image, the same features are extracted and input into the learned regression module to gain a quality score. Since the assigned subjective scores are involved in these methods, they are also called “opinion-aware” methods. On the contrary, the “opinion-unaware” methods don’t need subjective scores for calibrating the quality model. In [35], Mittal *et al.* extracted features from an image and fitted them with a multivariate Gaussian model (MVG), then the quality of the distorted image was derived by the MVG difference from the fitted MVG. Zhang *et al.* extended this method to IL-NIQE by introducing more statistical features, i.e., gradients, colors and Log-gabor filter responses [36].

Although the above mentioned IQA methods have already been verified to work effectively on the well-known image databases, such as LIVE [37], CSIQ [38], TID [39], etc. While to the real photographs or camera images, they may become ineffective even failed on this occasion. The main reason maybe lies in that the distorted images in the popular databases are rather ideal, i.e., the distortion is always artificially generated and the distortion types are also limited. While in real camera images, the distortions are much more complicated and harder to model. Therefore, there remains high research potential in developing new IQA algorithms for the camera images. Toward this end, in this work, we design an effective blind approach for the camera images from two concerns, which are the low-level human vision characteristics and the high-level brain activities. Specifically, on the one hand, we compute the locally normalized luminance coefficients and their products along four orientations based on the observation that these statistics will be altered at the presence of distortion. Therefore, these statistics serve as the low-level statistical features in our algorithm. On the other hand, the brain theory, especially the free-energy principle, manifests the brain intends to explain the encountered scenes through an inner generative model and actively generate corresponding projections for the scenes. Then the quality of perceptions can be measured by how well the brain explains the image, i.e., the agreement between the image itself and its projection in the brain, which inspires us to employ this agreement or discrepancy to extract the corresponding high-level statistical features. With the extracted low-level and high-level features, we design a neural network to combine all the features and map them to the image quality. Experimental results demonstrate the low-level and high-level statistical features can play complementary role in computing the quality of camera images. Additionally, by testing our method on two camera image databases, we verify its effectiveness in quality estimation and superiority over the mainstream IQA approaches.

We organize the rest of this paper as follows: In Section II, the detailed principle of the proposed method for camera image quality evaluation is presented. In Section III, we present the experimental results and deliver necessary analysis. We summarize this paper in Section IV.

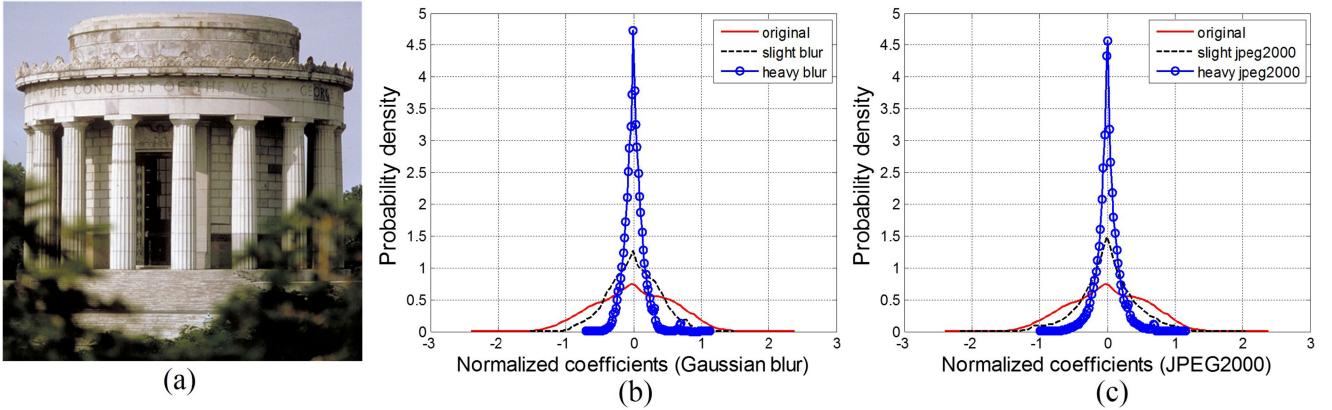


Fig. 1. Histograms of the normalized coefficients (NC) of an original image with its corresponding distorted images. (a) The distortion-free image. (b) The NC histograms of the images with gaussian blur. (c) The NC histograms of the images with JPEG2000.

II. PROPOSED METHOD

A. Low-Level Statistical Features

The first concern in our method is that natural images without distortions will show regularity in statistics, such as the coefficients in the transform domain. While such NSS regularity will be destroyed because of the introduced distortions, which makes the image “unnatural”. This claim has been widely adopted in the design of NR IQA algorithms [8], [34], [35]. Rather than exploring the NSS regularity in the transform domain, in this paper, we extract NSS features in the spatial domain of the image. Specifically, we compute the locally normalized coefficients by subtracting the local mean and then dividing the local standard deviation, which is adopted in [8], [35]. Such operation can de-correlate the image signals effectively and remove their redundant information, which conforms to the behaviors in the early human vision [8]. Given an image I , its locally normalized coefficients I' can be calculated as

$$I'(x, y) = \frac{I(x, y) - \mu(x, y)}{\sigma(x, y)} \quad (1)$$

with $I(x, y)$ and $I'(x, y)$ denoting the original and normalized luminance values at position (x, y) . $\mu(x, y)$ and $\sigma(x, y)$ stands for the mean and standard deviation of a local patch centered at (x, y) , which are respectively calculated by:

$$\mu(x, y) = \sum_{s=-S}^S \sum_{t=-T}^T \omega_{s,t} I(x+s, y+t) \quad (2)$$

$$\sigma(x, y) = \sqrt{\sum_{s=-S}^S \sum_{t=-T}^T \omega_{s,t} [I(x+s, y+t) - \mu(x, y)]^2} \quad (3)$$

where $\omega = \{\omega_{s,t} \mid s = -S, \dots, S; t = -T, \dots, T\}$ denotes a 2D circularly-symmetric Gaussian weighting filter. In Fig. 1, we show an example of the calculated normalized coefficient distributions of an original image with its corresponding distorted versions from CSIQ database [38]. In this figure, (a) refers to the original image, (b) shows the coefficients distribution comparison between the original image and its Gaussian

blurred images. Similarly, (c) gives the distribution comparison between the original image and the corresponding JPEG2000-compressed images. By observing this figure, we can find that no matter Gaussian blur or jp2k compression, the original coefficients distribution are changed to sharper than the original distribution which implies more values become zero as a result. This is due to blur or jp2k eases the structures in the original image and makes it smoother. Therefore, we can conclude that the introduced distortion changes the empirical distribution of the normalized coefficients as the distributions of the distorted images are deviated from that of the original image. In addition, more severe distortion will cause greater changes to the distribution of the coefficients, which can be found from the observation that the distribution of the heavily distorted image deviates more from the original image than the slightly distorted image. To extract effective NSS features from the normalized coefficient distribution for quality estimation, researchers usually fit the distribution with the zero-mean generalized Gaussian distribution (GGD), then the GGD parameters are estimated and serve as the quality-aware features. While the fitting process and parameter estimation will inevitably introduce error in feature extraction. Instead, here we extract features directly from the coefficients distribution. Specifically, take another look at Fig. 1(b) and (c), we find the distribution variation of the original image and distorted images mainly concentrate among the interval of $[-2, 2]$ at the horizontal axis. Therefore, we directly cut out the part of the histogram of $[-2, 2]$ as our features. Specifically, we divide $(-2, 2)$ into 20 bins at the step of 0.2 and utilize the number of elements in each bin resulting in a vector of 20 elements.

Furthermore, as stated in [8] and [35], the distribution of the products of pairs of adjacent normalized coefficients can also be disturbed because of the distortion. Here we calculate the adjacent products along four directions, which are vertical, horizontal, main-diagonal and second-diagonal, namely the products of $I'(x, y)I'(x+1, y)$, $I'(x, y)I'(x, y+1)$, $I'(x, y)I'(x+1, y+1)$ and $I'(x, y)I'(x+1, y-1)$ respectively. Then we have four distributions of the four kinds of products. Like feature extraction of the normalized coefficients, we also take part of the histogram of these four distributions and get four 20-dimension

vectors. By considering the color information in quality evaluation, we perform NSS feature extraction in R, G, B channels respectively. Then the features in these three channels are concatenated into the final feature vector, which serves as the low-level NSS feature vector for quality evaluation.

B. Free-Energy Principle-Induced High-Level Statistical Features

Besides concerning the early human vision characteristics for quality evaluation, the high-level brain activities for visual perception should also be taken into consideration. Considering this, our second category of features for quality estimation are derived from the free-energy principle, which attempts to account for the human brain activities when exposed to the outside world [40], [41]. A basic precondition of the free-energy principle lies in that the human perception is controlled by an generative model in brain, with which the human brain can actively produce the prediction of the visual scenes while avoid the uncertainty maximally in the meantime. In practice, the brain generative model that manipulates visual perception is often modeled as parametric which can explain the outside world through adjusting its parameters, denoted by \mathcal{M} . For convenience, we utilize \mathbf{m} as the vector which is composed of the parameters in \mathcal{M} . Then, the ‘surprise’ of an input image I can be calculated by integrating the joint distribution $P(I, \mathbf{m})$ over the parameter space of \mathbf{m} :

$$-\log P(I) = -\log \int P(I, \mathbf{m}) d\mathbf{m} \quad (4)$$

we introduce an assistant item $\tilde{P}(\mathbf{m}|I)$ into the right component of the above equation while still maintain the equivalence as:

$$-\log P(I) = -\log \int \tilde{P}(\mathbf{m}|I) \frac{P(I, \mathbf{m})}{\tilde{P}(\mathbf{m}|I)} d\mathbf{m} \quad (5)$$

here $\tilde{P}(\mathbf{m}|I)$ can be thought of as the posterior distribution of image I , which is the approximate posterior distribution to the true posterior distribution $P(\mathbf{m}|I)$ when exposed to I . For explaining I , the human brain tries to reduce the divergence between the approximate $\tilde{P}(\mathbf{m}|I)$ and the true $P(\mathbf{m}|I)$ as much as possible. Based on Jensen’s inequality, the above equation can be written as:

$$-\log P(I) \leq -\int \tilde{P}(\mathbf{m}|I) \log \frac{P(I, \mathbf{m})}{\tilde{P}(\mathbf{m}|I)} d\mathbf{m} \quad (6)$$

On the basis of statistical physics and thermodynamics [42], we define the right part of the above equation as “free energy”, namely:

$$F(\mathbf{m}) = -\int \tilde{P}(\mathbf{m}|I) \log \frac{P(I, \mathbf{m})}{\tilde{P}(\mathbf{m}|I)} d\mathbf{m} \quad (7)$$

Obviously, $F(\mathbf{m})$ denotes an upper bound of I ’s ‘surprise’. We can explain this through further derivation. As

$P(I, \mathbf{m}) = P(\mathbf{m}|I)P(I)$, equation (7) can be rewritten as:

$$\begin{aligned} F(\mathbf{m}) &= \int \tilde{P}(\mathbf{m}|I) \log \frac{\tilde{P}(\mathbf{m}|I)}{P(\mathbf{m}|I)P(I)} d\mathbf{m} \\ &= -\log P(I) + \int \tilde{P}(\mathbf{m}|I) \log \frac{\tilde{P}(\mathbf{m}|I)}{P(\mathbf{m}|I)} d\mathbf{m} \\ &= -\log P(I) + \mathbf{KL}(\tilde{P}(\mathbf{m}|I) \| P(\mathbf{m}|I)) \end{aligned} \quad (8)$$

where $\mathbf{KL}(\cdot)$ denotes the Kullback-Leibler divergence, which is nonnegative. It is easily found that the free energy of $F(\mathbf{m})$ is greater than or equal to $-\log P(I)$ which accounts for the ‘surprise’. When perceiving image I , the human brain intends to minimize the divergence of $\mathbf{KL}(\tilde{P}(\mathbf{m}|I) \| P(\mathbf{m}|I))$.

To apply the free-energy principle in IQA task, we need to determine the internal model at first. However, the difficulty is that real configuration of the internal model is beyond our knowledge [4]. Toward this end, IQA researchers resorted to existing image models to simulate the internal generative model. In this paper, we employ the linear auto-regressive (AR) model to approximate the internal generative model due to its flexibility to represent natural images, described as:

$$x_n = \mathcal{S}^k(x_n)\mathbf{c} + e_n \quad (9)$$

where x_n is the pixel in question, $\mathcal{S}^k(x_n)$ contains k nearest neighbor pixels of x_n , $\mathbf{c} = (c_1, c_2, \dots, c_k)^T$ represents k coefficients to compute x_n , “T” refers to the transpose operation. e_n gives the representation error. To calculate the coefficients \mathbf{c} , the following optimization equation is needed to be solved as:

$$\mathbf{c}^* = \underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{Sc}\|_2 \quad (10)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_k)^T$ and $\mathbf{S}(i, :) = \mathcal{S}^k(x_i)$. Obviously, the solution of this equation is $\mathbf{c}^* = (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T \mathbf{x}$. Here \mathbf{c}^* represents the parameter vector \mathbf{m} of the internal generative model. Then x_n can be represented by $\mathcal{S}^k(x_n)\mathbf{c}^*$ and the brain prediction of image I can be calculated in this point-wise manner. In this regard, the posterior distribution $\tilde{P}(\mathbf{m}|I)$ is approximated by all the AR parameters’ distribution. In Fig. 3, we show an example of a natural image and the posterior distribution $\tilde{P}(\mathbf{m}|I)$ to illustrate. It can be found that $\tilde{P}(\mathbf{m}|I)$ reveals a very sharp appearance which indicates most values are near zero.

As the free-energy principle conjectures although the brain tries its best to explain the input image, the divergence between the image and the brain explanation still exists. This is due to the generative model in the brain isn’t universal, otherwise nothing will be surprising or interesting to the brain. Then such divergence should be closely associated with the perception quality, which thereby can be utilized for quality estimation of the image. Inspired by this, we define the prediction divergence by the prediction residual R , calculated as:

$$R = I - I_p \quad (11)$$

where I_p represents the brain explanation of I , which is obtained by AR prediction in equation (9). “-” refers to the operation of element-wise subtraction. To illustrate that the prediction residual R can characterize the image quality levels, we give an

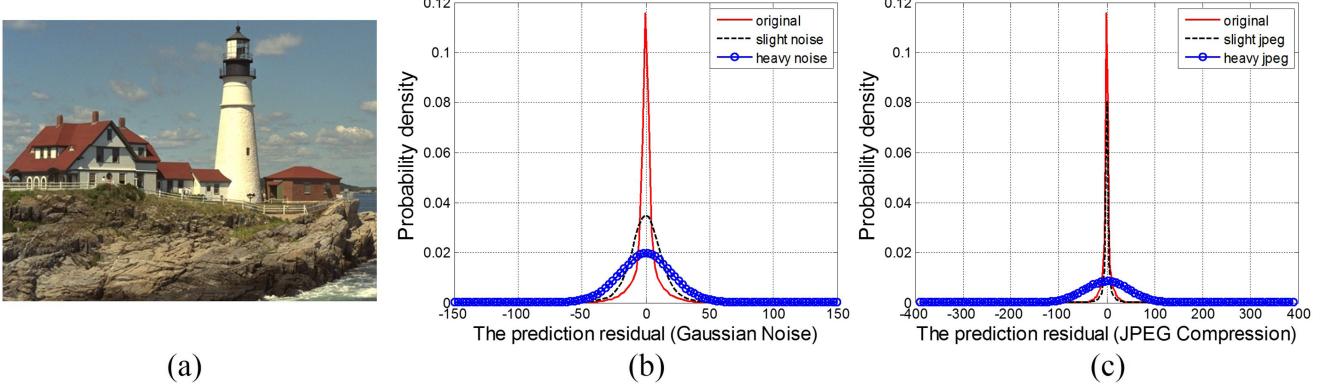


Fig. 2. Histograms of the prediction residual R_s of an original image with its corresponding distorted images. (a) The distortion-free image. (b) The histograms of R_s of the images with gaussian noise. (c) The histograms of R_s of the images with JPEG compression.

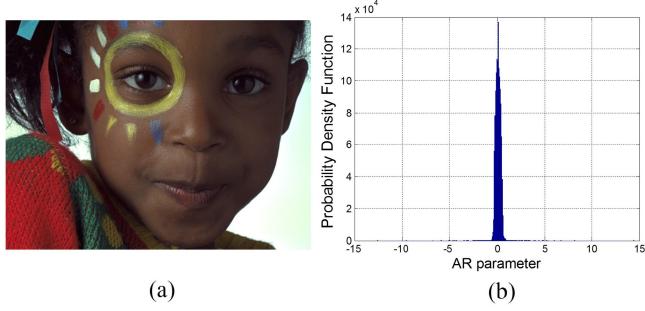


Fig. 3. Illustration of the posterior distribution of $\tilde{P}(\mathbf{m}|I)$. (a) A natural image. (b) The corresponding $\tilde{P}(\mathbf{m}|I)$ of (a) in terms of AR parameters.

example of R_s s of an original image and its distorted versions with different distortion levels in Fig. 2, where (a) is the original image, (b) shows the histogram comparison of R_s s of the original image and the gaussian noise images, (c) shows the histogram comparison of R_s s of the original image and the JPEG compression images. We investigate two distortion levels which are slight distortion and heavy distortion. From this figure, we can see that the Gaussian noise and jpeg compression both make the original residual distribution smoother with longer tails. This is because the Gaussian noise and jpeg compression will introduce more high-frequency components, e.g., the blocking artifacts, to the original image. Such components are more difficult for AR representation, which leads to higher prediction residual. Therefore, the brain prediction residual is able to distinguish the distortion-free and distorted image. Likewise, different distortion levels can also be captured by the deviation degree of the residual distribution as observed in Fig. 2. Based on this, we can extract quality-aware features accordingly.

Similar as the NSS features extraction, we divide the interval $[-50, 50]$ into 100 bins at the step of 1 and take the number of elements falling into each bin leading to a 100-dimension vector, which stands for the high-level feature vector for quality calibration. Considering that multi-scale strategy is effective in objective IQA methods [11], we extract all the features (low-level NSS features and high-level free-energy-induced features) in five resolutions respectively. The operations we take are that

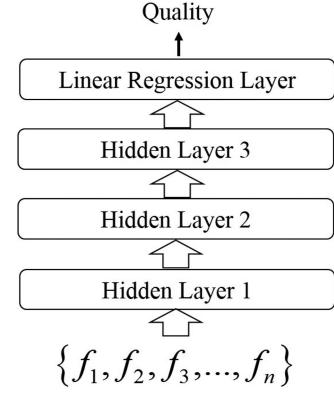


Fig. 4. The structure of our neural network for quality evaluation.

we apply a low-pass filter and then down-sample the filtered image by the factor of 2. These operations are conducted four times iteratively. At last, the features extracted in each scale are concatenated into the final feature vector.

C. Neural Network-Based Quality Assessment

As aforementioned, learning-based NR IQA methods usually learn a regression module which maps the quality-aware features to the quality level after feature extraction. In this process, SVR is mostly employed in the existing algorithms for its convenience and effectiveness. While SVR is also proposed that it can't approximate the complicated HVS optimally because of its shallow architecture [43], [44]. Instead, in this work, we design a neural network to convert the quality-aware features into the final quality score. Experiments in Section III-F will verify the designed neural network is more effective than SVR in quality prediction. We show the structure of our neural network in Fig. 4. It can be observed that the designed neural network includes three hidden layers and one linear regression layer. The hidden layers attempt to dig potential data information closely related to visual quality of the image. As the image quality score is represented by a continuous value, we adopt the linear regression layer to regress the output of the last hidden layer to the final image quality score, which represents the image quality. The bottom input of the neural network is the features we

extracted, denoted by $f_1, f_2, f_3, f_4, \dots, f_n$ ($n = 2000$) and the linear regression layer output the quality score of the image. The sizes of the three hidden layers are 200, 40, 6 respectively. Specifically, we train our neural network in three steps. The first step is an unsupervised pre-training, in which each hidden layer is treated as a sparse auto-encoder. The L-BFGS algorithm [45] is employed to train each sparse auto-encoder. The maximum iteration of L-BFGS is 1000. The sigmoid function is applied as the activation function in the hidden layers. The cost function for training each hidden layer is:

$$J_{\text{sparse}}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} KL(\rho \parallel \hat{\rho}_j) \quad (12)$$

where

$$\begin{aligned} J(W, b) &= \frac{1}{m} \sum_{j=1}^m J(W, b; x^{(i)}, y^{(i)}) + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \\ &= \frac{1}{m} \sum_{j=1}^m \frac{1}{2} \|h_{W,b}(x^{(i)}) - y^{(i)}\|^2 + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \end{aligned} \quad (13)$$

and

$$KL(\rho \parallel \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \quad (14)$$

where W is the network weight, b refers to the hidden layer bias, $h_{W,b}(\cdot)$ represents the output of each nerve cell in the hidden layer. ρ and $\hat{\rho}_j$ represent the mean activation and expected mean activation respectively. In this work, ρ is set to be 0.1, β is set to be 3, λ refers to the weight decay parameter, which is set to be 0.0001. After training each sparse auto-encoder, the linear regression layer is trained through the cost function as:

$$J(W, b) = \frac{1}{m} \sum_{i=1}^m \frac{1}{2} (Y - \text{Label})^2 + \frac{\lambda}{2} \sum_i W_i^2 \quad (15)$$

where Y means the output of the regression layer and Label denotes the subjective quality scores of the training images, which stand for the ground truth as our training goals. At last, we fine tune the whole network, the weights are initialized with the pre-trained values. Now the trained network can be utilized for quality estimation.

III. EXPERIMENTAL RESULTS

A. Prediction Performance Evaluation Indexes

In our experiments, to evaluate the performance of the objective IQA approaches, we employ four statistical indexes, which are to compute the correlation between subjective ratings and objective scores given by the IQA methods. They are Spearman Rank order Correlation coefficient (SRCC), Kendall's rank correlation coefficient (KRCC), Pearson's linear correlation coefficient (PLCC) and root mean square error (RMSE) respectively. In general, a better objective method is expected to achieve higher SRCC, KRCC and PLCC values, while lower RMSE value. Specifically, before calculating PLCC and RMSE, objective scores given by objective IQA methods are often mapped

to subjective scores by nonlinear regression [46]. Hence, we employ a logistic function with five parameters for implementation:

$$q(z) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2 \cdot (z - \beta_3))} \right) + \beta_4 \cdot z + \beta_5 \quad (16)$$

where z and $q(z)$ are the objective score and the mapped score. $\beta_1, \beta_2, \dots, \beta_5$ are parameters to be fitted through curve fitting.

B. Testing Image Databases

In this work, two real camera image databases are employed as the test bed for examining the performance of the proposed method. The first one is the CID2013 database [47], which was specifically proposed for evaluating NR IQA algorithms. The CID2013 database includes 474 naturally-distorted images captured by 79 different cameras or image signal processing pipelines. The included images can be classified into six sets (I–VI), each set includes six different scenes and every scene was captured by 12–14 different cameras. The second image database for testing our method is the realistic blur image database (BID) [48], which includes a total of 586 blurred images, which were photographed under different exposure time and lighting situations. The distorted images in BID can be classified into five blur cases, i.e., Unblurred, Simple motion blur, Complex motion blur, Out-of-focus blur and their combination. We give some example images from these two databases in Fig. 5, where the group of (a) is from CID2013 database and group of (b) is from BID database.

C. Overall Prediction Performance Comparison

With the two testing image databases, we randomly divide each database into nine subsets and adopt the nine-fold leave-one-out cross validation methodology to evaluate the prediction performance of our approach. Specifically, in each procedure, eight of the subsets are used for training and the remaining one subset is for testing. This process repeats nine times. The average performance evaluated by SRCC, KRCC, PLCC and RMSE are summarized in Table I and Table II. Table I lists the performance results on CID2013 database and Table II gives the performance results on BID database. We highlight the best performed method with boldface. As observed in Table I, we compare our approach with ten representative approaches, which are BIQI [33], BRISQUE [8], DIIVINE [9], NFERM [7], SISBLIM [49], LPSI [50], NIQE [35], ILNIQE [36], QAC [51], BIQME [52] and FRIQUEE [53] respectively. It is noted that all the compared methods belong to NR methods as there are no original images for the distorted images to reference so that the FR and RR methods can't be applied. Furthermore, we can divide the NR methods into two categories. The one is the training-based category, which contains BIQI, BRISQUE, DIIVINE, NFERM and FRIQUEE. The other one category is training-free, which includes SISBLIM, LPSI, NIQE, ILNIQE, QAC and BIQME. By observing Table I, we can have two meaningful findings. First, the methods which perform well on artificially-generated databases become ineffective or invalid on the real camera image database, such as BRISQUE, LPSI

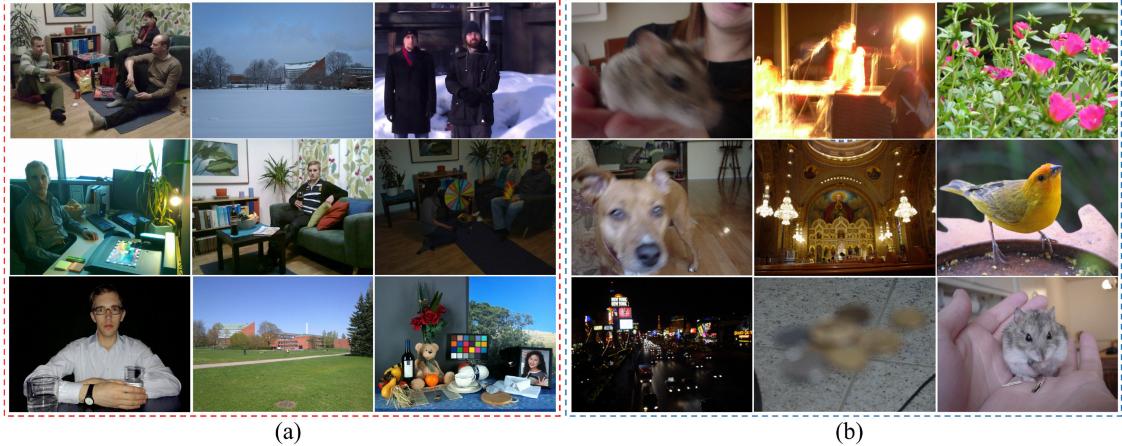


Fig. 5. Example distorted images from CID2013 and BID databases. (a) Images from CID2013 database. (b) Images from BID database.

TABLE I
OVERALL PREDICTION PERFORMANCE COMPARISON ON CID2013 DATABASE

Methods	SRCC	KRCC	PLCC	RMSE
BIQI [33]	0.7176	0.5259	0.7259	15.5711
BRISQUE [8]	0.4419	0.3052	0.4806	19.8548
DIIVINE [9]	0.4633	0.3159	0.5348	19.1310
NFERM [7]	0.6205	0.4493	0.6439	17.3217
SISBLIM [49]	0.6554	0.4780	0.7038	16.0826
LPSI [50]	0.3230	0.2167	0.4524	20.1914
NIQE [35]	0.6539	0.4630	0.6660	16.8891
ILNIQE [36]	0.3063	0.2099	0.4274	20.4688
QAC [51]	0.0299	0.0178	0.1739	22.2956
BIQME [52]	0.2418	0.1596	0.4777	19.8895
FRIQUEE [53]	0.5580	0.3889	0.5956	18.1872
Proposed	0.7830	0.6051	0.8005	12.6541

The best performer is in boldface.

and QAC etc. This reveals precise quality assessment for real camera images or photographs is still very challenging and reserves great potential for study. Second, our proposed method achieves notably better prediction performance than the other competing methods, which verifies our method can perform more effectively than the competitors. It is also noted that although FRIQUEE is designed for authentically distorted images, it still fails to attain desired results for quality prediction.

As to the results on BID database shown in Table II, besides comparing our method with the representative NR models, we also include the specific blurriness assessment methods for comparison as the BID database is a realistic blur image database. Those mainstream blurriness assessment methods are CPBD [54], ARISMC [55], FISH [56], GPSQ [57], JNB [58], LPC [59], S3 [60] and BIBLE [23]. From this table, similar conclusions can be drawn that the general-purpose NR methods cannot deliver desired results. In addition, although the sharpness methods are specialized for the blurred images, they are also unable to produce high performance for the real blurred im-

TABLE II
OVERALL PREDICTION PERFORMANCE COMPARISON ON BID DATABASE

Methods	SRCC	KRCC	PLCC	RMSE
BIQI [33]	0.4024	0.2730	0.4420	1.1230
BRISQUE [8]	0.0977	0.0629	0.2506	1.2120
DIIVINE [9]	0.3427	0.2300	0.3310	1.1814
NFERM [7]	0.4679	0.3182	0.4738	1.1025
SISBLIM [49]	0.2424	0.1588	0.3529	1.1714
LPSI [50]	0.0428	0.0315	0.1754	1.2325
NIQE [35]	0.4587	0.3091	0.4608	1.1111
ILNIQE [36]	0.4937	0.3412	0.5070	1.0791
QAC [51]	0.2995	0.2028	0.3237	1.1845
BIQME [52]	0.2360	0.1622	0.2297	1.2184
FRIQUEE [53]	0.4492	0.3009	0.4700	1.1050
CPBD [54]	0.0184	0.0130	0.2704	1.2053
ARISMC [55]	0.0200	0.0130	0.1314	1.2411
FISH [56]	0.4835	0.3315	0.4799	1.0983
GPSQ [57]	0.5099	0.3473	0.5315	1.0605
JNB [58]	0.1409	0.0982	0.2605	1.2087
LPC [59]	0.3161	0.2161	0.3901	1.1528
S3 [60]	0.4125	0.2805	0.4270	1.1320
BIBLE [23]	0.3606	0.2436	0.3842	1.1559
Proposed	0.6276	0.4496	0.6502	1.0126

The best performer is in boldface.

ages. Among them, GPSQ performs the best since it is designed for the real blurred images. It is worthy noting that our proposed method still outperforms all the other methods significantly on the BID database.

To statistically verify the superiority of our proposed method, we further inspect the statistical significance of the results given by IQA models through t-test. The t-test is conducted by computing the prediction residuals between the subjective ratings and the converted objective scores. We give the experimental

TABLE III
STATISTICAL SIGNIFICANCE RESULTS (T-TEST)

t-test	BIQI	BRISQUE	DIIVINE	NFERM	SISBLIM	LPSI	NIQE	ILNIQE	QAC	BIQME	FRIQUEE
CID2013	0	1	1	1	1	1	1	1	1	1	1
BID	1	1	1	1	1	1	1	1	1	1	1

1, 0, or -1 indicates our approach is statistically superior, indistinguishable, or inferior to the general-purpose NR method (with 95% confidence).

TABLE IV
STATISTICAL SIGNIFICANCE RESULTS (T-TEST)

t-test	CPBD	ARISMIC	FISH	GPSQ	JNB	LPC	S3	BIBLE
BID	1	1	1	1	1	1	1	1

1, 0, or -1 indicates our method is statistically superior, indistinguishable, or inferior to the sharpness assessment method (with 95% confidence).

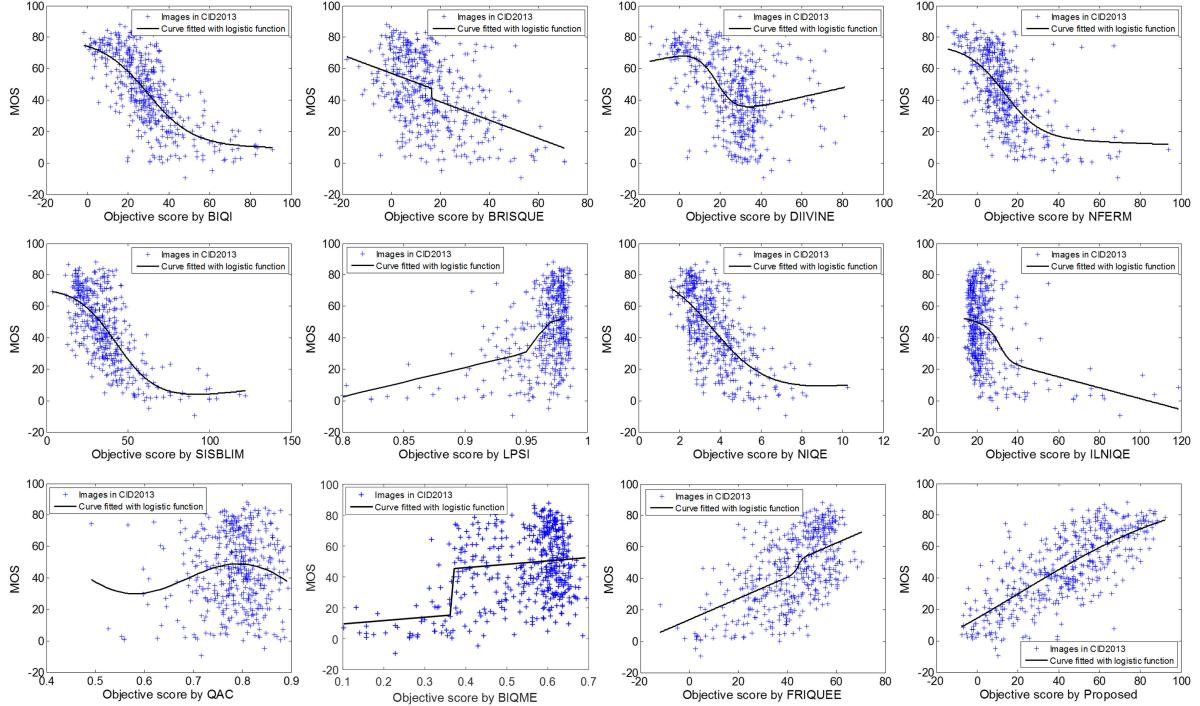


Fig. 6. The distribution diagrams of MOS values with respect to objective values on CID2013 dataset.

TABLE V
PERFORMANCE COMPARISON ON CID2013 DATABASE WHEN TRAINED ON BID DATABASE

Methods	SRCC	KRCC	PLCC	RMSE
BIQI [33]	0.2744	0.2223	0.3612	21.1119
BRISQUE [8]	0.3665	0.2472	0.4129	20.6206
DIIVINE [9]	0.1936	0.1359	0.4484	20.2363
NFERM [7]	0.2588	0.1702	0.5626	18.7172
FRIQUEE [53]	0.1068	0.0698	0.2300	22.0332
Proposed	0.2656	0.1938	0.3350	21.3318

The best performer is in boldface.

TABLE VI
PERFORMANCE COMPARISON ON BID DATABASE WHEN TRAINED ON CID2013 DATABASE

Methods	SRCC	KRCC	PLCC	RMSE
BIQI [33]	0.2636	0.2108	0.2770	1.2029
BRISQUE [8]	0.0041	0.0026	0.1404	1.2395
DIIVINE [9]	0.3863	0.2608	0.4096	1.1421
NFERM [7]	0.2723	0.1765	0.2980	1.1950
FRIQUEE [53]	0.0559	0.0351	0.2175	1.2220
Proposed	0.3087	0.2227	0.2553	1.2167

The best performer is in boldface.

results on CID2013 and BID in Table III and Table IV, where 1, 0 and -1 respectively indicates our approach is superior, indistinguishable and inferior to competing approaches statistically

in each column (with 95% confidence). From these two tables, we can find that only BIQI on CID2013 database is statistically indistinguishable from our proposed method. While in other

TABLE VII
PERFORMANCE COMPARISON ACCORDING TO SRCC, KRCC, PLCC AND RMSE VALUES OF THE TWO GROUPS OF FEATURES
ON CID2013 AND BID DATABASES

Feature Type	CID2013				BID			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
NSS	0.6109	0.4545	0.6555	15.9242	0.5365	0.3743	0.5496	1.0287
Free-energy	0.6831	0.5120	0.7312	14.6970	0.3710	0.2507	0.3913	1.1316
Combination	0.7830	0.6051	0.8005	12.6541	0.6276	0.4496	0.6502	1.0126

We highlight the best performance with boldface.

TABLE VIII
PERFORMANCE COMPARISON ACCORDING TO SRCC, KRCC, PLCC AND RMSE VALUES BETWEEN NEURAL NETWORK AND SVR

Model Type	CID2013				BID			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
SVR	0.6159	0.4510	0.6550	16.2794	0.4920	0.3437	0.5304	1.0364
Neural Network	0.7830	0.6051	0.8005	12.6541	0.6276	0.4496	0.6502	1.0126

We highlight the best performance with boldface.

cases, our method is superior to all of the compared models statistically.

For visualization, we also provide the distribution diagrams of subjective MOS values with respect to objective values on CID2013 database in Fig. 6, in which we denote the distorted images with blue “+” and the black curves are obtained in the curve fitting process in (16). One can see the blue “+” of our method gather evenly and close to the black curve and the curve also exhibits almost a straight line, which manifest the better correlation between the scores given by our method and the subjective judgements for the image quality.

D. Cross-Dataset Performance Comparison

In previous subsection, we compared the performance of the objective methods in each dataset as we partition each dataset into training dataset and testing dataset. In this subsection, we would like to test the prediction performance by cross-dataset validation, in which we train the objective method on one dataset and test it on the other dataset. In our compared methods, BIQI, BRISQUE, DIIVINE, NFERM and FRIQUEE belong to the learning-based methods, therefore, we compare their performance with ours through cross-dataset validation. The performance comparison is given in Table V and Table VI. Among them, Table V lists the performance comparison on CID2013 database with the methods trained on BID database, Table VI lists the results on BID database with the methods trained on CID2013 database.

It is observed that the performance of the learning-based methods all degrade significantly in Table V and Table VI. This is because the distortions in these two databases are much different. Therefore, the quality model trained on one database can't cover the distortions of the other database so that the prediction results are likely inaccurate. Such results also reveal the common disadvantage of the learning-based methods, namely the generalization capability of the learning-based methods is

TABLE IX
PERFORMANCE OF THE PROPOSED METHOD W.R.T. VARIATIONS OF THE STEP VALUES IN FEATURE EXTRACTION ON CID2013 DATABASE

v_1	SRCC	v_2	SRCC
0.1	0.7822	1	0.7830
0.2	0.7830	2	0.7829
0.3	0.7829	3	0.7825
0.4	0.7827	4	0.7826
0.5	0.7818	5	0.7806

relatively weak [36]. As observed in these two tables, our proposed method earns moderate results among the compared methods. These results once again manifest there remains much study room for precise and robust quality assessment for the real camera images.

E. Analysis of the Features' Contribution in Our Proposed Method

In our proposed method, we design two kinds of statistical features for characterizing the quality degradations, which are the low-level NSS features and the free-energy induced high-level features. We are interested in the contribution of each type of the features for final quality assessment. Therefore, in this experiment, we separately use the NSS features, the free-energy induced features and their combination for quality prediction. Other configurations in the proposed method are all fixed and the test methodology is the same as in Section III-C. The experimental results on CID2013 and BID databases are listed in Table VII, where we bold the best results. It is observed that on CID2013 database, the free-energy induced features can be more effective than the NSS features as the performance of the former is above the latter. On the contrary, the NSS features can lead to better prediction performance on the BID database. These observations are due to that CID2013 database contains compli-

TABLE X
PERFORMANCE OF THE PROPOSED METHOD W.R.T. VARIATIONS OF THE HIDDEN LAYER SIZE IN NEURAL NETWORK
ON CID2013 DATABASE

s_1	SRCC	s_2	SRCC	s_3	SRCC
100	0.7829	30	0.7818	4	0.7829
200	0.7830	40	0.7830	5	0.7829
300	0.7818	50	0.7791	6	0.7830
400	0.7793	60	0.7783	7	0.7813
500	0.7767	70	0.7781	8	0.7814

cated distorted images, which may induce more thinking or brain activities for viewers to perform quality assessment. Therefore, the high-level features from the brain activities play the leading role in quality determination on the CID2013 database. While to the BID database, the images are mainly degraded by the distortion of blur, which makes it much easier for humans to judge the image quality. Therefore, the humans tend to resort to more low-level vision characteristics for quality evaluation, which accounts for the low-level NSS features are more effective than the high-level features on the BID database. These experimental results can reflect that our proposed method is consistent with human perceptions for the image quality. It is worth noting that the combination of the NSS features and the free-energy induced features outperforms any one of them on both databases, which proves that the low-level and high-level features can work cooperatively in evaluating the image quality so that the combination of them can boost the final quality evaluation performance accordingly.

F. Prediction Performance Comparison Between SVR and Neural Network

To map the quality-aware feature vector onto the final quality level, we design a neural network to implement this goal. In this section, we want to compare the proposed neural network with the traditional SVR for quality prediction through necessary experiments. In experiments, we fed the extracted features into the neural network and SVR respectively and get the corresponding quality score. We report the testing results on CID2013 and BID databases with Table VIII, where we highlight the best performance in boldface. From this table, we can clearly see that all the statistical index values of the designed neural network are notably better than that of SVR, which proves the superiority of our designed neural network over SVR in approximating the process of converting the quality-aware features into quality level in human brain.

G. Sensitivity to Parameter Variations

In this subsection, we discuss about the sensitivity of the proposed method to the parameter variations. Here, we investigate the parameters of the step value in the stage of feature extraction, the approximation method of the internal model and the sizes of the three hidden layers in the neural network. In implementation, we set the parameter values empirically, but the performance of the proposed method is robust to the parameters within a moderate range. We conducted experiments

on the CID2013 database by varying the parameter values as follows: the step value in the low-level feature extraction is from 0.1 to 0.5 with the step size being 0.1, the step value in the high-level feature extraction is from 1 to 5 with the step size being 1, the approximation method is from AR model to sparse representation adopted in [21], the size of the first hidden layer is from 100 to 500 with the step size being 100, the size of the second hidden layer is from 30 to 70 with the step size being 10, the size of the third hidden layer is from 4 to 8 with the step size being 1. In our experiments, when we vary one parameter, we fix other parameters as their default values. The test strategy is the same as subSection III-C. Here, we denote the step value in the low-level feature extraction as v_1 , the step value in the high-level feature extraction as v_2 , the size of the first hidden layer as s_1 , the size of the second hidden layer as s_2 , the size of the third hidden layer as s_3 . The experimental results in terms of SRCC are reported in Table IX and Table X, respectively. From Table IX, we observe that the performance of the proposed method changes slightly as the step values vary, which indicates the proposed method is robust to parameter variations. Similar conclusion can be drawn from Table X. From these results, we can conclude the proposed method is insensitive to the parameter variations within a moderate range.

For approximating the internal model in high-level feature extraction, we compare the adopted AR model and sparse representation model [21], the configuration of sparse representation follows [21] strictly. The SRCC values of AR and sparse representation on CID2013 are 0.7830 and 0.8069 respectively, indicating sparse representation is more effective than AR in approximating the internal model, which coincides with the results in [21]. Therefore, we may resort to sparse representation for simulating the internal model in future work.

IV. CONCLUSION

In this work, a blind image quality evaluator has been introduced which aims to accommodate the camera images in reality. In the proposed method, we extracted the quality-aware features from the low-level human vision characteristics and the high-level brain activities in free-energy principle respectively. On one hand, the low-level features can well characterize the image quality degradations. On the other hand, the brain theory, particularly the free-energy theory reveals that the divergence between the visual scene and the generation by the brain can denote the quality of perceptions, which inspires us to design the high-level quality-aware features. After feature extraction, we designed a neural network that integrating all the features and converting

them into the final quality. Through extensive experiments, we verify that our proposed method can deliver superior prediction performance over state-of-the-art quality models. What's more, the rationalities of the components in the proposed method have also been proved with reliable experiments.

REFERENCES

- [1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [2] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [3] L. Zhang, Y. Shen, and H. Li, "Vsi: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014.
- [4] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41–52, Jan. 2012.
- [5] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 2012.
- [6] 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.
- [7] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.
- [8] 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.
- [9] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
- [10] 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.
- [11] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE 37th Asilomar Conf. Signals, Syst., Comput.*, 2003, vol. 2, pp. 1398–1402.
- [12] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, "Structural similarity weighting for image quality assessment," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2013, pp. 1–6.
- [13] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
- [14] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro-and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, May 2017.
- [15] L. Li *et al.*, "Sparse representation-based image quality index with adaptive sub-dictionaries," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3775–3786, Aug. 2016.
- [16] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.
- [17] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 7, pp. 971–982, Jul. 2011.
- [18] H. Liu and I. Heynderickx, "Studying the added value of visual attention in objective image quality metrics based on eye movement data," in *Proc. 16th IEEE Int. Conf. Image Process.*, 2009, pp. 3097–3100.
- [19] Z. Wang *et al.*, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, Jun. 2006.
- [20] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012.
- [21] Y. Liu *et al.*, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379–391, Feb. 2018.
- [22] Y. Liu, G. Zhai, X. Liu, and D. Zhao, "Perceptual image quality assessment combining free-energy principle and sparse representation," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2016, pp. 1586–1589.
- [23] L. Li *et al.*, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybernetics*, vol. 46, no. 1, pp. 39–50, Jan. 2016.
- [24] D. Zoran and Y. Weiss, "Scale invariance and noise in natural images," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 2209–2216.
- [25] L. Li, Y. Zhou, K. Gu, W. Lin, and S. Wang, "Quality assessment of dibr-synthesized images by measuring local geometric distortions and global sharpness," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 914–926, Apr. 2018.
- [26] B. Hu *et al.*, "No-reference quality assessment of compressive sensing image recovery," *Signal Process., Image Commun.*, vol. 58, pp. 165–174, 2017.
- [27] L. Li *et al.*, "Image sharpness assessment by sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1085–1097, Jun. 2016.
- [28] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. IEEE Int. Conf. Image Process.*, 2002, pp. 57–60.
- [29] L. Li *et al.*, "No-reference quality assessment of deblurred images based on natural scene statistics," *IEEE Access*, vol. 5, pp. 2163–2171, Jan. 2017.
- [30] L. Li *et al.*, "Perceptual quality evaluation for image defocus deblurring," *Signal Process., Image Commun.*, vol. 48, pp. 81–91, 2016.
- [31] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 529–539, Apr. 2010.
- [32] H. Liu and I. Heynderickx, "A perceptually relevant no-reference blockiness metric based on local image characteristics," *EURASIP J. Adv. Signal Process.*, vol. 2009, Art. no. 263540, 2009.
- [33] 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.
- [34] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
- [35] 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.
- [36] 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.
- [37] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," 2006. [Online]. Available: <http://live.ece.utexas.edu/research/quality>
- [38] E. C. Larson and D. Chandler, "Categorical image quality (CSIQ) database," 2010. [Online]. Available: <http://vision.okstate.edu/csiq>
- [39] N. Ponomarenko *et al.*, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. 4th Eur. Workshop Vis. Inf. Process.*, 2013, pp. 106–111.
- [40] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *J. Physiol. Paris*, vol. 100, no. 1, pp. 70–87, 2006.
- [41] K. Friston, "The free-energy principle: A unified brain theory?" *Nature Rev. Neurosci.*, vol. 11, no. 2, pp. 127–138, 2010.
- [42] M. Moore, "Statistical mechanics: A set of lectures," *Phys. Bull.*, vol. 25, no. 5, 1974, Art. no. 191.
- [43] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1275–1286, Jun. 2015.
- [44] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Deep learning network for blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 511–515.
- [45] A. Ng, J. Ngiam, C. Y. Foo, Y. Mai, and C. Suen, "UFLDL tutorial," 2012. [Online]. Available: <http://ufldl.stanford.edu/>
- [46] A. M. Rohaly *et al.*, "Final report from the video quality experts group on the validation of objective models of video quality assessment," ITU-T Standards Contribution COM, 2000, pp. 9–80.
- [47] 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.
- [48] A. Ciancio *et al.*, "No-reference blur assessment of digital pictures based on multifeature classifiers," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 64–75, Jan. 2011.
- [49] 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, Sep. 2014.

- [50] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 339–343.
- [51] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 995–1002.
- [52] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1301–1313, Apr. 2018.
- [53] D. Ghadiyaram and A. C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *J. Vis.*, vol. 17, no. 1, pp. 32–32, 2017.
- [54] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2678–2683, Sep. 2011.
- [55] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218–3231, Oct. 2015.
- [56] P. V. Vu and D. M. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation," *IEEE Signal Process. Lett.*, vol. 19, no. 7, pp. 423–426, Jul. 2012.
- [57] Y. Liu *et al.*, "Quality assessment for real out-of-focus blurred images," *J. Vis. Commun. Image Represent.*, vol. 46, pp. 70–80, 2017.
- [58] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.
- [59] R. Hassen, Z. Wang, and M. M. Salama, "Image sharpness assessment based on local phase coherence," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2798–2810, Jul. 2013.
- [60] C. T. Vu, T. D. Phan, and D. M. Chandler, "S3: A spectral and spatial measure of local perceived sharpness in natural images," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 934–945, Mar. 2012.



Yutao Liu received the B.S. and M.S. degrees in computer science from Harbin Institute of Technology (HIT), Harbin, China, in 2011 and 2013, respectively. He is currently working toward the Ph.D. degree in the School of Computer Science and Technology, HIT.

From 2014 to 2016, he was a Research Assistant with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China.. His research interests include image processing and image quality assessment.



Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively.

He is currently an Associate Professor with Beijing University of Technology, Beijing, China. His research interests include image analysis, environmental perception, quality assessment, machine learning, and big data analysis.

Dr. Gu received the Best Paper Award at the IEEE International Conference on Multimedia and Expo in 2016, and the Excellent Ph.D. Thesis Award from the Chinese Institute of Electronics in 2016. He was the leading Special Session Organizer in VCIP 2016 and ICIP 2017, and serves as a Guest Editor in the *Digital Signal Processing* journal. He is currently an Associate Editor of IEEE ACCESS and IET Image Processing and is the Reviewer for 20 top SCI journals.



Shiqi Wang (M'15) received the B.S. degree in computer science from Harbin Institute of Technology, Harbin, China, in 2008, and the Ph.D. degree in computer application technology from Peking University, Beijing, China, in 2014.

From March 2014 to March 2016, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. From April 2016 to April 2017, he was a Research Fellow with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong, Kowloon Tong, Hong Kong. He has proposed more than 30 technical proposals to ISO/MPEG, ITU-T, and AVS standards. His research interests include image/video compression, analysis, and quality assessment.



Debin Zhao (M'11) received the B.S., M.S., and Ph.D. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 1985, 1988, and 1998, respectively.

He is currently a Professor with the Department of Computer Science, Harbin Institute of Technology. He has authored or coauthored over 200 technical articles in refereed journals and conference proceedings in the areas of image and video coding, video processing, video streaming and transmission, and pattern recognition.



Wen Gao (M'92–SM'05–F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

From 1991 to 1995, he was a Professor of computer science with Harbin Institute of Technology, Harbin, China. He was a Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He is currently a Professor of computer science with Peking University, Beijing, China. He has authored five books and more than 600 technical articles in refereed journals and conference

proceedings in image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics.

Dr. Gao serves the editorial board for several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, EURASIP Journal of Image Communications, and the Journal of Visual Communication and Image Representation. He was a Chair of a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE ICME and ACM Multimedia, and also served on the advisory and technical committees of numerous professional organizations.