

Toward a No-Reference Quality Metric for Camera-Captured Images

Runze Hu^{ID}, Member, IEEE, Yutao Liu^{ID}, Ke Gu^{ID}, Member, IEEE, Xiongkuo Min^{ID}, Member, IEEE,
and Guangtao Zhai^{ID}, Senior Member, IEEE

Abstract—Existing no-reference (NR) image quality assessment (IQA) metrics are still not convincing for evaluating the quality of the camera-captured images. Toward tackling this issue, we, in this article, establish a novel NR quality metric for quantifying the quality of the camera-captured images reliably. Since the image quality is hierarchically perceived from the low-level preliminary visual perception to the high-level semantic comprehension in the human brain, in our proposed metric, we characterize the image quality by exploiting both the low-level image properties and the high-level semantics of the image. Specifically, we extract a series of low-level features to characterize the fundamental image properties, including the brightness, saturation, contrast, noiseness, sharpness, and naturalness, which are highly indicative of the camera-captured image quality. Correspondingly, the high-level features are designed to characterize the semantics of the image. The low-level and high-level perceptual features play complementary roles in measuring the image quality. To infer the image quality, we employ the support vector regression (SVR) to map all the informative features to a single quality score. Thorough tests conducted on two standard camera-captured image databases demonstrate the effectiveness of the proposed quality metric in assessing the image quality and its superiority over the state-of-the-art NR quality metrics. The source code of the proposed metric for camera-captured images is released at <https://github.com/YT2015?tab=repositories>.

Index Terms—Camera-captured image, deep neural network (DNN), image quality assessment (IQA), no-reference (NR)/blind.

Manuscript received 17 May 2021; revised 15 September 2021; accepted 6 November 2021. Date of publication 30 November 2021; date of current version 17 May 2023. This work was supported in part by the Young Talent Project of Ocean University of China and in part by the National Natural Science Foundation of China under Grant 62076013, Grant 62021003, Grant 61831015, and Grant 61927809. This article was recommended by Associate Editor S. Chen. (*Runze Hu and Yutao Liu contributed equally to this work.*) (*Corresponding author: Yutao Liu.*)

Runze Hu is with the Department of Information Science and Technology, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China (e-mail: hrzlpk2015@gmail.com).

Yutao Liu is with the School of Computer Science and Technology, Ocean University of China, Qingdao 266100, China (e-mail: liuyutao2008@gmail.com).

Ke Gu is with the Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke@bjut.edu.cn).

Xiongkuo Min is with the Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: minxiongkuo@gmail.com).

Guangtao Zhai is with the Institute of Image Communication and Network Engineering and the MoE Key Laboratory of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhaiguangtao@sjtu.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCYB.2021.3128023>.

Digital Object Identifier 10.1109/TCYB.2021.3128023

I. INTRODUCTION

CAMERA-CAPTURED images become increasingly important in our daily lives. Especially, during the period of the pandemic, communications by means of the camera-captured images can avoid face-to-face contact and hence, preventing the spread of the virus potentially. However, due to different shooting conditions, camera equipment, or shooting skills, the acquired images are of jagged or even poor quality, which may severely affect the communication efficiency. Therefore, a reliable quality metric is eagerly required for quantifying, monitoring, and controlling the quality of the camera-captured images in the image-based applications.

Since the camera-captured images are usually watched by human beings, the most credible manner for judging their quality is subjective quality evaluation by observers. However, subjective quality test is time consuming and unwieldy, which cannot satisfy the real-time requirements. In contrast, the objective image quality assessment (IQA) conducted by computationally efficient models is much more appealing for the image quality evaluation task. On the basis of the approachability of the original pristine image, current objective IQA approaches can be divided into three types, that is: 1) full-reference (FR); 2) reduced-reference (RR); and 3) no-reference (NR) methods. In detail, FR IQA methods refer to those that compute the image quality by fully consulting its original counterpart [1], [2]. RR IQA extracts partial information from the image for quality estimation [3]–[5]. As the original image is fully or partially available, FR and RR IQA are capable of delivering high prediction performance. However, when the original image is absent or unavailable, both of the FR and RR methods become invalid. In such condition, NR IQA metrics, also known as blind IQA (BIQA) metrics, which do not require the original image in quality estimation, are developed to predict the image quality [6]–[9]. In reality, it is always intractable to acquire the perfect original image of the camera-captured image. Therefore, BIQA is the most appropriate way to evaluate the camera-captured image quality.

A. Related Work

Early attempts in the BIQA research are distortion-specific, concentrating on estimating the degree of specific distortions in the image for quality assessment. For example, Wang *et al.* [10] established an NR quality model for evaluating the JPEG compressed images by measuring the blurring and blockiness level of the image. Pan *et al.* [11] proposed

a block-by-block method to quantify the blocking artifacts in the image. Zhai *et al.* [12] exploited natural image statistics to deduce the image noise level. Liu *et al.* [13] proposed a patch-based noiseness estimation model and proposed that the noise-level parameter should be modulated according to the image complexity. Liu *et al.* [14] extracted the phase congruency and gradient magnitude features to characterize the blur degree of the out-of-focused blurred images. Ferzli and Karam [15] proposed a blind sharpness metric by estimating the just noticeable blur near the image edges. Gu *et al.* [16] learned an NR model for evaluating the quality of the enhanced images with big data. Yue *et al.* [17] proposed an effective and efficient blind quality measure for the contrast-distorted images. Although the above-mentioned BIQA approaches deliver high prediction accuracy in estimating the specific distortion degrees, they are still limited in evaluating the quality of the camera-captured images.

Recent years have seen the prosperity of the general-purpose or universal BIQA methods, which predict the image quality without restricting to a particular type of distortion in advance. Representative general-purpose BIQA approaches are reviewed as follows. Moorthy and Bovik [18] evaluated the image quality by characterizing the statistical regularities in the wavelet transform domain. In [19], natural scene statistics (NSS) features of the image were extracted to capture the image quality degradations. Saad *et al.* [20] exploited the DCT coefficients and designed the quality-aware features to measure the image quality. Wu *et al.* [21] proposed a fast NR quality metric by analyzing the local binary patterns of the query image. Ye and Doermann [22] proposed a BIQA method, in which a visual codebook was constructed to encode the image and the informative features were thus extracted for quality evaluation. In [23], a quality-aware clustering (QAC) method was proposed by learning a group of centroids to encode the image quality. In [24], a natural image quality evaluator (NIQE) was devised, in which the quality-aware features were fitted with a multivariate Gaussian (MVG) model, and then the MVG model parameters were extracted for quality computation. Zhang *et al.* [25] proposed an integrated local feature-enriched NIQE (IL-NIQE) by characterizing more statistical features. Liu *et al.* [6] constructed an unsupervised BIQA method by characterizing the image structure, naturalness, and perception characteristics, respectively. Li *et al.* [26] proposed a contourlet energy statistics-based opinion-unaware BIQA method, which measures the perceptual degradation level based on the variation in subband energies of contourlet transforms of images. Ghadiyaram and Bovik [27] designed an NR model, in which rich statistical features were extracted from multiple domains of the image to characterize the image quality. Wu *et al.* [28] proposed an efficient BIQA approach, which extracted multiple-domain features to simulate the hierarchical structure of the visual cortex perception and employed label transfer strategy to infer the image quality.

The deep learning technologies based on the convolutional neural network (CNN) have gained great success in various computer vision tasks [29], [30]. Inspired by this,

IQA researchers have investigated incorporating CNN into BIQA and proposed some successful BIQA metrics. For example, Ma *et al.* [29] proposed an end-to-end BIQA model, which optimized the distortion type identification and quality prediction in a joint manner. In [31], a two-stream CNN model was trained to predict the image quality, which can extract rich features in quality computation. For characterizing the image quality degradations comprehensively, Wu *et al.* [32] designed hierarchical features, including the low-level structural features and the high-level semantic features extracted from the deep residual network, based on which a hierarchical feature degradation BIQA method was built. In [33], a great deal of quality-discriminable image pairs was generated, based on which a new discriminable image pairs-inferred quality index was constructed. In [34], a set of quality-aware features was deployed to characterize the image quality, and then, the general regression neural network was adopted to learn the regression from the feature space to the image quality score. Bosse *et al.* [35] presented a framework of deep neural networks (DNNs), which can be trained for both FR IQA and NR IQA.

Although the aforesaid general-purpose BIQA methods can be applied to evaluate the quality of the camera-captured image directly, their prediction performance still remains unsatisfactory. The major reasons can be analyzed in the following. Above all, as indicated in [36], the camera-captured image is usually afflicted by complex mixtures of multiple distortions, leading to that the quality degradation condition of the camera-captured image is considerably complicated. However, many existing general-purpose BIQA approaches, such as [18], [19], [21], and [24], only characterized single or a small quantity of low-level image properties to estimate the image quality, which is far from adequate for quality evaluation. Second, the high-level semantics of the image was often ignored in those existing BIQA metrics, which may also hinder their prediction accuracy. For example, without considering the high-level semantics, those BIQA methods are unable to distinguish the intrinsic flat regions and the blurry regions in the image, which have quite divergent perceptual quality [37]. Third, as for the deep learning-based BIQA metrics, they always need a great deal of subjectively annotated images in training a reliable prediction model. If the requirement for sufficient training data cannot be satisfied, their generalization capability will be restricted significantly.

B. Contributions of This Article

To address the aforementioned issues, we in this article propose a new blind metric to quantify the quality of the camera-captured image. It is known that the human brain perceives the image quality in a hierarchical manner, which is from the preliminary visual perception to the high-level semantic comprehension [38]. More specifically, in the low-level visual perception in the primary visual areas, the human brain mainly perceives the early features of the input visual scenes, such as local edge, brightness, noiseness, etc. Then, more complicated and regional information, for example, contour and shape, is

perceived in the following visual areas. After integrating and abstracting the features from the previous visual perception processes, the human brain is able to perceive the high-level semantic information of the image, such as object and categories [32], [39]. Inspired by this, we propose to evaluate the image quality by characterizing both the preliminary visual perception and the semantic comprehension, which coincides with the quality perception process in the human brain. In the preliminary visual perception, the human brain tends to perceive some low-level image properties, such as the brightness, noiseness, sharpness, etc. [9], [12], [16]. Unlike those BIQA methods that only characterize very limited image properties, we hereby characterize plentiful image properties, that is, the brightness, saturation, contrast, noiseness, sharpness, and naturalness, to thoroughly describe the image quality. Dedicated features are carefully designed to represent these quality-related properties, which we regard as the low-level perceptual features. On the other hand, for characterizing the semantic comprehension, we resort to the powerful pretrained DNNs to extract the corresponding semantic features, which we regard as the high-level perceptual features for quality evaluation. Specifically, the high-level features are extracted from the top layers of a specific DNN pretrained on ImageNet [40]. Such high-level features related to the image category have also been verified to be highly effective to characterize the image quality [41]. To derive the image quality level, we adopt the support vector regression (SVR) to integrate all the features and map them to the final quality score [42], [43]. We name the proposed NR quality metric as blind camera-captured images' quality index, or BCQI for abbreviation. The experimental results on two standard camera-captured image databases demonstrate the effectiveness of BCQI and its superiority over the existing representative NR quality metrics.

The contributions of this article can be summarized as follows.

- 1) First, we exploit a number of essential low-level image properties, including brightness, saturation, contrast, noiseness, sharpness, and naturalness, which can be more comprehensive than the existing BIQA metrics in estimating the camera-captured image quality.
- 2) Second, we investigate the high-level semantics of the image for quality evaluation, which is rarely involved in the existing BIQA metrics. Specifically, a number of famous DNN models, such as VGG [44], SqueezeNet [45], GoogleNet [46], ResNet [47], etc., that extract the high-level semantics of the image for quality evaluation are thoroughly investigated.
- 3) Finally, extensive experiments performed on two typical camera-captured image databases demonstrate that the proposed BCQI metric outperforms the existing state-of-the-art BIQA metrics by a large margin.

The remainder of this article is arranged as follows. Section II details the proposed NR IQA metric of the camera-captured images. Section III testifies the effectiveness of the proposed method with extensive experiments and detailed analysis. Finally, we make a conclusion for this article in Section IV.

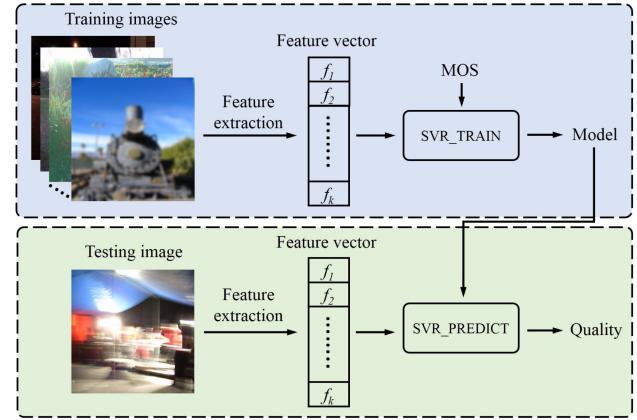


Fig. 1. BCQI framework for camera-captured images.

II. PROPOSED NR IQA METRIC FOR THE CAMERA-CAPTURED IMAGES

A. Overview of the Proposed NR IQA Method

In this article, we measure the image quality by characterizing both the low-level image properties and the high-level semantics. After extracting the corresponding low-level and high-level perceptual features, we concatenate them together and generate a quality-aware feature vector that characterizes the image quality. For deriving the image quality score, we train a quality prediction model through SVR with this quality-aware feature vectors and the associated subjective mean opinion scores (MOSSs) on a training set. Mathematically, given a training set Ω , we extract the feature vector $\mathbf{f}_{I_i} = \{f_1, f_2, \dots, f_k\}$ of image $I_i \in \Omega$ that contains k features. Letting MOS_{I_i} be the MOS value of I_i , the quality prediction model can be obtained by

$$\mathbf{M} = \text{SVR_TRAIN}([\mathbf{f}_{I_i}], [\text{MOS}_{I_i}]), \quad I_i \in \Omega \quad (1)$$

where \mathbf{M} refers to the quality prediction model. Given a new query image \hat{I} to be assessed, we extract its feature vector $\mathbf{f}_{\hat{I}}$ and use SVR to predict its quality with the trained quality prediction model \mathbf{M} , which can be described as

$$Q = \text{SVR_PREDICT}(\mathbf{f}_{\hat{I}}, \mathbf{M}) \quad (2)$$

where Q refers to the predicted quality score of \hat{I} . For more intuitive understanding of the proposed quality assessment scheme, we present the framework of the proposed NR quality metric in Fig. 1. The details of the proposed BIQA metric for the camera-captured images are described as follows.

B. Characterization of the Image Low-Level Properties

In the proposed metric, we characterize a group of fundamental image properties, including the brightness, saturation, contrast, noiseness, sharpness, and naturalness, in order to describe the camera-captured image quality comprehensively.

1) Brightness and Saturation Evaluation: Brightness and saturation are the two most fundamental image properties that influence the image quality. Both low and high brightness or saturation values will degrade the image quality [16]. To

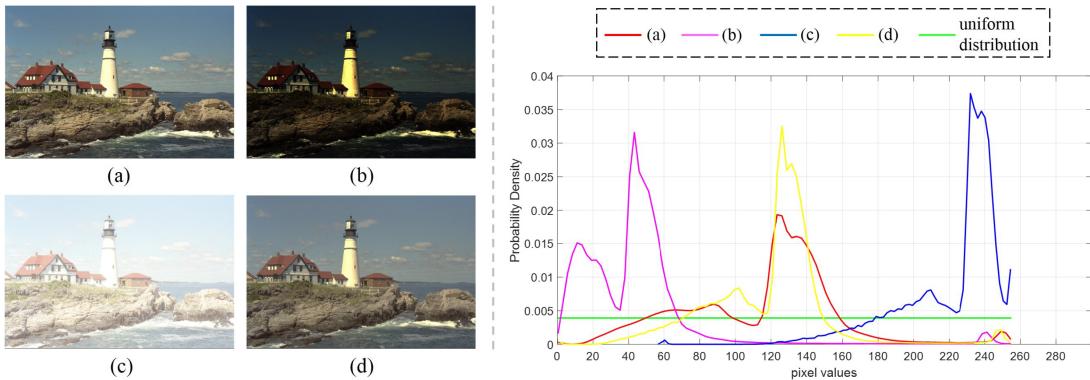


Fig. 2. Histogram distribution of images with different contrast, where (a) refers to a high-contrast image, while (b)–(d) refer to the relatively low-contrast images.

characterize the brightness and saturation of a given camera-captured image I , we first transform I from the RGB color space to the HSI color space as these two factors can be more conveniently measured within the HSI space, expressed as

$$\tilde{I} = \mathcal{T}(I) \quad (3)$$

where \tilde{I} refers to the version of I in the HSI space and $\mathcal{T}(\cdot)$ denotes the operation of transforming I from the RGB space to the HSI space. Specifically, in the HSI space, H-channel represents the hue of the image, S-channel represents the saturation, and the I-channel indicates the brightness of the image. Here, we take the mean values of the I-channel and S-channel to characterize the overall brightness and the saturation of I , respectively, written as

$$b = \frac{1}{K} \sum_{i=1}^K \tilde{I}_I(i), \quad s = \frac{1}{K} \sum_{i=1}^K \tilde{I}_S(i) \quad (4)$$

where b and s refer to the brightness and saturation levels of I . K refers to the total pixel number of \tilde{I} . \tilde{I}_I and \tilde{I}_S represent the I- and S-channels of \tilde{I} , respectively.

2) Contrast Measurement: Contrast is an essential low-level property of an image, which is highly correlated with the image quality [16], [48]. In image processing, the image histogram is a good indicator to the image contrast. For example, the histogram of a low-contrast image tends to exhibit more skewed or diverged against the uniform distribution. Therefore, a histogram equalization algorithm was often applied to modify the image histogram from lopsided to balanced, which can improve the image contrast effectively. In Fig. 2, we show the histogram distributions of images with different contrast, where Fig. 2(a) refers to a high-contrast image, while Fig. 2(b)–(d) refer to the relatively low-contrast images. Their histogram distributions are visually shown in the right part of Fig. 2. From the histogram distributions, we can clearly observe the distributions of the pixel values of the image. For example, as we can see, Fig. 2(c) is of higher brightness. Correspondingly, its pixel values concentrate around 240 as illustrated by the blue line of (c), which shows the pixel value distribution intuitively. More importantly, the histogram distribution in red of (a) with high contrast is more balanced or closer to the uniform distribution in green than that of the low-contrast images. These observations manifest that the

divergence degree between the histogram distribution of the camera-captured image and the uniform distribution provides an effective measure of the contrast of the camera-captured image.

For characterizing the difference between the histogram distribution of a camera-captured image and the uniform distribution, we employ the famous K-L divergence, which is widely adopted in measuring the distance between two given distributions. Specifically, assuming \mathbf{m} and \mathbf{n} as the histogram distribution of a camera-captured image and the uniform distribution, the K-L divergence between \mathbf{m} and \mathbf{n} is defined as follows:

$$D_{KL}(\mathbf{m} \parallel \mathbf{n}) = - \int m(x) \log n(x) dx + \int m(x) \log m(x) dx. \quad (5)$$

One problem here is that the K-L divergence measure is asymmetric, which may induce instability in the BIQA framework. Therefore, we employ another symmetrized and more stable measure of the Jensen–Shannon (J-S) divergence to characterize the discrepancy between the camera-captured image histogram distribution and the uniform distribution, which is directly derived from the K-L divergence, defined as

$$D_{JS}(\mathbf{m}, \mathbf{n}) = \frac{D_{KL}(\mathbf{m} \parallel \Lambda) + D_{KL}(\mathbf{n} \parallel \Lambda)}{2} \quad (6)$$

in which $\Lambda = 1/2(\mathbf{m} + \mathbf{n})$. We take $D_{JS}(\mathbf{m}, \mathbf{n})$ as the quality-aware feature to characterize the contrast of an input camera-captured image accordingly.

3) Noiseness Measurement: The images acquired in a dark environment likely to suffer from noiseness, leading to the degradation of image quality. In the literature, quite a few dedicated works have been proposed to estimate the noiseness level of the image [12], [13], [49]. As indicated in [49], the kurtosis values of the marginal distributions in natural clean images keep unchanged over scales, while those values tend to be modified by the introduced noise. These observations manifest that the variation degree of those kurtosis values can well indicate the noiseness level of the camera-captured image.

Specifically, assuming a clean image x and its noisy version y , according to [49], the kurtosis of y can be calculated via

$$\kappa_y = \frac{\kappa_x(\alpha) - 3}{\left(1 + \frac{\sigma_n^2}{\sigma_x^2}\right)^2} + 3 \quad (7)$$

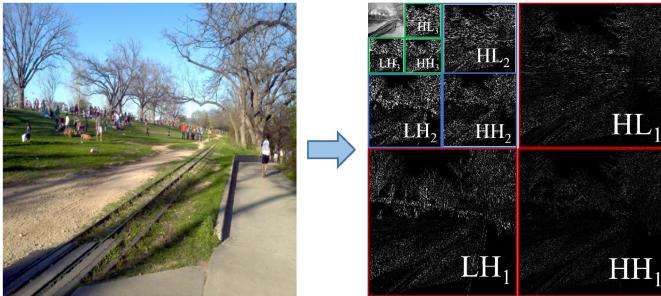


Fig. 3. Camera-captured image from the LIVEC database [36] and its decomposed three-level wavelet subbands. For convenient observation, we amplify the wavelet coefficients in each subband by four times.

where κ_x and κ_y denote the kurtosis values of x and y , α is the shape parameter of the distribution of x , and σ_x^2 and σ_n^2 represent the variances of x and the noise, respectively. Then, the noise variance $\hat{\sigma}_n^2$ and the kurtosis $\hat{\kappa}_x$ of the clean image can be estimated by minimizing

$$\hat{\kappa}_x, \hat{\sigma}_n^2 = \underset{\kappa_x, \sigma_n^2}{\operatorname{argmin}} \sum_{i=2}^{N^2} \left| \frac{\kappa_x - 3}{\left(1 + \frac{\sigma_n^2}{\hat{\sigma}_n^2 - \sigma_n^2}\right)^2} + 3 - \hat{\kappa}_{y_i} \right| \quad (8)$$

where $\hat{\sigma}_{y_i}^2$ and $\hat{\kappa}_{y_i}$ are the variance and kurtosis of image y_i , which is the i th response image by convolving image y with every filter of the $N \times N$ DCT basis. Here, we take the estimated noise variance $\hat{\sigma}_n^2$ to measure the noiseness level of the camera-captured image.

4) *Sharpness Assessment*: Sharpness is possibly the most important low-level property of a camera-captured image. Generally, there are mainly two kinds of approaches to evaluate the sharpness level of the image. One is through measuring the structures or edges in the image, such as [14] and [15]. The other evaluates the image sharpness by characterizing the spectral energies in the transform domain of the image [50], [51]. Here, we employ an effective and efficient strategy for sharpness assessment by characterizing the log energy of the wavelet coefficients of the image [51].

We at first decompose the camera-captured image into three-level wavelet subbands with the Cohen–Daubechies–Faueraue 9/7 filters [52], denoted as $\{LH_l, HL_l, HH_l | l = 1, 2, 3\}$, as illustrated in Fig. 3. Then, we calculate the log energy of each subband at each level l

$$E_{XY_l} = \log_{10} \left(1 + \frac{1}{M_{XY_l}} \sum_{i,j} C_{XY_l}^2(i,j) \right) \quad (9)$$

where E_{XY_l} refers to the log energy of subband XY_l . XY refers to LH , HL , or HH . M_{XY_l} counts the number of the coefficients in XY_l . $C_{XY_l}(i,j)$ represents the coefficient at (i,j) . Following that, we calculate the total log energy of the wavelet coefficients in each level as follows:

$$E_l = \lambda_1 \times E_{HL_l} + \lambda_2 \times E_{LH_l} + \lambda_3 \times E_{HH_l} \quad (10)$$

where E_l refers to the total energy of the wavelet coefficients at level l . λ_1 , λ_2 , and λ_3 are the weights assigned to the log energy of each XY subband, satisfying $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Since

the HH subband covers a higher radial spatial frequency, which is more important for sharpness assessment, we empirically set $\lambda_1 = \lambda_2 = 0.1$ and $\lambda_3 = 0.8$ to emphasize the log energy of the HH subband. Finally, we calculate the total log energy of the entire image by aggregating the log energy of the wavelet coefficients at each level as

$$E_T = \sum_{l=1}^3 \omega_l E_l \quad (11)$$

where E_T represents the total log energy of the input image and ω_l is the weight at level l . We set $\omega_1 = 4/7$, $\omega_2 = 2/7$, and $\omega_3 = 1/7$ to emphasize the log energy of finer scales. E_T is then used to assess the sharpness of the camera-captured image.

5) *Naturalness Estimation*: Naturalness depicts the commonality belonging to the majority of natural images, such as the NSS regularities [19], [24]. However, violating such commonality will make the image look unnatural, reducing the image quality as well. For quantifying the commonality of natural images, characterization of the NSS regularities was often employed in the BIQA research [18], [19], [24]. Here, we formulate the regularity of the locally mean subtracted and contrast normalized (MSCN) coefficients to characterize the image naturalness as in [19]. Specifically, the MSCN coefficients of an image I can be calculated by

$$I'(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1} \quad (12)$$

where (i,j) denotes the pixel coordinate, and $I'(i,j)$ refers to the MSCN coefficient at (i,j) . $\mu(i,j)$ and $\sigma(i,j)$ are the mean and standard deviation of the patch centered around (i,j) , calculated as

$$\mu(i,j) = \sum_{s=-S}^S \sum_{t=-T}^T \ell_{s,t} I(i+s, j+t) \quad (13)$$

$$\sigma(i,j) = \sqrt{\sum_{s=-S}^S \sum_{t=-T}^T \ell_{s,t} [I(i+s, j+t) - \mu(i,j)]^2} \quad (14)$$

where $\ell = \{\ell_{s,t} | s = -S, \dots, S; t = -T, \dots, T\}$ refers to a 2-D circularly symmetric Gaussian weighting filter. As for a natural image, the MSCN coefficients distribution well conforms to the zero-mean generalized Gaussian distribution (GGD), while the GGD model parameters vary according to the variation of the image naturalness degree [19]. Accordingly, we first model the MSCN coefficients distribution with the zero-mean GGD, defined as

$$g(\lambda; \alpha, \beta) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|\lambda|}{\beta}\right)^\alpha\right) \quad (15)$$

in which α and β refer to the GGD model parameters. $\Gamma(\cdot)$ refers to the gamma function, defined as

$$\Gamma(x) = \int_0^\infty \phi^{x-1} e^{-\phi} d\phi, \quad x > 0. \quad (16)$$

In implementation, we employ the moment matching-based algorithm [53] to estimate α and β , which will be used to characterize the naturalness level of the camera-captured image.

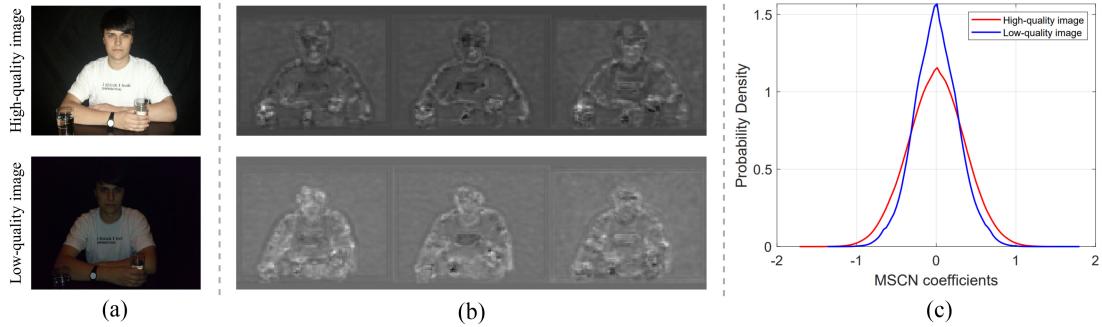


Fig. 4. MSCN coefficients distributions and high-level semantic feature maps of the high-quality and low-quality images taken from the CID2013 database [54]. (a) High-quality and low-quality images. (b) High-level semantic feature maps. (c) MSCN coefficients distributions.

Fig. 4(c) shows the MSCN coefficients distributions of the high-quality and low-quality images taken from the CID2013 database [54]. Note that the high-quality image is with less distortion, while the low-quality image is with more complex distortion. In Fig. 4(c), we can observe a notable modification in the MSCN coefficients distributions of two images. This clearly demonstrates the impact of distortion on the low-level MSCN statistics. Therefore, it is effective to characterize the MSCN statistics in naturalness estimation for quality assessment.

C. Characterization of the Image High-Level Semantics

As indicated in [37], the neglect of the high-level semantics of the image in quality evaluation may lead to some undesirable results, such as assigning a clear blue sky with a poor quality score, which is inconsistent with the subjective quality evaluation. Therefore, we further characterize the high-level semantics of the image for evaluating the quality of the camera-captured image more precisely. In order to characterize the image semantics, one of the popular techniques in the literature is to feed the image into a pretrained DCNN model and extract the output of the top layers as the high-level features [37], [55]. The rationale behind this approach lies in that the DCNN model can mimic the human visual perception mechanism by cascading the convolutional layers and the pooling layers for extracting the high-level semantics of the image [47].

In this work, we also adopt the DCNN-based methodology to extract the high-level perceptual features. Since the pretrained DCNN model generally requires an image patch of fixed size as input [46], [47], we extract the central patch \mathbf{p}_c of size $N \times N$ of the image and feed it into the pretrained DCNN model to extract its high-level semantics, denoted as

$$\mathbf{f}_h = \text{DCNN}(\mathbf{p}_c, L; \mathbf{W}) \quad (17)$$

where \mathbf{f}_h refers to the high-level features, L indicates the layer to extract the features, and \mathbf{W} refers to the parameters of the DCNN model. Note that the central patch may not represent the entire image. There are other ways to extract the patches, which can represent the image better, such as dividing the image into patches and then extracting the high-level features of each patch. However, this methodology leads to a high-dimension feature vector and much computational time. Here, we only consider the central patch based on the following

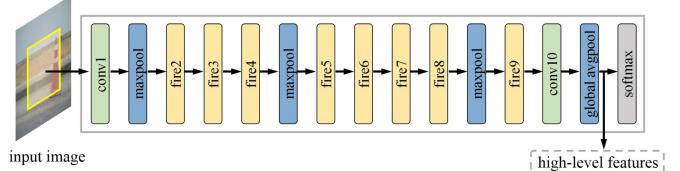


Fig. 5. Extraction of the high-level semantic features of an input image using the SqueezeNet [45].

two considerations. First, the central patch is extracted from the query image directly, which contains the original quality information of the image. Second, visual saliency research indicates that more visual attention of human eyes is focused on the central area of the image [56], [57]. In addition, merely extracting the high-level features of the central patch can also achieve relatively high prediction performance (please refer to Section III-B). Therefore, extracting the high-level feature of the central patch is an effective and efficient strategy. Other more effective methods of extracting patches that can represent the image better will be investigated in future research. In implementation, we take the activations of the layer before the last softmax layer, which gives rise to a 1000-dimension vector that characterizes the high-level semantics of the image.

For convenient understanding, in Fig. 5, we illustrate the extraction process of the high-level perceptual features with SqueezeNet [45]. We input the central patch of size 227×227 from the image into the neural network and take the output of the layer "global avgpool" as our high-level semantic features. In Fig. 4(b), we show the semantic feature maps obtained from the convolutional layer "conv10" in SqueezeNet. The layer of "conv10" is configured with 1000 filters and subsequently outputs 1000 feature maps. Here, we show the first three feature maps of the two images to visually present the impact of distortions on the high-level semantic features of the image. From this figure, it can be seen that the feature maps of the high-quality image and the low-quality image exhibit obvious differences, which intuitively demonstrate the significant impact of distortions on the high-level semantics and the potential effectiveness of employing semantic features in quality evaluation.

The extracted high-level semantic features are combined with the low-level perceptual features to produce the quality-aware vector whose dimensions are summarized in Table I.

TABLE I
SUMMARY OF THE QUALITY-AWARE FEATURES

Index	Category	Dimension
1	Brightness	1
2	Saturation	1
3	Contrast	1
4	Noiseness	1
5	Sharpness	1
6	Naturalness	2
7	Semantic information	1000

D. Quality Evaluation

To evaluate the camera-captured image quality, a regression model that maps from the feature space to the image quality score is needed. Toward this end, any regressor can be employed. In this article, we employ the widely adopted SVR to learn such regression from the feature space to the quality score, which is good at processing high-dimensional data [58] and has been pervasively utilized by many successful BIQA works [19], [27]. Specifically, given a set of training data $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_k, y_k)\}$, where \mathbf{x}_i , $i = 1, 2, \dots, k$, is the feature vector containing the above-designed low-level and high-level perceptual features and y_i is the associated subjective MOS value. We expect to learn a function f that predicts the image quality score with the feature vector \mathbf{x} , written by

$$f(\mathbf{x}) = \langle \omega, \mathbf{x} \rangle + b \quad (18)$$

where $\langle \cdot, \cdot \rangle$ refers to the inner product, ω is the weight vector, and b is a bias value. By introducing two slack variables ξ_i and ξ_i^* , we can obtain ω and b by solving the following problem:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 + C \left(\sum_{i=1}^k \xi_i + \sum_{i=1}^k \xi_i^* \right) \\ \text{s.t.} \quad & \langle \omega, \mathbf{x}_i \rangle + b - y_i \leq \varepsilon + \xi_i \\ & y_i - \langle \omega, \mathbf{x}_i \rangle - b \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, k \end{aligned} \quad (19)$$

where C refers to a constant for balancing the two terms. In implementation, we employ the radial basis function (RBF) kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$ to solve the above optimization problem. Detailed derivation of SVR can be referred to [58]. After obtaining the regression model, we use it to predict the quality score of a given camera-captured image.

III. EXPERIMENT VALIDATION

A. Camera-Captured Image Databases and Performance Evaluation Protocol

We evaluate the performance of the proposed BCQI on two typical camera-captured image databases, which are the LIVE in the wild image quality challenge database (LIVEC) [36] and the CID2013 database [54]. Specifically, the LIVEC database contains 1162 images that were captured by a wide variety of contemporary mobile camera devices, such as smartphones and tablets. The CID2013 database is composed of 474 images captured by 79 different cameras that range from low quality to high quality.

In our experiments, we utilize four widely used mathematical measures to quantify the prediction performance of BCQI, which are the Pearson's linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SRCC), Kendalls rank correlation coefficient (KRCC), and root mean square error (RMSE), recommended by [65]. Specifically, the above four measures are calculated between the objective quality scores of the BIQA algorithms and the subjective quality scores annotated by human beings. A superior BIQA metric is expected to achieve the absolute values of SRCC, KRCC, and PLCC close to 1 and RMSE value close to 0. According to [65], prior to calculating the SRCC, KRCC, PLCC, and RMSE values, the objective scores are required to be fitted onto the subjective quality scores through a five-parameter logistic function, defined as

$$z(o) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2 \cdot (o - \beta_3))} \right) + \beta_4 \cdot o + \beta_5 \quad (20)$$

in which $z(o)$ denotes the fitted quality score, o denotes the predicted objective score of BIQA metrics, and β_1 – β_5 refer to the model parameters, which are determined in curve fitting.

B. Overall Prediction Performance Comparison

In our experiments, 15 state-of-the-art BIQA metrics, including BRISQUE [19], DIIVINE [18], NFERM [61], BMPRI [62], FRIQUEE [27], DBCNN [63], HyperIQ [64], NPQI [60], MEON [29], LPSI [21], dipIQ [33], NIQE [24], IL-NIQE [25], SNP-NIQE [6], and SISBLIM [59], are compared with our proposed BCQI. We divide all the BIQA methods into two types of the supervised and the unsupervised methods. Among them, the first eight methods of BRISQUE, DIIVINE, NFERM, BMPRI, FRIQUEE, DBCNN, HyperIQ, MEON, and the proposed BCQI belong to the supervised methods, which need to train the quality model with the subjective quality scores. The other seven methods of SISBLIM, LPSI, dipIQ, NIQE, IL-NIQE, SNP-NIQE, and NPQI, which do not require the subjective scores for constructing the quality model, belong to the unsupervised methods. Among the BIQA metrics, DBCNN, HyperIQ, MEON, and dipIQ are the deep learning-based BIQA metrics. As the supervised methods need the training process for quality estimation, each database was randomly divided into two sets, including: 1) a training set that contains 80% images for calibrating the prediction model and 2) a testing set that contains the other 20% images for evaluating the calibrated BIQA model. Note that the training set and the testing set are completely nonoverlapped. For those supervised methods, we train the quality model on the training set and evaluate its performance on the testing set. We repeat this process 1000 times to mitigate the performance bias. As for the unsupervised BIQA approaches, we test them on the same testing set as that used for the supervised methods. The mean performance of such 1000-time experiments of all the BIQA metrics on LIVEC and CID2013 is listed in Table II, where the first, second, and third highest performance results on each evaluation index of these two databases are emphasized with red, blue, and magenta, respectively. For a comprehensive comparison, we introduced an additional rank

TABLE II
PREDICTION PERFORMANCE COMPARISON MEASURED BY SRCC, KRCC, PWRC, PLCC, AND RMSE ON LIVEC AND CID2013 DATABASES.
THE FIRST, SECOND, AND THIRD HIGHEST PERFORMANCE VALUES IN EACH DATABASE
ARE HIGHLIGHTED WITH RED, BLUE, AND MAGENTA, RESPECTIVELY

BIQA Metrics	Type	LIVEC					CID2013				
		SRCC	KRCC	PWRC	PLCC	RMSE	SRCC	KRCC	PWRC	PLCC	RMSE
SISBLIM [59]	unsupervised	0.4288	0.2964	5.6912	0.4953	17.4983	0.6518	0.4785	11.8332	0.7065	15.8679
LPSI [21]	unsupervised	0.0867	0.0555	2.2376	0.3224	19.0912	0.3223	0.2183	4.8769	0.4762	19.7482
dipIQ [33]	unsupervised	0.1788	0.1222	2.3799	0.2990	19.2419	0.1112	0.0697	4.8742	0.3406	21.0810
NIQE [24]	unsupervised	0.4495	0.3073	6.2348	0.5068	17.3911	0.6502	0.4640	11.6844	0.6748	16.5665
IL-NIQE [25]	unsupervised	0.4384	0.2986	6.0817	0.5138	17.3078	0.3063	0.2116	4.1781	0.4547	19.9830
SNP-NIQE [6]	unsupervised	0.4634	0.3157	6.4264	0.5277	17.1360	0.7110	0.5158	13.4301	0.7286	15.3745
NPQI [60]	unsupervised	0.4755	0.3276	6.3108	0.5049	17.4007	0.7658	0.5662	14.8934	0.7857	13.8794
BRISQUE [19]	supervised	0.5228	0.3639	6.5182	0.5610	16.6950	0.7353	0.5453	14.2801	0.7375	15.1526
DIIVINE [18]	supervised	0.5459	0.3779	7.2862	0.5876	16.3222	0.6891	0.5009	15.6448	0.6972	16.0949
NFERM [61]	supervised	0.5181	0.3548	6.7517	0.5493	16.8604	0.7079	0.5234	13.6265	0.7179	15.6191
BMPRI [62]	supervised	0.4595	0.3191	5.8938	0.5092	17.3568	0.6507	0.4770	12.1811	0.6798	16.4498
FRIQUEE [27]	supervised	0.6827	0.4923	8.8256	0.7128	14.1444	0.7375	0.5495	15.7993	0.7465	14.9233
MEON [29]	supervised	0.3683	0.2496	5.1535	0.4561	17.9278	0.3757	0.2537	5.9749	0.4186	20.4055
DBCNN [63]	supervised	0.6116	0.4333	8.4640	0.6435	15.4275	0.7533	0.5550	14.5441	0.7879	13.8119
HyperIQ [64]	supervised	0.7209	0.5247	9.7781	0.7380	13.6016	0.7762	0.5958	14.7496	0.8310	12.4666
BCQI (Pro.)	supervised	0.7506	0.5542	10.2320	0.7732	12.7866	0.8948	0.7248	17.5967	0.9080	9.3775

TABLE III
T-TEST RESULTS ON THE LIVEC AND THE CID2013 DATABASES. THE VALUE OF -1, 0, OR 1 IN THE TABLE REPRESENTS THE PROPOSED BCQI IS
INFERIOR, EQUAL, OR SUPERIOR TO ITS COMPETITORS STATISTICALLY (95% CONFIDENCE)

t-test	SISBLIM	LPSI	dipIQ	NIQE	IL-NIQE	SNP-NIQE	NPQI	BRISQUE	DIIVINE	NFERM	BMPRI	FRIQUEE	MEON	DBCNN	HyperIQ
LIVEC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CID2013	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

correlation statistics, namely, perceptually weighted rank correlation (PWRC) [66], which has been verified to be highly efficient in measuring the prediction performance of BIQA methods.

In Table II, it is clear to be observed that among all the BIQA metrics, the proposed BCQI achieves the best prediction performance on both of the two databases and outperforms the second best performer notably. These observations demonstrate the effectiveness and superiority of BCQI in characterizing the image quality explicitly. By comparison, those methods, such as LPSI, BRISQUE, DIIVINE, NIQE, etc., that only characterize a small quantity of low-level properties are observed to achieve undesirable even poor prediction performance. In addition, although FRIQUEE extracts a great deal of low-level image features for quality evaluation, without characterizing the high-level semantics, it can only deliver moderate prediction performance. For the deep learning-based BIQA methods, HyperIQ that integrates content understanding and perception rule learning achieves relatively better prediction performance. Due to characterizing both the low-level image properties and the high-level semantics, the proposed BCQI is able to deliver superior prediction performance to all the other BIQA metrics.

In addition, to visually show the correlation between the subjective quality scores in terms of MOS and the objective scores produced by the BIQA methods, we provide the scatter plots of the subjective scores against the objective scores on the LIVEC database in Fig. 6, where each blue “+” represents one image in the LIVEC database and the red curve is obtained by curve fitting. Without loss of generality, these experimental results were taken from one testing set from the 1000-time partitions. From Fig. 6, we observe that the image data points of

BCQI distribute more compactly and evenly around the fitted curve, which indicates the preferable prediction performance of BCQI vividly. For the other methods, some of the fitted curves are even broken lines, such as SISBLIM, LPSI, SNP-NIQE, etc. Furthermore, their data points also distribute more divergently or irregularly, which implies the weak correlation between the compared BIQA methods and the subjective quality scores.

C. Statistical Significance Test

To inspect the statistical significance of the objective BIQA models, we performed t-test on the prediction residuals between the subjective scores and the fitted objective scores on LIVEC and CID2013 databases [67]. We present the experimental results of t-test in Table III, where the symbol of -1, 0, or 1 indicates that the proposed BCQI is inferior, equal, or superior to its competitors statistically (95% confidence). By observing this table, we can find that all the values are “1” on each image database, which verifies that the proposed BCQI is superior to all other BIQA methods statistically.

D. Cross-Database Validation

Having testified the BIQA methods on each database, we would like to evaluate their prediction performance through the cross-database validation methodology, which is used to test the generalization capability of the objective BIQA metrics. The experiments of cross-database validation were performed as follows. For the supervised methods, we trained the quality models on one image database and tested them on the other database. For the unsupervised BIQA approaches, we directly tested them on the corresponding databases. The experimental

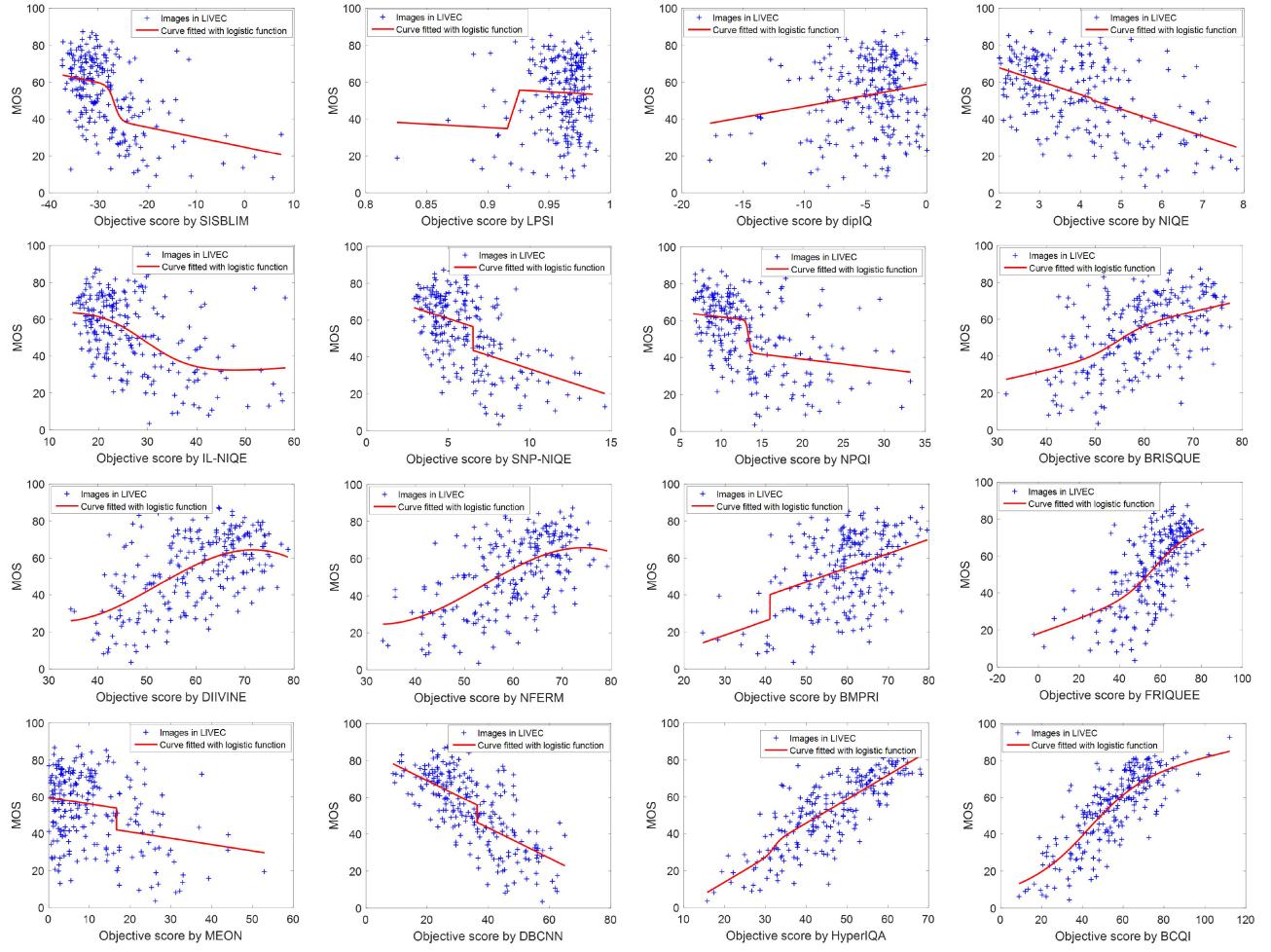


Fig. 6. Scatter plots of subjective scores against objective scores delivered by BIQA methods on the LIVEC database.

TABLE IV

CROSS-DATABASE VALIDATION MEASURED BY SRCC, KRCC, PLCC, AND RMSE ON LIVEC AND CID2013 DATABASES. THE FIRST, SECOND, AND THIRD HIGHEST PERFORMANCE VALUES IN EACH DATABASE ARE HIGHLIGHTED WITH RED, BLUE, AND MAGENTA, RESPECTIVELY

BIQA Metrics	Type	LIVEC (Trained on CID2013)				CID2013 (Trained on LIVEC)			
		SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
SISBLIM [59]	unsupervised	0.4283	0.2952	0.4948	17.6378	0.6554	0.4780	0.7038	16.0826
LPSI [21]	unsupervised	0.0832	0.0523	0.2999	19.3621	0.3230	0.2167	0.4524	20.1914
dipIQ [33]	unsupervised	0.1771	0.1205	0.3174	19.2468	0.0927	0.0545	0.3664	21.0660
NIQE [24]	unsupervised	0.4509	0.3072	0.4963	17.6201	0.6539	0.4630	0.6660	16.8891
IL-NIQE [25]	unsupervised	0.4395	0.2985	0.5043	17.5269	0.3063	0.2099	0.4274	20.4688
SNP-NIQE [6]	unsupervised	0.4654	0.3161	0.5196	17.3410	0.7157	0.5055	0.7150	15.8287
NPQI [60]	unsupervised	0.4754	0.3259	0.4921	17.6692	0.7704	0.5660	0.7775	14.2392
BRISQUE [19]	supervised	0.3816	0.2624	0.4348	18.2774	0.5806	0.4189	0.5708	18.5903
DIIVINE [18]	supervised	0.4559	0.3110	0.4862	17.7361	0.6003	0.4298	0.6095	17.9492
NFERM [61]	supervised	0.4054	0.2793	0.4574	18.0484	0.6515	0.4601	0.6470	17.2630
BMPRI [62]	supervised	0.3123	0.2165	0.3585	18.9471	0.5074	0.3470	0.5614	18.7357
FRIQUEE [27]	supervised	0.5420	0.3767	0.5595	16.8218	0.6834	0.4986	0.6783	16.6367
MEON [29]	supervised	0.2435	0.1613	0.2938	19.4009	0.3466	0.2295	0.3540	21.1744
DBCNN [63]	supervised	0.4060	0.2768	0.4410	18.2164	0.6567	0.4638	0.6978	16.2175
HyperIQA [64]	supervised	0.4847	0.3349	0.5148	17.4001	0.6502	0.4666	0.7374	15.2934
BCQI (Pro.)	supervised	0.4967	0.3236	0.5225	17.2956	0.6912	0.5057	0.7456	15.0876

results in terms of SRCC, KRCC, PLCC, and RMSE values on LIVEC and CID2013 are listed in Table IV clearly, in which the first, second, and third highest performance results in each database are emphasized with red, blue, and magenta, respectively. From these two tables, we can obtain some

meaningful observations. First, as the unsupervised methods do not require the training process, cross-database validation has little effect on their prediction performance. Second, the performance of all the supervised BIQA methods decreases to a certain extent in such cross-database validation, which is

TABLE V
PERFORMANCE EVALUATION OF THE LOW-LEVEL AND HIGH-LEVEL PERCEPTUAL FEATURES ON LIVEC AND CID2013 DATABASES

Feature type	LIVEC				CID2013			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
Low-level perceptual features	0.5996	0.4187	0.6143	16.0748	0.7555	0.5745	0.7800	14.0292
High-level perceptual features	0.6215	0.4464	0.6433	14.6258	0.7509	0.5693	0.7447	14.9571
Combination	0.7506	0.5542	0.7732	12.7866	0.8948	0.7248	0.9080	9.3775

TABLE VI
PERFORMANCE EVALUATION OF EACH TYPE OF THE LOW-LEVEL PERCEPTUAL FEATURES ON LIVEC AND CID2013 DATABASES

Low-level Features	LIVEC				CID2013			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
Brightness	0.2169	0.1488	0.2973	19.2682	0.4770	0.3388	0.5666	18.4831
Saturation	0.0531	0.0360	0.1459	19.9697	0.3079	0.2173	0.4401	20.1621
Contrast	0.2565	0.1729	0.3014	19.2411	0.2572	0.1780	0.4636	19.8948
Noiseness	0.1383	0.0942	0.2466	19.4371	0.2065	0.1464	0.3276	21.2228
Sharpness	0.4854	0.3317	0.5413	16.9636	0.7342	0.5468	0.7611	14.5508
Naturalness	0.3938	0.2726	0.4500	18.0117	0.4730	0.3393	0.5203	19.1724

not surprising as there exist some differences between different image databases, such as the subjective scoring criteria, the testing environment, etc., which will have an inevitable effect on the supervised BIQA methods. Third, we can also observe that the performance of the proposed BCQI is inferior to that of FRIQUEE on LIVEC database. This is mainly because the image number of CID2013 is small; however, the feature dimension of BCQI is larger than FRIQUEE. Therefore, BCQI needs more images for training the quality model than FRIQUEE. This can be verified by the performance comparison on the CID2013 database. When we trained the BIQA models on the large scale LIVEC database, BCQI outperforms FRIQUEE significantly as indicated by the PLCC values.

E. Ablation Study of BCQI

1) *Performance Evaluation of Each Type of Features in BCQI:* In the proposed BCQI, we exploit the low-level and high-level perceptual features to characterize the image low-level properties and semantics, respectively. Hence, we are interested in being aware of the individual contribution of each type of features in the quality evaluation. Toward this end, we used each type of features to construct the quality model and applied it to predict the image quality, respectively. The experiments were performed on the LIVEC and CID2013 databases. We employed the same train-test strategy described in Section III-B to testify each type of the features in BCQI. The experimental results in terms of SRCC, KRCC, PLCC, and RMSE are listed in Table V.

From this table, we observe that both the low-level and high-level perceptual features are effective in characterizing the image quality and thus, contribute to the overall prediction performance. It is observed that the high-level perceptual features contribute more than the low-level features on the LIVEC database, while the low-level features are more effective than the high-level features on the CID2013 database. The reason for such an observation can be analyzed as follows. For the LIVEC database, the images in it are of complicated distortions and the photographed scenes are not restricted.

For assessing the quality of such complicated images, the human brain has to evoke more high-level visions to judge the image quality [37]. Therefore, the high-level features of the proposed method that have stronger abstraction ability can be more effective than the low-level features in characterizing the image quality. However, in the CID2013 database, the photographed scenes are significantly restricted, that is, it only contains nine photographed scenes. The distortions of the images in CID2013 are also much simpler than that in LIVEC database. What is more, in subjective experiments during the construction of CID2013 database, different from LIVEC database, subjects were requested to rate not only the overall quality but also the low-level properties of the image in the meantime, such as contrast, saturation, sharpness, etc. Consequently, the ratings for the low-level properties may have a greater influence on the subjective rating for the overall image quality. Therefore, the low-level features in our proposed method that characterize the low-level image properties, such as the brightness, saturation, contrast, noiseness, sharpness, etc., contribute more than the high-level features in quality evaluation on the CID2013 database. In Table V, it is also observed that the combination of these two types of features earns the best prediction performance, which strongly verifies that the two types of features that characterize the low-level properties and the high-level semantics play complementary roles in image quality evaluation.

The low-level features in BCQI characterize different image properties, respectively, and we further examine different types of the low-level features in quality estimation. Similarly, we employed each type of the low-level features that characterize different image properties to train the quality model and then test its prediction performance. The experimental results are listed in Table VI. By observing this table, we find that the sharpness feature earns the best prediction performance among all the low-level features, outperforming other features significantly, especially on the CID2013 database, which implies that sharpness assessment plays the leading role in the quality evaluation of the camera-captured image. In addition, the naturalness features that characterize the commonality of natural

TABLE VII
PERFORMANCE EVALUATION OF THE PROPOSED BCQI WITH WHICH A LOW-LEVEL FEATURE IS REMOVED

Removed low-level features	LIVEC				CID2013			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
Brightness	0.7138	0.5193	0.7321	13.7090	0.8025	0.6050	0.8272	12.5869
Saturation	0.7462	0.5510	0.7693	12.8560	0.8410	0.6572	0.8698	10.9347
Contrast	0.7107	0.5163	0.7297	13.8867	0.8457	0.6638	0.8624	10.9496
Noiseness	0.7201	0.5256	0.7424	13.2834	0.8491	0.6712	0.8721	10.8217
Sharpness	0.6764	0.4569	0.6859	15.0363	0.7778	0.5734	0.7832	13.8617
Naturalness	0.6989	0.4886	0.7093	14.2772	0.8053	0.6160	0.8308	12.2663
BCQI (Pro.)	0.7506	0.5542	0.7732	12.7866	0.8948	0.7248	0.9080	9.3775

TABLE VIII
PERFORMANCE EVALUATION OF DIFFERENT NSS-BASED NR-IQA METHODS FOR NATURALNESS ESTIMATION

BIQA metrics	LIVEC				CID2013			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
BCQI with NIQE	0.7540	0.5577	0.7781	12.6655	0.8964	0.7270	0.9094	9.3051
BCQI with IL-NIQE	0.7550	0.5587	0.7801	12.6148	0.8939	0.7236	0.9074	9.4062
BCQI (Pro.)	0.7506	0.5542	0.7732	12.7866	0.8948	0.7248	0.9080	9.3775

images can achieve relatively higher performance among all the low-level features. It is also interesting to find that the brightness feature earns the second best performance on the CID2013 database, which indicates that only measuring the brightness of the image can evaluate the quality of the images in CID2013 to some extent. By comparison, other types of features, such as saturation, contrast, and noiseness, however, perform poorly in characterizing the image quality solely.

To further investigate the individual contribution of the low-level features, we performed experiments by removing each type of the low-level features. The experimental results are listed in Table VII. From this table, we can observe that removing any single feature will lead to the performance degradation of BCQI to some extent, which confirms the role of each feature in quality evaluation as the designed features characterize different properties related to the image quality. By further observation of Table VII, on the LIVEC database, removing the saturation feature brings minimal performance degradation. However, experiments on CID2013 database demonstrate that removing the saturation feature may cause more significant performance loss. In addition, on the LIVEC database, removing other features can lead to more performance degradation than the saturation feature, whereas, on the CID2013 database, removing any single feature will induce a noticeable performance degradation. From the above analysis, we can conclude that it is better to reserve all the features in our proposed BCQI, which can characterize the image quality comprehensively.

2) *Performance Evaluation of Different NSS-Based NR-IQA Methods for Naturalness Estimation:* NSS-based NR-IQA methods, such as NIQE or ILNIQE, can be employed to characterize the image naturalness directly. In this section, we further investigate how BCQI performs when using NSS-based NR-IQA methods for naturalness estimation. Toward this end, we replaced the original naturalness estimation technique of BCQI with the NIQE and IL-NIQE methods. The prediction performance comparison is reported in Table VIII. As observed, NIQE is capable of improving the prediction

performance on these two databases, whereas IL-NIQE can only improve the prediction performance on the LIVEC database. Although both NIQE and IL-NIQE methods demonstrate their potential to further improve the BCQI, such an improvement is not significant. Besides, NIQE and IL-NIQE will introduce more features and higher computational time. Hence, by comprehensively considering the prediction performance and efficiency, it is more suitable to adopt the original naturalness estimation method to construct the blind image quality metric.

3) *Performance Evaluation of Different Deep Neural Networks for Extracting the Semantic Features:* For extracting the high-level semantic features of the camera-captured image, we utilized the pretrained DNN models. Here, we comprehensively tested eight famous DNN models for the extraction of the semantics for quality evaluation, including the SqueezeNet [45], GoogleNet [46], Inceptionv3 [68], DenseNet [69], ResNet [47], ShuffleNet [70], AlexNet [71], and VGG [44], respectively. The extracted high-level semantic features combined with the low-level features were fed into SVR to predict the image quality score. In Table IX, we tabulate the prediction performance of different DNN models with their model sizes in MATLAB pretrained version. The performance is evaluated by SRCC, KRCC, PLCC, and RMSE. It is observed that the prediction performance of GoogleNet, Inceptionv3, and AlexNet in extracting the semantics feature for quality evaluation is relatively lower than that of other DNN models on these two databases. While other DNN models perform similarly in characterizing the image quality. As can be observed, among all the DNN models, SqueezeNet is able to achieve relatively high prediction accuracy. Besides, the model size of SqueezeNet is the minimum with only 4.6 MB, which is very promising for the integration into the mobile devices. Therefore, we choose SqueezeNet as our default model for extracting the high-level semantics in BCQI.

4) *Investigation of New Framework for Image Quality Evaluation:* Apart from the original framework of using SVR

TABLE IX
PERFORMANCE EVALUATION OF DIFFERENT DNNs MODELS ADOPTED BY BCQI ON LIVEC AND CID2013 DATABASES

Deep neural network	Model size	LIVEC				CID2013			
		SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
SqueezeNet [45]	4.6 MB	0.7506	0.5542	0.7732	12.7866	0.8948	0.7248	0.9080	9.3775
GoogleNet [46]	27 MB	0.7183	0.5266	0.7400	13.5573	0.8812	0.7063	0.8971	9.8948
Inceptionv3 [68]	89 MB	0.7108	0.5214	0.7488	13.3632	0.8675	0.6893	0.8847	10.4384
DenseNet [69]	77 MB	0.7480	0.5542	0.7775	12.6791	0.9069	0.7396	0.9180	8.8836
ResNet [47]	44 MB	0.7399	0.5488	0.7756	12.7233	0.8831	0.7066	0.8993	9.7991
ShuffleNet [70]	6.3 MB	0.7459	0.5533	0.7783	12.6531	0.8893	0.7161	0.9016	9.6892
AlexNet [71]	227 MB	0.7266	0.5331	0.7497	13.3400	0.8797	0.7048	0.8940	10.0407
VGG [44]	515 MB	0.7422	0.5490	0.7702	12.8583	0.8840	0.7095	0.8978	9.8663

TABLE X
PERFORMANCE EVALUATION OF DIFFERENT FRAMEWORKS FOR IMAGE QUALITY EVALUATION

Regression techniques	LIVEC				CID2013			
	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
Adapted SqueezeNet	0.7555	0.5622	0.7875	12.2182	0.9005	0.7346	0.9202	9.0658
SVR	0.7506	0.5542	0.7732	12.7866	0.8948	0.7248	0.9080	9.3775

TABLE XI
TIME COST COMPARISON OF THE BIQA METHODS (SECONDS)

SISBLIM	LPSI	dipIQ	NIQE	IL-NIQE	SNP-NIQE	NPQI	BRISQUE
3.60	0.03	1.93	0.12	2.74	1.52	1.03	0.15
DIIVINE	NFERM	BMPRI	FRIQUEE	MEON	DBCNN	HyperIQA	BCQI
12.55	16.46	0.93	58.42	1.13	0.87	0.93	0.98

to integrate all the features for evaluating the image quality, fully connected layers can also be employed to predict the image quality score. Toward this end, we adapted the original architecture of the SqueezeNet from two aspects. First, we added a fully connected layer after the “global avgpool” layer, which connected the high-level and low-level features. Second, we changed the neurons on the output layer to a single neuron that outputs the image quality score. We trained the entire framework for quality evaluation. Note that the train-test strategy was kept the same as that in Section III-B. The experimental results are shown in Table X, where the best performer is highlighted with boldface. From this table, we can find that the new framework can deliver better prediction performance than the original scheme of BCQI on the two databases, which verifies that more effective features can be extracted in this new framework for image quality evaluation. This attempt provides a feasible way to improve our method effectively.

F. Computational Efficiency Evaluation

Computational efficiency of an objective BIQA model is an important index for practical application. Thus, we turn to examine the computational efficiency of the proposed BCQI at last. To ensure a fair comparison, we chose a standard image of size 500×500 from the LIVEC database, and then ran all the BIQA methods on it and recorded the running time of each method, respectively. Note that our computer equipment is with a 3.0-GHz Intel Core i5-9500 CPU, 8-GB RAM, and a graphics card NVIDIA 1080 Ti with 11G graphics RAM. The experiments were conducted on MATLAB R2019b and PyTorch. The running time in terms of seconds is listed in

Table XI clearly. As can be observed, the proposed BCQI works in less than 1 s, which is much more promising than the FRIQUEE, NFERM, DIIVINE methods, etc. These experimental results demonstrate that BCQI has great potential for practical applications.

IV. CONCLUSION

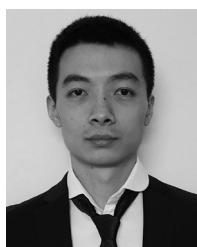
This article has concentrated on the challenging issue of blind quality assessment of the camera-captured images and delivered a novel BIQA metric, which is called BCQI. The proposed BCQI metric quantifies the image quality by characterizing both the low-level image properties and the high-level semantics of the image comprehensively. Dedicated low-level and high-level features are thus extracted to describe the essential image properties, that is, the brightness, saturation, contrast, noiseness, sharpness and naturalness, and the semantics, respectively. The low-level and high-level perceptual features perform synergistically in characterizing the image quality. Finally, SVR is adopted to map all the extracted features to the final image quality score. Extensive tests performed on two representative camera-captured image databases manifest that the proposed BCQI metric yields promising performance in terms of both prediction accuracy and computational efficiency.

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, pp. 600–612, 2004.
- [2] 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.

- [3] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379–391, Feb. 2018.
- [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, pp. 41–52, 2012.
- [5] 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. (ISCAS)*, Montreal, QC, Canada, 2016, pp. 1586–1589.
- [6] Y. Liu *et al.*, "Unsupervised blind image quality evaluation via statistical measurements of structure, naturalness, and perception," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 4, pp. 929–943, Apr. 2020.
- [7] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4559–4565, Dec. 2017.
- [8] Y. Liu and X. Li, "No-reference quality assessment for contrast-distorted images," *IEEE Access*, vol. 8, pp. 84105–84115, 2020.
- [9] Y. Liu, G. Zhai, X. Liu, and D. Zhao, "Quality assessment for out-of-focus blurred images," in *Proc. Vis. Commun. Image Process. (VCIP)*, Singapore, 2015, pp. 1–4.
- [10] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. Int. Conf. Image Process.*, Rochester, NY, USA, Sep. 2002, pp. 477–480.
- [11] F. Pan *et al.*, "A locally-adaptive algorithm for measuring blocking artifacts in images and videos," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 3, Vancouver, BC, Canada, May 2004, pp. 925–928.
- [12] G. Zhai, A. Kaup, J. Wang, and X. Yang, "A dual-model approach to blind quality assessment of noisy images," in *Proc. Picture Coding Symp. (PCS)*, San Jose, CA, USA, Dec. 2013, pp. 29–32.
- [13] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Trans. Image Process.*, vol. 22, pp. 5226–5237, 2013.
- [14] Y. Liu, K. Gu, G. Zhai, X. Liu, D. Zhao, and W. Gao, "Quality assessment for real out-of-focus blurred images," *J. Vis. Commun. Image Represent.*, vol. 46, pp. 70–80, Jul. 2017.
- [15] 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, pp. 717–728, 2009.
- [16] 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.
- [17] G. Yue, C. Hou, T. Zhou, and X. Zhang, "Effective and efficient blind quality evaluator for contrast distorted images," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2733–2741, Aug. 2019.
- [18] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, pp. 3350–3364, 2011.
- [19] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, pp. 4695–4708, 2012.
- [20] 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, pp. 3339–3352, 2012.
- [21] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Quebec City, QC, Canada, 2015, pp. 339–343.
- [22] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, pp. 3129–3138, 2012.
- [23] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, 2013, pp. 995–1002.
- [24] 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.
- [25] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, pp. 2579–2591, 2015.
- [26] C. Li, T. Guan, Y. Zheng, B. Jin, X. Wu, and A. Bovik, "Completely blind image quality assessment via contourlet energy statistics," *IET Image Process.*, vol. 15, no. 2, pp. 443–453, 2021.
- [27] 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, p. 32, 2017.
- [28] Q. Wu *et al.*, "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 3, pp. 425–440, Mar. 2016.
- [29] K. Ma, W. Liu, K. Zhang, Z. Duamu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.*, vol. 27, pp. 1202–1213, 2018.
- [30] K. Gu, Z. Xia, J. Qiao, and W. Lin, "Deep dual-channel neural network for image-based smoke detection," *IEEE Trans. Multimedia*, vol. 22, no. 2, pp. 311–323, Feb. 2020.
- [31] Q. Yan, D. Gong, and Y. Zhang, "Two-stream convolutional networks for blind image quality assessment," *IEEE Trans. Image Process.*, vol. 28, pp. 2200–2211, 2019.
- [32] J. Wu, J. Zeng, W. Dong, G. Shi, and W. Lin, "Blind image quality assessment with hierarchy: Degradation from local structure to deep semantics," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 353–362, Jan. 2019.
- [33] 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, pp. 3951–3964, 2017.
- [34] C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Trans. Neural Netw.*, vol. 22, no. 5, pp. 793–799, May 2011.
- [35] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, pp. 206–219, 2018.
- [36] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, pp. 372–387, 2016.
- [37] D. Li, T. Jiang, and M. Jiang, "Exploiting high-level semantics for no-reference image quality assessment of realistic blur images," in *Proc. 25th ACM Int. Conf. Multimedia*, 2017, pp. 378–386.
- [38] D. C. Van Essen and J. H. R. Maunsell, "Hierarchical organization and functional streams in the visual cortex," *Trends Neurosci.*, vol. 6, pp. 370–375, Jan. 1983.
- [39] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "Learning sparse representation for objective image retargeting quality assessment," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1276–1289, Apr. 2018.
- [40] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, 2009, pp. 248–255.
- [41] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 586–595.
- [42] R. Hu, V. Monebhurrun, R. Himeno, H. Yokota, and F. Costen, "An adaptive least angle regression method for uncertainty quantification in FDTD computation," *IEEE Trans. Antennas Propag.*, vol. 66, no. 12, pp. 7188–7197, Dec. 2018.
- [43] R. Hu, Y. Liu, Z. Wang, and X. Li, "Blind quality assessment of nighttime image," *Displays*, vol. 69, Sep. 2021, Art. no. 102045.
- [44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [45] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size," 2016, *arXiv:1602.07360*.
- [46] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, 2015, pp. 1–9.
- [47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.
- [48] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.
- [49] D. Zoran and Y. Weiss, "Scale invariance and noise in natural images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Kyoto, Japan, 2009, pp. 2209–2216.
- [50] 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, pp. 934–945, 2012.
- [51] 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.
- [52] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 45, pp. 485–560, Jun. 1992.

- [53] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.
- [54] 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, pp. 390–402, 2015.
- [55] S. Gu, J. Bao, D. Chen, and F. Wen, "GIQA: Generated image quality assessment," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*, Aug. 2020, pp. 369–385.
- [56] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Kyoto, Japan, 2009, pp. 2106–2113.
- [57] L. Zhang, Z. Gu, and H. Li, "SDSP: A novel saliency detection method by combining simple priors," in *Proc. IEEE Int. Conf. Image Process.*, Melbourne, VIC, Australia, 2013, pp. 171–175.
- [58] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.
- [59] 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.
- [60] Y. Liu, K. Gu, X. Li, and Y. Zhang, "Blind image quality assessment by natural scene statistics and perceptual characteristics," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 16, no. 3, pp. 1–91, 2020.
- [61] 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.
- [62] X. Min, G. Zhai, K. Gu, Y. Liu, and X. Yang, "Blind image quality estimation via distortion aggravation," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 508–517, Jun. 2018.
- [63] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 36–47, Jan. 2020.
- [64] S. Su *et al.*, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, USA, 2020, pp. 3667–3676.
- [65] 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 COM 9-80-E, 2000.
- [66] Q. Wu, H. Li, F. Meng, and K. N. Ngan, "A perceptually weighted rank correlation indicator for objective image quality assessment," *IEEE Trans. Image Process.*, vol. 27, pp. 2499–2513, 2018.
- [67] R. Hu, V. Monebhurrun, R. Himeno, H. Yokota, and F. Costen, "A statistical parsimony method for uncertainty quantification of FDTD computation based on the PCA and ridge regression," *IEEE Trans. Antennas Propag.*, vol. 67, no. 7, pp. 4726–4737, Jul. 2019.
- [68] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2818–2826.
- [69] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 2261–2269.
- [70] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 6848–6856.
- [71] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst. Vol. 1*, 2012, pp. 1097–1105.



Runze Hu (Member, IEEE) received the B.S. degree in computer science from North China Electric Power University, Baoding, China, in 2014, and the M.Sc. degree in computer science and the Ph.D. degree in electrical and electronics engineering from the University of Manchester, Manchester, U.K., in 2016 and 2020, respectively.

His current research interests include artificial intelligence, uncertainty quantification techniques, and image processing.



Yutao Liu received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2011, 2013, and 2018, respectively.

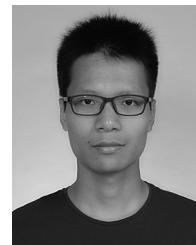
From 2014 to 2016, he was with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China, as a Research Assistant. From 2018 to 2021, he was a Postdoctoral Fellow with Tsinghua University, Beijing, China. He is currently an Associate Professor with the Ocean University of China, Qingdao, China. His research interests include image quality assessment, perceptual image processing, and computer vision.



Ke Gu (Member, IEEE) 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 a Professor with Beijing University of Technology, Beijing, China. His research interests include environmental perception, image processing, quality assessment, and machine learning.

Prof. Gu received the Best Paper Award from the IEEE TRANSACTIONS ON MULTIMEDIA and the Best Student Paper Award at the IEEE International Conference on Multimedia and Expo in 2016. He was the Leading Special Session Organizer in the VCIP 2016 and the ICIP 2017 and serves as a Guest Editor for *Digital Signal Processing*. He is currently an Associate Editor for SPIC, Displays, Entropy, and IET-IPR. He is a reviewer for 20 top SCI journals.



Xiongkuo Min (Member, IEEE) received the B.E. degree from Wuhan University, Wuhan, China, in 2013, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2018.

From January 2016 to January 2017, he was a visiting student with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He is currently a Postdoctoral Fellow with Shanghai Jiao Tong University. His research interests include visual quality assessment, visual attention modeling, and perceptual signal processing.

Dr. Min received the Best Student Paper Award at IEEE ICME 2016.



Guangtao Zhai (Senior Member, IEEE) received the B.E. and M.E. degrees from Shandong University, Shandong, China, in 2001 and 2004, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2009.

He is currently a Research Professor with the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University. From 2008 to 2009, he was a visiting student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he was a Postdoctoral Fellow from 2010 to 2012. From 2012 to 2013, he was a Humboldt Research Fellow with the Institute of Multimedia Communication and Signal Processing, Friedrich Alexander University of Erlangen-Nuremberg, Erlangen, Germany. His research interests include multimedia signal processing and perceptual signal processing.

Dr. Zhai received the Award of National Excellent Ph.D. Thesis from the Ministry of Education of China in 2012.