Blind Image Quality Assessment Using Local Consistency Aware Retriever and Uncertainty Aware Evaluator

Qingbo Wu, *Member, IEEE*, Hongliang Li, *Senior Member, IEEE*, King N. Ngan, *Fellow, IEEE*, and Kede Ma, *Student Member, IEEE*

***Abstract*— Blind image quality assessment (BIQA) aims to automatically predict the perceptual quality of a digital image without accessing its pristine reference. Previous studies mainly focus on extracting various quality-relevant image features. By contrast, the explorations on highly efficient learning model are still very limited. Motivated by the fact that it is difficult to approximate a complex and large data set via a global parametric model, we propose a novel local learning method for BIQA to improve quality prediction performance. More specifically, we search for the perceptually similar neighbors of a test image to serve as its unique training set. Unlike the widely used k nearest neighbors principle, which only measures the similarity between the testing and training samples, the local consistency of the selected training data is also considered to generate smoother sample space. The image quality is estimated via a sparse Gaussian process. As an additional benefit, the uncertainty of the predicted score is jointly inferred, which can subsequently drive more robust perceptual image processing applications, such as deblocking investigated in this paper. Extensive experiments demonstrate that the proposed learning model leads to consistent quality prediction improvements over many state-of-the-art BIQA algorithms.**

***Index Terms*— Blind image quality assessment, local consistency aware retriever, uncertainty aware evaluator.**

I. Introduction

ACCURATELY predicting the human perception of image quality is of fundamental importance in evaluating and improving the user experience of the multimedia systems. In the past decades, booming development of digital imaging and internet technologies has prompted the extreme growth of internet pictures, which makes the subjective evaluation impractical and expensive. As a result, the objective image quality assessment (IQA) models that can automatically assess the perceptual quality of digital images are highly desirable, which can also be subsequently deployed to monitor image acquisition, optimize image processing systems and design perception-friendly display devices [1]–[7].

Recently, the blind image quality assessment (BIQA) has become an active research area due to the inaccessibility of pristine reference image in many real-world applications. Early research on BIQA usually assumes that the distortion type is known. Then, the image features or measures are designed to quantify the obviousness of specific distortion artifacts, such as, blocking [8], [9], ringing [10], [11] and blurring [12]–[14]. Since these models are developed for some specific distortion types, their application scope would be fairly limited.

Therefore, more and more researchers turn their attention to the general purpose BIQA, where no assumption about distortion type is needed. Through unremitting efforts in recent years, many representative general purpose BIQA algorithms [15]–[23] have been proposed, which deliver good prediction performance via a supervised learning framework. Particularly, during the training phase, these approaches map the labeled images to a quality-relevant feature space, and utilize a parametric regression function to approximate the distribution of all training samples. In the test phase, the test image is also mapped to the same feature space, and then loaded to the offline trained regression model, which produces a continuous score to indicate the predicted quality.

Under the same training framework, recent BIQA methods mainly focus on developing a variety of quality-aware image features. Moorthy and Bovik [15], [19] modeled the distribution of wavelet coefficients and used estimated parameters to summarize the natural scene statistics (NSS) that are assumed to be quality-relevant. Saad *et al.* [16] explored the statistics of DCT coefficients to measure the image quality variation. Mittal *et al.* [18] found that the NSS of the locally normalized intensity coefficients are highly correlated with human perception. Xue *et al.* [22] extended the local contrast analysis by computing the joint statistics of gradient magnitude and Laplacian of Gaussian responses. Gu *et al.* [20] combined three types of image features based on a free energy theory. Wu *et al.* [23] proposed the multi-channel fused images features to simulate the hierarchical structure of human vision system. Zhang *et al.* [24] utilized the semantic

Manuscript received June 11, 2016; revised December 7, 2016; accepted May 24, 2017. Date of publication June 1, 2017; date of current version September 13, 2018. This work was supported by the National Natural Science Foundation of China under Grant 61601102, Grant 61525102, and Grant 61502084. This paper was recommended by Associate Editor D. Mukherjee. *(Corresponding author: Qingbo Wu.)*

Q. Wu and H. Li are with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: wqb.uestc@gmail.com; hlli@uestc.edu.cn).

K. N. Ngan is with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, and also with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: knngan@ee.cuhk.edu.hk).

K. Ma is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: k29ma@uwaterloo.ca).

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/TCSVT.2017.2710419

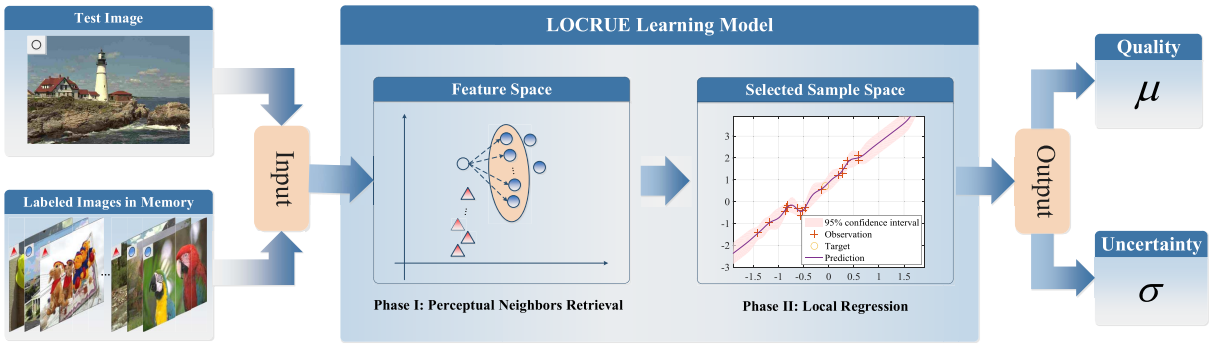


Fig. 1. Diagram of the proposed LOCRUE learning model. The symbol of an image in the feature space has been labeled in its top left corner. The symbols with the same shape and color share the equal perceptual quality, and vice versa. The local regression produces two outputs, *i.e.*, the predicted quality score μ and its associated uncertainty value σ.

obviousness to enhance the performance of low-level NSS features. In addition, some learning based image representations are also discussed in [25]–[27]. Benefitting from these studies, more and more quality relevant visual information has been developed to constantly increase the prediction accuracy of BIQA models.

However, not as active as the research on image feature extraction, only very limited learning models are explored for the BIQA task, even though its performance is crucial in determining the quality prediction accuracy. In existing approaches [15], [18]–[22], [25], [26], the support vector regression (SVR) [28], [29] dominates the prediction function learning procedure, which is acknowledged as an efficient tool in processing high-dimensional data [29]–[31]. Gao *et al.* [21] improved the regression performance by incorporating a feature fusion scheme – multiple kernel learning [32], [33] into the SVR. In addition to the SVR-like geometric approaches [34], [35], some statistical learning models are also adopted. Saad *et al.* [16] investigated the performance of a simple probabilistic predictive model based on the naive Bayes theory. Limited by the accuracy in fitting the joint distribution of the high dimensional features and the image quality labels [36]–[38], this probabilistic model does not bring consistent performance gains compared with the SVR. Li *et al.* [17] tried to learn a quality evaluator with a generalized regression neural network, which delivers better prediction performance [17]. To achieve more elaborate inference, Hou *et al.* [39] introduced additional hidden variables in the probabilistic model, which generates a deeper network. With the help of the well-designed hierarchical framework, this statistical learning model could estimate the image quality more accurately. Recently, the studies on deep neural networks (DNN) are also discussed for BIQA, and promising results have been reported in [40] and [41]. Although the DNN is quite popular in other research areas such as image classification and object detection, its huge computational load and data hungry property bring great challenges in implementing BIQA on limited computational resources and subject-rated images.

The predictive models mentioned above are all developed based on global learning, which optimizes a single parametric function by minimizing the mean fitting error across all training samples. An appealing aspect of this scheme lies in its one-pass training process, which could store the offline trained model in a small memory and directly apply it to all test samples [42]. However, when the training data present a complex distribution, the adoption of global learning is usually computationally intensive and analytically intractable [42], [43]. For these reasons, an alternative solution – local or lazy learning becomes more attractive, which could accelerate the learning procedure and even achieve more accurate approximation [44]–[48]. Particularly, the local learning follows a divide-and-conquer strategy, and decomposes the complex global fitting problem into a simpler query-specific approximation [42], [43], which tries to interpret each test sample with its local neighbors. Inspired by aforementioned studies, Wu *et al.* [23] investigated a simple local learning instance, which is referred to as label transfer (LT), to estimate the quality of an image by pooling the quality labels of its k nearest neighbors (kNN). Although developed from a straightforward distance based pooling scheme, the LT achieved competitive prediction performance, which encourages us to conduct a more detailed exploration of local learning.

In this paper, we propose a novel local learning model, which is characterized by utilizing a LOcal Consistency-aware Retriever and an Uncertainty-aware Evaluator (LOCRUE). As shown in Fig. 1, we tackle the BIQA in two phases. In Phase I, the perceptual neighbors of a test image are first retrieved from all labeled samples in the feature space, which constructs a query-specific local training set. In Phase II, the regression function is learned from previous selected training samples, and produces the estimated quality score μ and its associated uncertainty value σ for current test image. In comparison with the previous LT scheme [23], the proposed LOCRUE shows significant advantages, which are summarized into the following three points.

1) *Probabilistic Interpretation*: A posterior probability model is developed to measure the perceptual similarity of two images with the given feature distance. Specifically, we collect the ternary results from a subjective test by comparing the quality of two distorted images, in which the subjects are required to judge which image is better or they are perceptually equal to each other. Then, the probabilistic relationship between the feature

distance and perceptual similarity is learned by kernel density estimate (KDE) [49], which provides more clear and accurate similarity interpretation than LT [23].

2) *Local Consistency*: A local consistency constraint is embedded into the perceptual neighbor retrieval process. Besides minimizing the dissimilarity of the feature vectors between the test image and its selected labeled samples, the proposed model also tries to guarantee that the local neighbors share close image quality, which could produce a smoother sample space and boost the local learning.
3) *Predictive Uncertainty*: Due to the fact that humans naturally show uncertainty in making decisions [50]–[55], we integrate both the subjective mean opinion scores (MOS) and their associated standard derivations across multiple subjects into the learning model via the sparse Gaussian process (GP) [56]. By doing so, we are able to predict the image quality and its associated uncertainty, which are measured by the mean and standard derivation values of the estimated GP. This two dimensional output is more consistent with human judgements, and very beneficial for improving the robustness of perception-driven applications.

Through experiments on LIVE II [57], TID2013 [58], VCL@FER [59], and CSIQ [60] databases, it is verified that the proposed learning model efficiently improves the prediction performance of existing BIQA algorithms as using the same set of features as input. Moreover, it is very robust to small training sets.

The rest of this paper is organized as follows. We introduce the proposed LOCRUE model in Section II. The experimental results are presented in Section III. Section IV discusses a perception-driven deblocking application based on LOCRUE. Finally, we conclude this paper in Section V.

II. Methodology

As discussed in the introduction section, the local learning aims at interpreting a test or query image by its local neighbors, which consists of perceptual neighbors retrieval and local regression processes. In this section, we describe how to generate a smooth sample space by combing a probabilistic similarity measure with the local human opinion consistency constraint. In the following, a sparse Gaussian process based probabilistic model is introduced to implement the local regression which produces both the image quality score and its associated uncertainty value.

A. Local Consistency Aware Perceptual Neighbors Retrieval

1) Probabilistic Similarity Measure: Let I_q and I_c denote the query image and a candidate image, respectively. Meanwhile, let x_q and x_c denote their feature vectors. In a typical image retrieval system, the similarity of these two images is usually measured by the difference of x_q and x_c in terms of a distance metric, which is denoted by $D(x_q, x_c)$ [61]–[63]. Particularly, a smaller $D(x_q, x_c)$ indicates a higher similarity between I_q and I_c.

Due to the semantic gap, this feature distance based similarity measurement usually performs poorly with irrelevant image features, that provide misleading information or retrieval results for the query image [64]. To address this problem, we develop an efficient posterior probability model to measure the similarity of two images with given feature distance.

Let l denote a similarity related binary label. We set l to 1 if the perceptual quality of I_q and I_c are equal to each other. Otherwise, l is set 0. Our target is to estimate the posterior probability distribution of l with given feature distance $D(x_q, x_c)$, *i.e.*, $p\big(l|D(x_q, x_c)\big)$. For short, we use D to represent $D(x_q, x_c)$ in the following text. Based on the Bayes' theorem, $p(l|D)$ can be expressed as

$$p(l|D) = \frac{p(D|l) \cdot p(l)}{\int_{-\infty}^{\infty} p(D|l)p(l)dl}. \tag{1}$$

Training the likehood $p(D|l)$ and the prior $p(l)$ requires the semantic labels l from human subjects. Thanks to the large-scale database in our previous works [65], [66], we have collected 1440 groups of ternary labels in the pairwise comparisons of two distorted images across 30 subjects. More specifically, each subject is asked to judge which image is better or if they are perceptually equal to each other. In this paper, we set l to 0 if more than 15 subjects believe that a pair of images are perceptually different.

With the help of our labeled pairwise training data, we can obtain the raw density $T_{D,l}$, which is given by

$$T_{D,l} = \frac{1}{N_l} \sum_{i=1}^{N_l} \delta(D - d_i) \tag{2}$$

where N_l denotes the number of pairwise samples with the label l, d_i is the feature distance of the ith pair of images, and $\delta(\cdot)$ is the Dirac delta function.

To deduce the probability distribution $p(D|l)$ from previous finite samples, the kernel density estimation (KDE) [49] is employed here, which represents the estimated probability distribution as

$$\hat{p}(D|l) = \int_{-\infty}^{\infty} T_{D-s,l} k(s) ds \tag{3}$$

where $k(s)$ is a kernel function for smoothing the sampled raw density. In our implementation, the widely used Gaussian kernel is employed, *i.e.*,

$$k(s) = \frac{1}{\sqrt{2\pi}\omega} e^{-\frac{s^2}{2\omega^2}} \tag{4}$$

where ω is the variable kernel bandwidth whose optimization can be found in [49].

For clarity, a KDE example for inferring the distribution of $p(D|l = 1)$ has been show in Fig. 2, where the quality aware features from [18] are used for image representation and the Euclidean distance is employed in calculating D. By substituting Eq. (3) into Eq. (1), we obtain the probabilistic description for measuring the perceptual equality of two images. Particularly, a larger $p(l|D)$ means a higher confidence in making the judgement for l. It is seen that $p(D|l = 1)$ is very small when D approaches zero. This result is consistent with our assumption about the existence of

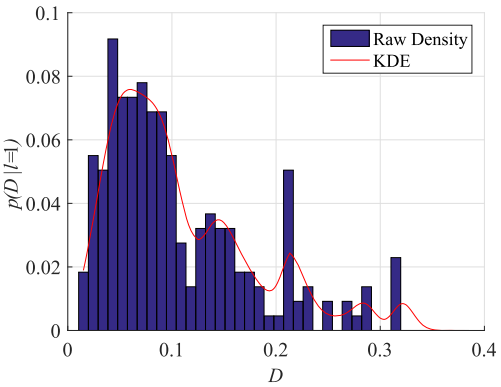


Fig. 2. An example of KDE result for the likehood $p(D|l=1)$.

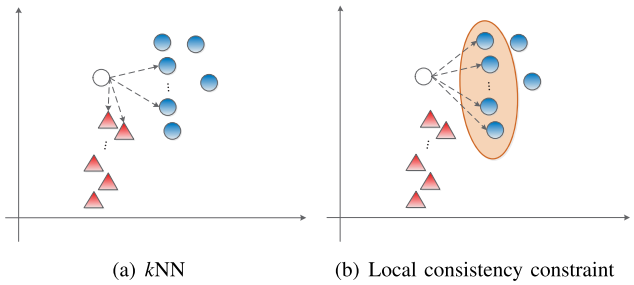


Fig. 3. A toy example of the comparison between different image retrieval schemes. The symbols with the same shape share the similar perceptual quality, and vice versa. (a) kNN. (b) Local consistency constraint.

semantic gap between the image feature and its semantic label. That is, a smaller D does not necessarily lead to more similar perceptual quality between two samples, and the probabilistic model could provide a reliable similarity description.

2) Local Consistency Based Image Retrieval: With the proposed similarity measure, we would discuss how to collect the perceptual neighbors of a test image. Specifically, our target is to separate all candidate training images into two compact groups, in which the selected samples are labeled by $l=1$ and the unselected samples are labeled by $l=0$.

In comparison with the classical kNN scheme, the significant difference of our image retrieval method lies in its local consistency constraint. For clarity, a toy example has been show in Fig. 3. Since the kNN only measures the differences between the testing and training samples, some perceptually dissimilar data are easily selected when they are close to the test image in the feature space. To address this problem, we introduce the penalty for the dissimilarity of all selected samples, which are bounded by the ellipse in Fig. 3 (b). In this way, a trade-off could be made between approaching the test image and keeping local consistency among the selected candidates.

As discussed in many computer vision works [67]–[69], this local consistency constrained labeling task can be formulated as a graph based energy minimization problem. In this study, the set of all candidate training images can be represented as a graph, and each candidate training sample is denote by a node. The diagram of the label decision process is shown in Fig. 4. There are two additional terminal nodes *sink* $\mathcal{T}$

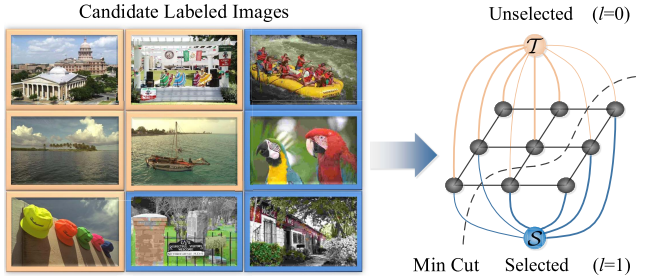


Fig. 4. Perceptual neighbors retrieval using graph cut. The cost of each edge is represented by its thickness. The dashed curve represents the minimum cut by separating the nodes into $\mathcal{S}$ and $\mathcal{T}$.

and *source* $\mathcal{S}$ located on the top and bottom of this graph, respectively. The node $\mathcal{T}$ represents the decision "unselected" (*i.e.*, $l=0$) and the node $\mathcal{S}$ denotes the decision "selected" (*i.e.*, $l=1$). They are linked with the candidate image nodes by the colorful edges whose capacities Φ_i correspond to the penalty for assigning the terminal labels (*i.e.*, $l=0$ or 1) to the ith candidate image. Meanwhile, the candidate image nodes are also linked with each other by the black edges whose capacities $\Psi_{i,j}$ correspond to the penalty for the inconsistency between two neighbors.

Let l_i denote the label assigned to the ith candidate image and $L=\{l_1,\cdots,l_N\}$ denote the label set for the whole graph. The graph based energy function can be represented by

$$E(L)=\sum_{i=1}^{N}\Phi_i(l_i)+\sum_{i,j\in\Omega}\Psi_{i,j}(l_i,l_j)\cdot\delta(l_i-l_j) \tag{5}$$

where Ω is the set of neighboring candidate images.

For our perceptual neighbors retrieval task, we utilize the negative log of the proposed similarity measure to describe the cost of assigning the label l_i to the current node, which corresponds to the unary term in Eq. (5), *i.e.*,

$$\Phi_i(l_i)=-\ln\left(p(l_i|D_i)\right) \tag{6}$$

where D_i denotes the feature distance between the ith candidate image and the test image, and a higher conditional probability $p(l_i|D_i)$ would produce a smaller cost $\Phi_i(l_i)$.

The pairwise term $\Psi_{i,j}$ in Eq. (5) corresponds to our local consistency constraint which prefers collecting the training data with similar perceptual quality. In the case of the BIQA task, the high-level semantic information (*i.e.*, human opinion scores) are available for all training images. Therefore, we can accurately measure the inconsistency of two candidate images with the difference between their image quality, *i.e.*,

$$\Psi_{i,j}=|S_i-S_j| \tag{7}$$

where S_i and S_j denote the human opinion scores of the ith and jth candidates, respectively.

After building the energy function, we can solve the labels L based on the maximum flow/minimum cut algorithm [67]. Then, all the candidates labeled by "selected" are used for the local training data of the test image. We illustrate the top-5 image retrieval results using different schemes in Fig. 5, where the quality-aware features from [18] are used for image

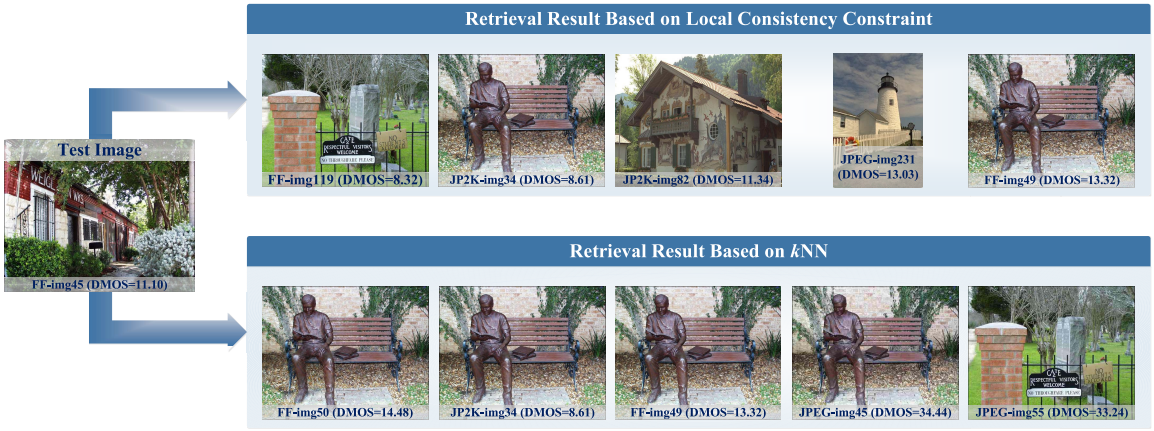


Fig. 5. Top-5 image retrieval results based on two different schemes. From left to right, the similarity between the test image and the selected candidate gradually reduces in terms of two retrieval schemes. All image names and their associated DMOSs are labeled on the bottom.

representation. Both the test and candidate images are from the LIVE II database [57]. It is clear that our local consistency constrained scheme generates smoother perceptual neighbors, whose mean DMOSs are very close to the test image and present small variations. By contrast, the kNN scheme collects five perceptually different neighbors, even they show the similar or the same visual contents.

B. Sparse Gaussian Process Based Local Regression

With the locally consistent training data, we further discuss how to learn a query-specific regressor. In view of the uncertainty of humans in making decisions [50]–[52], we consider all labeled training data as noisy observations. For the ith training image, let y_i denote its noisy target in terms of the subjective opinion score, x_i denote its feature vector, and $f(\cdot)$ represent the latent predictive function that we are going to learn. Without loss of generality, the noise ε_i on y_i is assumed to follow an additive Gaussian process, which can be represented by

$$y_i = f(x_i) + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma_u) \tag{8}$$

where σ_u denotes the standard derivation of the noise induced by the uncertainty of human subjects, and the expectation value $E(y_i)$ corresponds to the MOS or DMOS assigned to the ith training sample.

As discussed in [70] and [71], the Gaussian process regression could efficiently work with the noisy training data via a probabilistic framework, which considers the predictive function f following a multivariate Gaussian distribution, *i.e.*,

$$p(\mathbf{f}|\mathbf{X}) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K}) \tag{9}$$

where $\mathbf{f} = \{f_1, \cdots, f_N\}$ and $\mathbf{X} = \{x_1, \cdots, x_N\}$. $\mathbf{K}$ is a $N \times N$ covariance matrix, whose entries $\mathbf{K}(i, j)$ are derived from the parametric kernel function $\mathbb{K}_{\boldsymbol{\Theta}}(x_i, x_j)$, *i.e.*,

$$\mathbb{K}_{\boldsymbol{\Theta}}(x_i, x_j) = \alpha \exp(-\frac{1}{2}\|\beta^T \cdot (x_i - x_j)\|^2) \tag{10}$$

where the scalar α and vector β construct the hyperparameters $\boldsymbol{\Theta} = \{\alpha, \beta\}$ of a GP.

The training process aims to solve the optimal $\boldsymbol{\Theta}$ which can maximize the log of the following marginal likelihood

$$p(\mathbf{y}|\mathbf{X}, \theta) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K} + \sigma_u^2\mathbf{I}) \tag{11}$$

where $\mathbf{y} = \{y_1, \cdots, y_N\}$ and $\mathbf{I}$ is an identity matrix.

The high computational load limits the usage of this standard GP regression in our local learning model, which requires a high efficiency online training. To accelerate the training process, the sparse GP discussed in [56] is utilized to replace the marginal likelihood in Eq. (11) by

$$p(\mathbf{y}|\mathbf{X}, \hat{\mathbf{X}}, \boldsymbol{\Theta}) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{K}_{NM}\mathbf{K}_M^{-1}\mathbf{K}_{MN} + \boldsymbol{\Lambda} + \sigma_u^2\mathbf{I}) \tag{12}$$

where $\hat{\mathbf{X}} = \{\hat{x}_1, \cdots, \hat{x}_M\}$ is the pseudo input with M elements and satisfies $M < N$. $\mathbf{K}_M$ is a $M \times M$ matrix, whose element is given by $\mathbb{K}_{\boldsymbol{\Theta}}(\hat{x}_i, \hat{x}_j)$. $\mathbf{K}_{NM}$ is a $N \times M$ matrix whose element is calculated by $\mathbb{K}_{\boldsymbol{\Theta}}(x_i, \hat{x}_j)$. Correspondingly, $\mathbf{K}_{MN} = \mathbf{K}_{NM}^T$. $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \cdots, \lambda_N)$ is a diagonal matrix, whose element is given by

$$\lambda_i = \mathbb{K}_{\boldsymbol{\Theta}}(x_i, x_i) - \mathbf{k}_i^T\mathbf{K}_M^{-1}\mathbf{k}_i \tag{13}$$

where $\mathbf{k}_i^T$ is the ith row vector of $\mathbf{K}_{NM}$.

By means of the sparse pseudo input, the training complexity in maximizing Eq. (12) could be reduced to $\mathcal{O}(M^2N)$ [56]. For a given test image, we use x_* and y_* to denote its feature vector and the associated human opinion score, respectively. The BIQA in our local regression is equivalent to calculate the mean μ_* and variance σ_*^2 of y_*, which can be given by

$$\begin{aligned} \mu_* &= \mathbf{k}_*^T\mathbf{Q}_M^{-1}\mathbf{K}_{MN}(\boldsymbol{\Lambda} + \sigma_u^2\mathbf{I})^{-1}\mathbf{y} \\ \sigma_*^2 &= \mathbb{K}_{\boldsymbol{\Theta}}(x_*, x_*) - \mathbf{k}_*^T(\mathbf{K}_M^{-1} - \mathbf{Q}_M^{-1})\mathbf{k}_* + \sigma_u^2 \end{aligned} \tag{14}$$

where $\mathbf{k}_*$ is a M-dimension column vector and its ith element is $\mathbb{K}_{\boldsymbol{\Theta}}(x_*, \hat{x}_i)$. In addition, $\mathbf{Q}_M = \mathbf{K}_M + \mathbf{K}_{MN}(\boldsymbol{\Lambda} + \sigma_u^2\mathbf{I})^{-1}\mathbf{K}_{NM}$.

More specifically, the μ_* represents our estimated subjective mean opinion score (*i.e.*, MOS or DMOS), and σ_*^2 is the associated uncertainty value. A higher σ_*^2 means a more significant variation across multiple subjects in rating the quality of an image.

III. Experimental Results

A. Databases and Protocols

We evaluate the performance of the proposed LOCRUE method on four publicly available IQA databases including LIVE II [57], TID2013 [58], VCL@FER [59], and CSIQ [60], in which both the subjective mean opinion scores (e.g., DMOS or MOS) and their associated standard derivations are assigned to each image.

The LIVE II database collects 29 pristine images and simulates five distortion types, *i.e.*, JPEG2000 compression (JP2K), JPEG compression (JPEG), additive white noise (WN), Gaussian blur (GB) and fast fading Rayleigh channel (FF), which generates 779 distorted images. The TID2013 database consists of 25 reference images and simulates 24 distortion types to obtain 3000 distorted images. In VCL@FER database, there are 23 reference images and 552 distorted versions contaminated by 4 distortion types including JP2K, JPEG, WN and GB. For CSIQ [60] database, 30 original images are collected and contaminated by six artifacts to generate 866 distorted images. It is noted that there are 19 overlapped reference images between LIVE II and TID2013, which share the same visual contents. For the VCL@FER and CSIQ databases, their reference images are completely different from each other and unlike the visual contents in both LIVE II and TID2013.

Similar to the criterion in [16], [18], [22], and [39], for the TID2013 and CSIQ databases, we only consider four common distortion types, *i.e.*, JP2K, JPEG, WN and GB, which also appear in both LIVE II and VCL@FER. Two commonly used measures are employed for evaluating the performance of different BIQA algorithms, *i.e.*, Spearman rank-order correlation coefficient (SRC) and Pearson linear correlation coefficient (LCC). In view of the nonlinearity induced by the subjective rating process, we follow the recommendations of video quality experts group (VQEG) [72], and map the predicted quality scores to the human opinion score via a four-parameter monotonic logistic function in calculating the SRC and LCC indices. Let μ denote the predicted quality score and μ_m denote its nonlinear mapping result, the nonlinear mapping function is given by

$$\mu_m = \beta_2 + \frac{\beta_1 - \beta_2}{1 + \exp(-\frac{\mu - \beta_3}{|\beta_4|})} \tag{15}$$

where $\beta_1 \sim \beta_4$ are the parameters to be fitted.

B. Implementation Details

Since we focus on developing high efficiency learning method, the BIQA is implemented by incorporating existing quality-aware image features into the proposed LOCRUE model. Particularly, the image features from the latest seven research are investigated, *i.e.*, BIQI [15], BLIINDS II [16], BRISQUE [18], CORNIA [25], M_3 [22], NFERM [20] and TCLT [23], which have achieved state-of-the-art prediction performance in the general purpose BIQA task.

For comparison, the most popular learning model – SVR is utilized as the baseline and the kernel selection follows the recommendations in the literatures [15], [16], [18], [20], [22], [23], [25]. According to the instruction in [73], the grid search is utilized to determine the optimal SVR parameters for each BIQA metric. In addition, the nonparametric LT method in [23] is also involved in this experiment. As suggested in [23], we set the neighbor number to 5 and utilize the chi-square distance metric to measure the similarity of two images. For our LOCRUE model, the noise variance σ_u^2 is determined by the mean variance of human opinion scores in the local training data, which are collected for each test image.

To learn the regression models for different BIQA metrics, we follow the same criteria in [15], [16], [18], [20], [22], [23], and [25] and partition each database into the non-overlapped training and testing sets. More specifically, we randomly select part of the reference images and their distorted versions to build the training set. The rest images in each database are used for testing, which ensures that the test images present different visual contents with respect to the training samples. In this section, we investigate the performance of a learning model under three partition sizes, whose training set would occupy 80%, 50% and 20% of all images in each database. We repeat each random partition 100 times on each IQA database. Then, the median SRC and LCC are reported for evaluation.

C. Effect of Pseudo Input M

As discussed in Section II-B, the sparse GP is utilized in our method to balance the regression accuracy and computational complexity with a parameter M. To investigate the effect of M in LOCRUE, we test ten possible values of $M = s \cdot N$, N is the total number of perceptual neighbors and $s \in \{0.1, 0.2, \cdots, 1\}$ is a scale factor. Both the SRC performance and running time are reported under each s, which are represented by SRC_s and T_s, representatively. To facilitate the observation, two relative measures are employed,

$$\begin{aligned} \Delta\mathrm{SRC}_s &= \mathrm{SRC}_1 - \mathrm{SRC}_s \\ \Delta T_s &= T_s / T_1 \end{aligned} \tag{16}$$

where a smaller $\Delta\mathrm{SRC}_s$ means that the accuracy of sparse GP regression is closer to the standard GP with the setting of $M = s \cdot N$. Meanwhile, a smaller ΔT_s means the training time of the sparse GP is less than the standard one.

We repeat 100 times of random train-test splitting tests on the LIVE II database, and the training set takes up 80% of all images. The median SRC and running time across 100 trials are used for SRC_s and T_s. The results are shown in Figs. 6 and 7. For all quality-aware features, the regression accuracy of sparse GP gradually approaches to the standard GP with an increasing M, where $\Delta\mathrm{SRC}_s$ gets close to 0 as s grows from 0 to 1 in Fig. 6. Unsurprisingly, the computational complexity of sparse GP also increases with M, where ΔT_s goes up with s in Fig. 7. From Fig. 6, we can find that the $\Delta\mathrm{SRC}_s$ converges when $s = 0.8$. Meanwhile, the running time of sparse GP has reduced by 40% with respect to the standard GP as shown in Fig. 7. To keep a good balance between the regression accuracy and the computational complexity, we set the pseudo input M to $0.8 \cdot N$ in the following experiments.

TABLE I
MEDIAN SRC AND LCC RESULTS ACROSS 100 TRIALS USING 80% DATA FOR TRAINING

Database	Method	BIQI		BLIINDS II		BRISQUE		CORNIA		M_3		NFERM		TCLT	
		SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC
LIVE II	SVR	0.800	0.808	0.926	0.924	0.936	0.942	0.940	0.931	0.946	0.942	0.946	0.952	0.933	0.923
	LT	0.765	0.755	0.924	0.932	0.925	0.932	0.937	0.933	0.945	0.953	0.948	**0.954**	0.930	0.933
	LOCRUE	**0.822**	**0.816**	**0.948**	**0.953**	**0.941**	**0.945**	**0.948**	**0.945**	**0.949**	**0.957**	**0.952**	0.940	**0.950**	**0.952**
TID2013	SVR	0.758	0.796	0.904	0.919	0.930	0.924	0.930	0.925	0.930	0.936	0.930	0.938	0.902	0.907
	LT	0.774	0.803	0.906	0.917	0.926	0.937	0.922	0.917	0.934	0.951	**0.940**	0.950	0.906	0.889
	LOCRUE	**0.804**	**0.809**	**0.912**	**0.931**	**0.935**	**0.943**	**0.937**	**0.932**	**0.936**	**0.947**	**0.940**	**0.951**	**0.938**	**0.947**
VCL@FER	SVR	0.749	0.732	0.861	0.854	0.859	0.839	0.890	0.886	0.896	0.878	0.898	0.893	0.913	0.910
	LT	0.765	0.752	0.865	0.850	0.876	0.862	0.881	0.883	0.889	0.884	0.898	0.886	0.878	0.874
	LOCRUE	**0.797**	**0.770**	**0.879**	**0.870**	**0.899**	**0.895**	**0.900**	**0.908**	**0.918**	**0.914**	**0.917**	**0.911**	**0.924**	**0.921**
CSIQ	SVR	0.681	0.680	0.926	0.945	0.934	0.940	0.936	0.946	0.945	**0.960**	0.932	**0.958**	0.928	0.947
	LT	0.658	0.695	0.919	0.934	0.930	0.943	0.920	0.937	0.935	0.950	0.924	0.955	0.931	0.945
	LOCRUE	**0.712**	**0.757**	**0.937**	**0.948**	**0.942**	**0.948**	**0.944**	**0.951**	**0.952**	**0.960**	**0.948**	**0.958**	**0.942**	**0.950**

TABLE II
MEDIAN SRC AND LCC RESULTS ACROSS 100 TRIALS USING 50% DATA FOR TRAINING

Database	Method	BIQI		BLIINDS II		BRISQUE		CORNIA		M_3		NFERM		TCLT	
		SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC
LIVE II	SVR	0.731	0.730	0.901	0.894	0.909	0.919	0.910	0.911	0.921	0.911	0.926	0.920	0.917	0.917
	LT	0.694	0.697	0.899	0.906	0.900	0.906	0.917	0.913	0.925	0.932	0.925	**0.930**	0.903	0.895
	LOCRUE	**0.743**	**0.735**	**0.913**	**0.916**	**0.921**	**0.926**	**0.929**	**0.935**	**0.929**	**0.934**	**0.932**	0.914	**0.928**	**0.925**
TID2013	SVR	0.717	0.730	0.882	0.891	0.904	0.892	0.903	0.902	0.912	0.916	0.909	0.922	0.883	0.885
	LT	0.750	**0.784**	0.888	0.907	0.904	**0.914**	0.909	0.910	**0.925**	0.934	0.924	0.917	0.865	0.823
	LOCRUE	**0.752**	0.775	**0.900**	**0.909**	**0.912**	0.906	**0.921**	**0.919**	**0.925**	**0.938**	**0.925**	**0.931**	**0.927**	**0.934**
VCL@FER	SVR	0.667	0.659	0.812	0.795	0.798	0.784	0.867	0.876	0.869	0.853	0.872	0.866	0.854	0.853
	LT	0.670	0.666	0.808	0.798	0.837	0.825	0.865	0.864	0.879	**0.873**	0.867	0.859	0.761	0.763
	LOCRUE	**0.696**	**0.700**	**0.829**	**0.819**	**0.851**	**0.849**	**0.887**	**0.892**	**0.883**	**0.873**	**0.891**	**0.885**	**0.888**	**0.887**
CSIQ	SVR	0.642	0.688	0.888	0.921	0.905	0.920	0.917	0.931	0.926	**0.949**	0.921	0.946	0.890	0.926
	LT	0.578	0.552	0.883	0.915	0.895	0.916	0.918	0.927	0.916	0.938	0.915	0.935	0.914	0.933
	LOCRUE	**0.666**	**0.709**	**0.917**	**0.928**	**0.921**	**0.931**	**0.929**	**0.939**	**0.933**	**0.949**	**0.934**	**0.954**	**0.922**	**0.942**

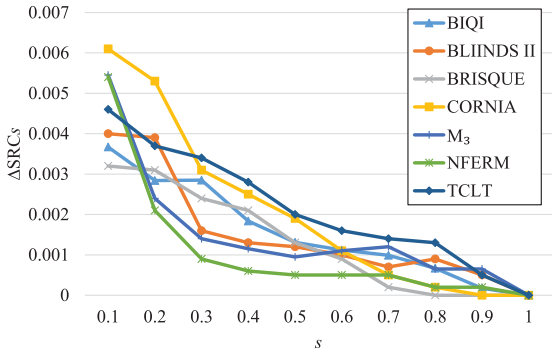


Fig. 6. The effect of M on the regression accuracy ($M = s \cdot N$).

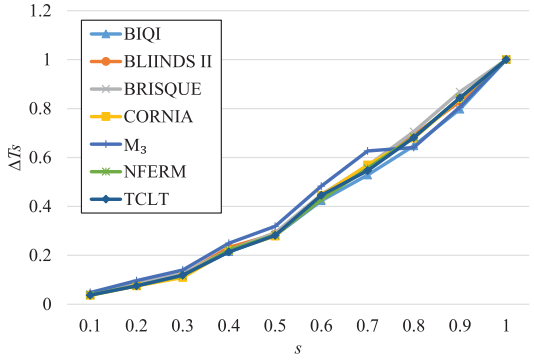


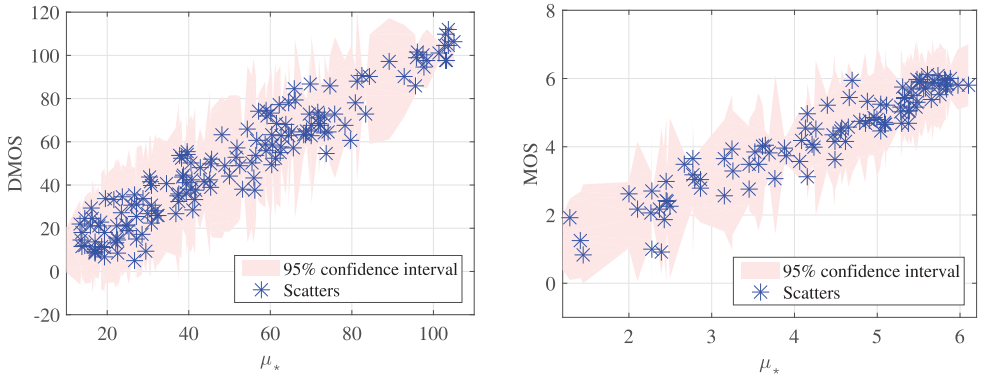
Fig. 7. The effect of M on the computational complexity ($M = s \cdot N$).

D. Consistency Evaluation

In order to evaluate the prediction accuracy of different learning methods, the median SRC and LCC results calculated under three training set sizes (*i.e.*, 80%, 50% and 20%) are shown in Tables I-III, respectively. For clarity, the best SRC and LCC results are highlighted by boldface in each column. It is seen that the SRC performance of our LOCRUE model outperforms both the SVR and LT methods across all IQA databases when incorporating with existing quality-aware image features. As shown in Table III, the highest SRC improvement between the LOCRUE and SVR is as large as 0.06, which is achieved by working with the BRISQUE feature on VCL@FER database. In comparison with the LT, the proposed LOCRUE model achieves a more significant SRC improvement in working with the TCLT feature on VCL@FER database, which is up to 0.127 as shown in Table II. Since the SVR focuses on minimizing the mean error across all training samples, the quality-irrelevant features would produce a more complex distribution in the feature space, which increases the difficulty in approximating all training samples with a

TABLE III
MEDIAN SRC AND LCC RESULTS ACROSS 100 TRIALS USING 20% DATA FOR TRAINING

Database	Method	BIQI		BLIINDS II		BRISQUE		CORNIA		M_3		NFERM		TCLT	
		SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC	SRC	LCC
LIVE II	SVR	0.539	0.530	0.837	0.826	0.857	0.868	0.871	0.870	0.872	0.850	0.861	0.846	0.864	0.864
	LT	0.570	**0.606**	0.853	0.863	0.861	0.870	0.864	0.876	0.887	0.895	0.875	**0.882**	0.874	0.869
	LOCRUE	**0.581**	0.596	**0.864**	**0.870**	**0.875**	**0.879**	**0.900**	**0.892**	**0.892**	**0.897**	**0.883**	0.877	**0.902**	**0.902**
TID2013	SVR	0.565	0.610	0.781	0.824	0.773	0.790	0.847	0.838	0.853	0.867	0.826	0.862	0.826	0.840
	LT	0.547	0.621	0.809	**0.826**	0.819	0.828	0.822	0.815	0.868	0.881	0.867	0.861	0.833	0.824
	LOCRUE	**0.600**	**0.641**	**0.815**	**0.826**	**0.838**	**0.832**	**0.869**	**0.873**	**0.872**	**0.891**	**0.869**	**0.878**	**0.864**	**0.879**
VCL@FER	SVR	0.592	0.596	0.730	0.728	0.725	0.681	0.792	0.781	0.764	0.705	0.764	0.759	0.800	0.796
	LT	0.562	0.571	0.696	0.688	0.762	0.752	0.805	0.816	0.776	**0.767**	0.800	0.794	0.739	0.722
	LOCRUE	**0.604**	**0.614**	**0.743**	**0.738**	**0.785**	**0.754**	**0.822**	**0.828**	**0.781**	0.732	**0.816**	**0.820**	**0.840**	**0.825**
CSIQ	SVR	0.559	0.572	0.800	0.842	0.856	**0.877**	0.863	0.886	0.861	0.893	0.858	0.880	0.848	0.870
	LT	0.579	**0.605**	0.828	**0.863**	0.854	0.873	0.856	0.836	0.882	**0.906**	0.842	0.860	0.875	0.891
	LOCRUE	**0.583**	0.570	**0.837**	0.854	**0.862**	0.876	**0.879**	**0.895**	**0.883**	0.889	**0.883**	**0.906**	**0.880**	**0.903**



(a) LIVE II (b) TID2013

(c) VCL@FER (d) CSIQ

Fig. 8. The scatter plots of the predicted quality index μ_* vs. the human opinion scores across four IQA databases.

global regression model. In contrast, the variation in the local region of the feature space is usually smoother, which facilitates the regression task in our LOCRUE model and delivers more accurate prediction. In addition, in comparison with our previous local learning model – LT, the proposed LOCRUE algorithm is superior in collecting perceptually similar neighbors, which is highly beneficial for learning a more accurate quality evaluator.

E. Uncertainty of Image Quality

As discussed in Section II-B, in addition to the predicted image quality, the proposed LOCRUE model also provides the uncertainty value σ_* for a given evaluation. More specifically, the estimated GP provides a 95% confidence interval $\mho_*$ for the estimated image quality μ_*, which is given by

$$\mho_* = [\mu_* - 2\sigma_*, \mu_* + 2\sigma_*]. \tag{17}$$

Fig. 8 shows the scatter plots of the predicted scores versus the DMOSs/MOSs across four databases, where the confidence interval $\mho_*$ has been highlighted by the magenta region. In this example, the BRISQUE feature is combined with our proposed LOCRUE model to estimate the image quality and the training set occupies 80% images in each database. It can be found

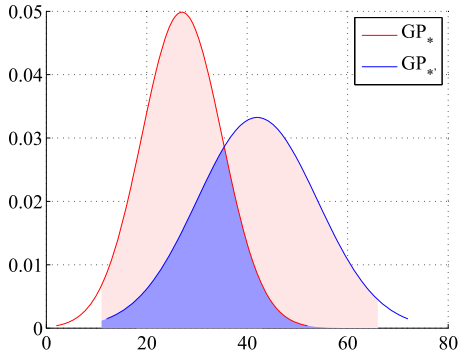


Fig. 9. Illustration of computing the overlap between the confidence intervals of two estimated GPs. The light blue region corresponds to the intersection $GP_*(\mho_*) \bigcap GP_{*'}(\mho_{*'})$. The combination of the light blue and red regions corresponds to the union $GP_*(\mho_*) \bigcup GP_{*'}(\mho_{*'})$.

that the region $\mho_*$ provides the alternative dynamic range for the quality prediction of an image. An important benefit of this confidence interval lies in its interpretability for the possibility of mistaking two perceptually similar image. More specifically, we don't require that two images are completely distinguishable in terms of a hard rank order. Alternatively, the possible rank is derived by the comparison between μ_* and $\mu_{*'}$. The possibility of mistaking these two images can be measured by their overlap $\mathcal{O}$ in terms of the confidence interval, *i.e.*,

$$\mathcal{O} = \frac{GP_*(\mho_*) \bigcap GP_{*'}(\mho_{*'})}{GP_*(\mho_*) \bigcup GP_{*'}(\mho_{*'})} \tag{18}$$

where GP_* and $GP_{*'}$ denote the probability density function for the estimated image quality μ_* and $\mu_{*'}$ respectively.

An illustration of $\mathcal{O}$ is shown in Fig. 9. A higher $\mathcal{O}$ means that the two images are more indistinguishable in terms of their perceptual quality. This overlap measurement is useful for detecting the confusing samples and improving the robustness of a perception-driven image processing system.

IV. Application to Perception-Driven Deblocking

To verify the effectiveness of the proposed uncertainty-aware BIQA method, we incorporate it into a perception-driven deblocking application. Particularly, the parametric shape adaptive DCT (SA-DCT) filter [74] is utilized to improve the perceptual quality of the compressed image via H.264/AVC codec.

In a conventional framework, a BIQA model works as a parameter evaluator (PE) to select the optimal filtering result which could produce the minimum estimated DMOS. In contrast, we implement the image filtering by a weighted fusion (WF) scheme, *i.e.*,

$$I_o = \sum_{i=1}^{N} w_i I_i \tag{19}$$

where I_o denotes the output image, $I_1 \sim I_N$ denote filtered images under different parameters, and w_i is the weight assigned to I_i.

TABLE IV
MS-SSIM Scores for Different Deblocking Schemes

Test Image	QP							
	38		42		46		50	
	PE	WF	PE	WF	PE	WF	PE	WF
bluegirl	0.9623	**0.9655**	0.9474	**0.9508**	0.9104	**0.9227**	0.8667	**0.8790**
hustler	**0.9638**	0.9609	0.9440	**0.9470**	0.9143	**0.9248**	0.8705	**0.8818**
lena	**0.9595**	0.9526	0.9335	**0.9342**	0.8959	**0.9033**	0.8418	**0.8525**
girl	0.9654	**0.9689**	0.9522	**0.9553**	0.9151	**0.9290**	0.8877	**0.9048**
flower	0.9596	**0.9623**	0.9431	**0.9464**	0.9133	**0.9181**	0.8516	**0.8604**
peppers	**0.9612**	0.9592	0.9400	**0.9442**	0.9096	**0.9207**	0.8571	**0.8762**
PartyScene	0.8804	**0.9282**	0.8589	**0.8899**	0.8035	**0.8153**	**0.6958**	0.6938
ducks_take_off	0.9510	**0.9696**	0.9330	**0.9427**	0.8779	**0.8863**	**0.7667**	0.7666
BQTerrace	0.8999	**0.9532**	0.8877	**0.9131**	0.8620	**0.8631**	0.8011	**0.8023**
rush_hour	0.9629	**0.9641**	0.9504	**0.9530**	0.9125	**0.9269**	0.8610	**0.8812**
Average	0.9466	**0.9585**	0.9290	**0.9377**	0.8915	**0.9010**	0.8300	**0.8399**

Let I_m denote the filtered image which produces the optimal estimated quality μ_m, where $I_m \in \{I_1, \cdots, I_N\}$. Let $\mathcal{O}_i$ denote the confidence interval overlap between I_m and I_i. Here, we assign a higher weight to the image whose quality rank is more easily mistaken with I_m. Then, the weight w_i can be defined by

$$w_i = \frac{\mathcal{O}_i}{\sum_{i=1}^{N} \mathcal{O}_i}. \tag{20}$$

For comparison, we use the SVR based BRISQUE metric to implement the PE based the image filtering, and the combination of BRISQUE features with our LOCRUE model is employed for conducting the WF scheme. Ten test images across different visual contents and resolutions are collected from two publicly available databases [75], [76], which are widely used for image processing and video coding. Then, the KTA2.4r1 [77] software is used to compress them under the Intra Only profile with the quantization setting QP = {38,42,46,50}. Similar with [18] and [23], the highly reliable full reference IQA metric MS-SSIM [78] is employed to measure the deblocking performances of PE and WF.

In Table IV, we show the detailed MS-SSIM scores for all test images, where the best results under each QP setting have been highlighted by boldface. It is seen that the proposed WF scheme delivers higher average MS-SSIM results across all QP settings in comparison with the traditional PE scheme. That is, we can produce human preferred deblocking results under different degradation degrees caused by H.264/AVC compression. For clarity, a visual comparison for the test image *Lena* compressed under the QP setting of 46, is shown in Fig. 10. It is seen that the PE based filtering result still presents obvious blocking artifact which is caused by mistaking the image quality processed with different parameters. Unlike the conventional method which tries to select one optimal filtering result, our uncertainty-aware evaluator provides a multi-frame fusion based framework to reduce the risk of selecting suboptimal parameter, which could generate better image quality as shown in Fig. 10 (d). The similar idea could also be extended to many other perception-driven image processing applications, such as, image contrast enhancement [79], [80], tone mapping for high dynamic range imaging [81], and color correction [82].

(a) Original image (b) H.264/AVC compression

(c) PE (d) WF

Fig. 10. Visual comparisons between different deblocking results. (a) Original image. (b) H.264/AVC compression. (c) PE. (d) WF

V. Conclusion

In this paper, we propose a novel BIQA algorithm which consists of a local consistency-aware retriever and the uncertainty-aware evaluator. Due to the advantage in capturing the smooth variation of a local region in the feature space, we can achieve better prediction performance in comparison with conventional global learning methods. In addition, by means of the probabilistic framework of our local regression model, the proposed method provides an efficient uncertainty description for the estimated image quality. Experiments on four benchmark IQA databases demonstrate that the proposed learning model efficiently improves the prediction accuracy of existing BIQA metrics. The application for auto deblocking also verifies its efficiency in improving the robustness of a perception-driven image processing system.

Acknowledgment

The authors would like to thank Prof. Zhou Wang for the valuable discussions about the uncertainty of human opinion score.

References

[1] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[2] Z. Wang, "Applications of objective image quality assessment methods [applications corner]," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 137–142, Nov. 2011.

[3] D. Schilling and P. C. Cosman, "Image quality evaluation based on recognition times for fast image browsing applications," *IEEE Trans. Multimedia*, vol. 4, no. 3, pp. 320–331, Sep. 2002.

[4] G.-S. Lin, Y.-T. Chang, and W.-N. Lie, "A framework of enhancing image steganography with picture quality optimization and anti-steganalysis based on simulated annealing algorithm," *IEEE Trans. Multimedia*, vol. 12, no. 5, pp. 345–357, Aug. 2010.

[5] F. Zhang, L. Ma, S. Li, and K. N. Ngan, "Practical image quality metric applied to image coding," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 615–624, Aug. 2011.

[6] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 838–842, Jul. 2015.

[7] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3951–3964, Aug. 2017.

[8] Z. Wang, A. C. Bovik, and B. L. Evan, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Sep. 2000, pp. 981–984.

[9] S. A. Golestaneh and D. M. Chandler, "No-reference quality assessment of JPEG images via a quality relevance map," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 155–158, Feb. 2014.

[10] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 1, pp. 1918–1927, Nov. 2005.

[11] H. Tao, 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.

[12] 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.

[13] T. Oh, J. Park, K. Seshadrinathan, S. Lee, and A. C. Bovik, "No-reference sharpness assessment of camera-shaken images by analysis of spectral structure," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5428–5439, Dec. 2014.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] X. Gao, F. Gao, D. Tao, and X. Li, "Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 12, pp. 2013–2026, Dec. 2013.

[22] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.

[23] 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.

[24] P. Zhang, W. Zhou, L. Wu, and H. Li, "SOM: Semantic obviousness metric for image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2394–2402.

[25] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1098–1105.

[26] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Real-time no-reference image quality assessment based on filter learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 987–994.

[27] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 995–1002.

[28] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.

[29] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[30] R. Clarke *et al.*, "The properties of high-dimensional data spaces: Implications for exploring gene and protein expression data," *Nature Rev. Cancer*, vol. 8, no. 1, pp. 37–49, 2008.

[31] V. V. Gavrishchaka and S. B. Ganguli, "Support vector machine as an efficient tool for high-dimensional data processing: Application to substorm forecasting," *J. Geophys. Res., Space Phys.*, vol. 106, no. A12, pp. 29911–29914, 2001.

[32] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *J. Mach. Learn. Res.*, vol. 12, pp. 2211–2268, Jul. 2011.

[33] A. Shrivastava, M. Rastegari, S. Shekhar, R. Chellappa, and L. S. Davis, "Class consistent multi-modal fusion with binary features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2282–2291.

[34] M. E. Mavroforakis and S. Theodoridis, "A geometric approach to support vector machine (SVM) classification," *IEEE Trans. Neural Netw.*, vol. 17, no. 3, pp. 671–682, May 2006.

[35] K. P. Bennett and E. J. Bredensteiner, "Duality and geometry in SVM classifiers," in *Proc. Int. Conf. Mach. Learn.*, 2000, pp. 57–64.

[36] 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.

[37] K.-S. Song, "A globally convergent and consistent method for estimating the shape parameter of a generalized Gaussian distribution," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 510–527, Feb. 2006.

[38] K. S. Song, "Globally convergent algorithms for estimating generalized gamma distributions in fast signal and image processing," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1233–1250, Aug. 2008.

[39] 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.

[40] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1733–1740.

[41] H. Tang, N. Joshi, and A. Kapoor, "Blind image quality assessment using semi-supervised rectifier networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2877–2884.

[42] G. Bontempi, "Local learning techniques for modeling, prediction and control," Ph.D. dissertation, Dept. IRIDIA, Univ. Libre Bruxelles, Brussels, Belgium, 1999.

[43] G. Bontempi, M. Birattari, and H. Bersini, "Lazy learning for local modelling and control design," *Int. J. Control*, vol. 72, nos. 7–8, pp. 643–658, 1999.

[44] C. Loader, *Local Regression and Likelihood*. New York, NY, USA: Springer, 2006.

[45] L. Bottou and V. Vapnik, "Local learning algorithms," *Neural Comput.*, vol. 4, no. 6, pp. 888–900, 1992.

[46] D. Nguyen-Tuong, J. R. Peters, and M. Seeger, "Local Gaussian process regression for real time online model learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1193–1200.

[47] D. Nguyen-Tuong, M. Seeger, and J. Peters, "Model learning with local Gaussian process regression," *Adv. Robot.*, vol. 23, no. 15, pp. 2015–2034, 2009.

[48] M. Wu and B. Schölkopf, "Transductive classification via local learning regularization," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2007, pp. 628–635.

[49] H. Shimazaki and S. Shinomoto, "Kernel bandwidth optimization in spike rate estimation," *J. Comput. Neurosci.*, vol. 29, nos. 1–2, pp. 171–182, 2010.

[50] S.-M. Chow, E. Ferrer, and F. Hsieh, Eds., *Statistical Methods for Modeling Human Dynamics: An Interdisciplinary Dialogue* (Notre Dame Series on Quantitative Methodology). New York, NY, USA: Taylor & Francis, 2012.

[51] K. R. Hammond, *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. Oxford, U.K.: Oxford Univ. Press, 2000.

[52] J. R. Busemeyer, Z. Wang, and J. T. Townsend, "Quantum dynamics of human decision-making," *J. Math. Psychol.*, vol. 50, no. 3, pp. 220–241, 2006.

[53] R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (MOS) revisited: Methods and applications, limitations and alternatives," *Multimedia Syst.*, vol. 22, no. 2, pp. 213–227, 2016.

[54] F. Ribeiro, D. Florêncio, C. Zhang, and M. Seltzer, "CROWDMOS: An approach for crowdsourcing mean opinion score studies," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2011, pp. 2416–2419.

[55] T. Hoßfeld, P. E. Heegaard, and M. Varela, "QoE beyond the MOS: Added value using quantiles and distributions," in *Proc. Int. Workshop Quality Multimedia Exper.*, May 2015, pp. 1–6.

[56] E. Snelson and Z. Ghahramani, "Sparse Gaussian processes using pseudo-inputs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, pp. 1257–1264.

[57] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, *LIVE Image Quality Assessment Database Release 2*, accessed on 2005. [Online]. Available: http://live.ece.utexas.edu/research/quality

[58] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process., Image Commun.*, vol. 30, pp. 57–77, Jan. 2015.

[59] A. Zarić *et al.*, "VCL@FER image quality assessment database," *Automatika-J. Control, Meas., Electron., Comput. Commun.*, vol. 53, no. 4, pp. 344–354, 2012.

[60] E. C. Larson and D. M. Chandler, *Categorical Image Quality (CSIQ) Database*, accessed on 2009. [Online]. Available: http://vision.okstate.edu/csiq

[61] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: An experimental comparison," *Inf. Retr.*, vol. 11, no. 2, pp. 77–107, Apr. 2008.

[62] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, Nov. 2000.
[63] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback–Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.
[64] Y. Sun, S. Todorovic, and S. Goodison, "Local-learning-based feature selection for high-dimensional data analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1610–1626, Sep. 2010.
[65] K. Ma *et al.*, "Group MAD competition? A new methodology to compare objective image quality models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1664–1673.
[66] K. Ma *et al.*, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 1004–1016, Feb. 2017.
[67] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
[68] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 82–96.
[69] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.
[70] C. E. Rasmussen, "Gaussian processes in machine learning," in *Advanced Lectures on Machine Learning*, 2004, pp. 63–71.
[71] C. K. I. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Proc. Adv. Neural Inf. Process. Syst.*, 1996, pp. 514–520.
[72] VQEG. (2000). *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase I (FR-TV)*. [Online]. Available: http://www.vqeg.org/
[73] K. R. Hammond, *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. Oxford, U.K.: Oxford Univ. Press, 2000.
[74] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1395–1411, May 2007.
[75] *CIPR Still Images*, accessed on 2002. [Online]. Available: http://www.cipr.rpi.edu/resource/stills/
[76] *Xiph.org Video Test Media*, accessed on 2009. [Online]. Available: https://media.xiph.org/video/derf/
[77] *VCEG KTA Reference Software*, accessed on 2015. [Online]. Available: http://iphome.hhi.de/suehring/tml/download/KTA
[78] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.
[79] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1480–1494, Sep. 2015.
[80] 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.
[81] K. Gu *et al.*, "Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 432–443, Mar. 2016.
[82] Y. Niu, H. Zhang, W. Guo, and R. Ji, "Image quality assessment for color correction based on color contrast similarity and color value difference," *IEEE Trans. Circuits Syst. Video Technol.* [Online]. Available: http://ieeexplore.ieee.org/document/7763834/

Qingbo Wu (S'12–M'13) received the B.E. degree in education of applied electronic technology from Hebei Normal University in 2009 and the Ph.D. degree in signal and information processing from University of Electronic Science and Technology of China in 2015. In 2014, he was a Research Assistant with the Image and Video Processing Laboratory, The Chinese University of Hong Kong. From 2014 to 2015, he was a Visiting Scholar with the Image and Vision Computing Laboratory, University of Waterloo. He is currently a Lecturer with the School of Electronic Engineering, University of Electronic Science and Technology of China. His research interests include image/video coding, quality evaluation, and perceptual modeling and processing.

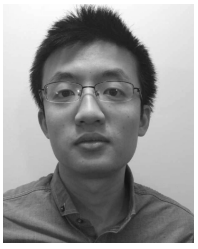
Hongliang Li (SM'12) received the Ph.D. degree in electronics and information engineering from Xi'an Jiaotong University, China, in 2005. From 2005 to 2006, he was a Research Associate with the Visual Signal Processing and Communication Laboratory, The Chinese University of Hong Kong, where he was a Post-Doctoral Fellow from 2006 to 2008. He is currently a Professor with the School of Electronic Engineering, University of Electronic Science and Technology of China. His research interests include image segmentation, object detection, image and video coding, visual attention, and multimedia communication system.

He has authored or co-authored numerous technical articles in well-known international journals and conferences. He is a Co-Editor of a Springer book, *Video Segmentation and Its Applications*. He was involved in many professional activities. He served as a Technical Program Co-Chair of the ISPACS 2009, a General Co-Chair of the ISPACS 2010, a Publicity Co-Chair of the IEEE VCIP 2013, a Local Chair of the IEEE ICME 2014, and a TPC Member in a number of international conferences, such as ICME 2013, ICME 2012, ISCAS 2013, PCM 2007, PCM 2009, and VCIP 2010. He serves as a Technical Program Co-Chair of the IEEE VCIP2016. He is a member of the Editorial Board of *Journal on Visual Communications and Image Representation* and the Area Editor of *Signal Processing: Image Communication* (Elsevier Science).

King N. Ngan (F'00) received the Ph.D. degree in electrical engineering from Loughborough University, U.K. He was a Full Professor with Nanyang Technological University, Singapore, and University of Western Australia, Australia. He has been a Chair Professor with University of Electronic Science and Technology, Chengdu, China, since 2012, under the National Thousand Talents Program. He holds honorary and visiting professorships of numerous universities in China, Australia, and South East Asia. He is currently a Chair Professor with the Department of Electronic Engineering, The Chinese University of Hong Kong.

He has authored extensively, including three authored books, seven edited volumes, over 400 refereed technical papers, and edited nine special issues in journals. He holds 15 patents in the areas of image/video coding and communications. He chaired and co-chaired a number of prestigious international conferences on image and video processing including the 2010 IEEE International Conference on Image Processing, and served on the advisory and technical committees of numerous professional organizations. He served as an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, *Journal on Visual Communications and Image Representation*, *Journal of Signal Processing: Image Communication* (EURASIP), and *Journal of Applied Signal Processing*.

Prof. Ngan is a Fellow of IET (U.K.) and IEAust (Australia). He was an IEEE Distinguished Lecturer from 2006 to 2007.



Kede Ma (S'13) received the B.E. degree from University of Science and Technology of China, Hefei, China, in 2012 and the M.A.Sc. degree from the University of Waterloo, ON, Canada, where he is currently working toward the Ph.D. degree in electrical and computer engineering. His research interests lie in perceptual image processing and computational photography.