Blind Image Quality Assessment Based on Rank-Order Regularized Regression

Qingbo Wu, *Member, IEEE*, Hongliang Li, *Senior Member, IEEE*, Zhou Wang, *Fellow, IEEE*, Fanman Meng, *Member, IEEE*, Bing Luo, Wei Li, and King N. Ngan, *Fellow, IEEE*

***Abstract*—Blind image quality assessment (BIQA) aims to estimate the subjective quality of a query image without access to the reference image. Existing learning-based methods typically train a regression function by minimizing the average error between subjective opinion scores and model predictions. However, minimizing average error does not necessarily lead to correct quality rank-orders between the test images, which is a highly desirable property of image quality models. In this paper, we propose a novel rank-order regularized regression model to address this problem. The key idea is to introduce a pairwise rank-order constraint into the maximum margin regression framework, aiming to better preserve the correct perceptual preference. To the best of our knowledge, this is the first attempt to incorporate rank-order constraints into margin-based quality regression model. By combing with a new local spatial structure feature, we achieve highly consistent quality prediction with human perception. Experimental results show that the proposed method outperforms many state-of-the-art BIQA metrics on popular publicly available IQA databases (i.e., LIVE-II, TID2013, VCL@FER, LIVEMD, and ChallengeDB).**

***Index Terms*—Image quality assessment, rank-order regularized regression.**

I. Introduction

With the rapid development of multimedia and networking technologies and services, the volume of digital image data has been growing explosively in recent years. With the wide spread of intelligent mobile devices (smartphones, tablets, etc.), it has become incredibly convenient for common consumers to acquire, edit and share images in their daily lives. However, various annoying distortions may degrade the subjective quality of an image during its acquisition, storage and transmission processes. Recently, perceptual image quality evaluation has received great attention due to its important role in the acquisition, compression, transmission, enhancement and display of images [1]–[6]. In many applications, the reference image and distortion information are unavailable, and thus the development of general purpose blind image quality assessment (BIQA) methods becomes highly desirable.

In practice, a BIQA model may be required to answer two fundamental questions: 1) How good the quality of an image is, i.e., the absolute perceptual quality; 2) Whether one image has better quality than another, i.e., the relative quality between images. To date, many BIQA algorithms have been proposed, but most of them focus on estimating the absolute perceptual quality, without specific mechanisms to maintain the correct rank-orders between images.

Existing general purpose BIQA approaches maybe roughly classified into four categories. The first class of methods typically train a regression function to map quality-sensitive features to the subjective quality scores. In [7], Mittal *et al.* introduced the mean subtracted contrast normalized (MSCN) coefficients to describe the natural scene statistics (NSS) in the spatial domain, and then support vector regression (SVR) is employed to generate the predicted quality score. In [8], Moorthy *et al.* developed NSS features in the wavelet domain and the subjective quality is generated with two steps, i.e, distortion identification and distortion-specific regression. Gao *et al.* [9] further enriched the NSS features in the wavelet domain and multiple kernel learning (MKL) is introduced to improve the SVR performance. The second class of approaches try to estimate the subjective image quality by pooling the human opinion scores from labeled samples. In [10], He *et al.* designed a sparse representation based method to determine the weights assigned to each differential mean opinion score (DMOS). In [11], Ye *et al.* discussed a clustering based weighting method. In [12], a KNN based label transfer model was explored for fusing multiple image quality labels. The third class of methods predict the perceptual image quality by directly measuring the similarity between the query image and the pristine image in the feature space. In [13], Mittal *et al.* utilized the probabilistic latent semantic analysis (pLSA) to learn the latent topics. For a query image, its estimated loadings across all topics are compared with all pristine images, whose average dot product is used as the predicted quality. In [14], Mittal *et al.* further developed a multivariate

Manuscript received September 21, 2016; revised February 12, 2017; accepted April 22, 2017. Date of publication May 2, 2017; date of current version October 13, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61601102, Grant 61525102, and Grant 61502084, and in part by the Natural Sciences and Engineering Research Council of Canada. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Lingfen Sun. *(Corresponding author: Qingbo Wu.)*

Q. Wu, H. Li, F. Meng, B. Luo, and W. Li are with the School of Electric Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: wqb.uestc@gmail.com; hlli@uestc.edu.cn; fmmeng@uestc.edu.cn; mathild1987@163.com; weili.cv@gmail.com).

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

K. N. Ngan is with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, China, 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).

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

Digital Object Identifier 10.1109/TMM.2017.2700206

Fig. 1. Diagram of rank-order regularized regression. The absolute opinion score constraint drives the regression function to improve the prediction accuracy of each image. The perceptual rank-order constraint refines the regression function to improve the consistency of the quality rank-order of each pair of images.

Gaussian (MVG) model to describe the statistics of MSCN coefficients. The predicted image quality is represented as the distance of the MVG features between the query and pristine images. The fourth class of approaches arise with the booming deep neural network, which aims at end-to-end optimization of the feature extractor and quality regressor simultaneously. In [15], Kang *et al.* explored a deep convolutional neural network (CNN) based method, where the deep learner is fed by the local patches extracted from the human-rated images. Similarly, Bosse *et al.* [16] utilized a popular CNN architecture in [17], and fine tuned it to adapt to the BIQA task.

Among the aforementioned algorithms, the first category of methods show the superiority in producing high prediction accuracy for perceptual image quality. Although different regression models [9], [18], [19] are explored for BIQA, they share one common feature. That is, only the absolute opinion score is utilized in training the perceptual quality regressor, and the relative perceptual rank-orders are not imposed. A mean prediction error based cost function is insensitive to mistaken rank-orders. For example, suppose there are two human opinion scores a and b in the training set ($a > b$). Each one has two alternative predictions with the same absolute error x, i.e., $\hat{a} = a \pm x, \hat{b} = b \pm x$, and $a - x < b + x$. When the predictions are $(a - x, b + x)$, it is clear that the rank-order is wrong. Apparently, one would prefer the predictions to be $(a - x, b - x)$, $(a + x, b + x)$ or $(a + x, b - x)$, which maintain the correct rank-order. However, existing methods are ignorant to this, because all combinations share the same mean prediction error x. Therefore, it is meaningful to adopt a robust regression model, which could enforce the correct perceptual rank-orders between the images.

In this paper, we propose a general purpose BIQA algorithm, which contains two main contributions:

1) We develop a novel *R*ank-order *R*egularized *R*egression (R^3) model. Its diagram is shown in Fig. 1, which includes two types of constraints. On the one hand, the absolute opinion score constraint attempts to associate each single image with an accurate quality score. On the other hand, the perceptual rank-order constraint tries to preserve the correct perception rank-order for each pair of images. In the proposed algorithm, we convert this pairwise constrained regression problem into a quadratic programming problem [20], [21] and use the interior-point algorithm [22], [23] to solve it.
2) A new image feature is introduced by computing the joint statistics of the spatial intensity variations and the distribution of the neighboring pixels. Both the multi-scale and Lab color perception properties are explored to enhance the discriminative power of this quality-aware feature.

The proposed regression model and image feature lead to significantly improved quality prediction accuracy of the proposed method against state-of-the-art BIQA metrics on the LIVE-II, TID2013, VCL@FER, LIVEMD and ChallengeDB databases.

The rest of this paper is organized as follows. Section II briefly reviews the SVR model and relevant rank learning methods. The proposed algorithm is presented in Section III. The experimental results are discussed in Section IV. Finally, Section V concludes this paper.

II. Related Work

In existing BIQA methods, SVR [19], [24] is one of the most widely used regression tools for bridging the gap between image features and the subjective image quality. Let $L = \{(x_i, y_i)\}$ denote the training set, which consists of n samples. $x_i \in \mathbb{R}^d$ is the input feature vector of the ith sample, and $y_i \in \mathbb{R}$ is the corresponding subjective quality label. SVR aims to learn a function $f(x_i)$, which maps x_i to y_i as closely as possible. $f(x_i)$ is typically a linear function, i.e.,

$$f(x_i) = \mathbf{w}^T \Phi(x_i) + b \tag{1}$$

where $\mathbf{w}$ and b denote the linear weight vector and the bias, respectively, and $\Phi(\cdot)$ represents the feature space transformation.

The parameters $\mathbf{w}$ and b can be learned by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{w},\xi_i,\xi_i^*} \ & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i (\xi_i + \xi_i^*) \\ s.t. \ & y_i - \mathbf{w}^T\Phi(x_i) - b \le \epsilon + \xi_i, \ \forall i \\ & \mathbf{w}^T\Phi(x_i) + b - y_i \le \epsilon + \xi_i^*, \ \forall i \\ & \xi_i \ge 0, \ \xi_i^* \ge 0 \end{aligned} \tag{2}$$

where ϵ denotes the tolerated precision, ξ_i and ξ_i^* denote the training loss.

It is noted that the above model does not explicitly penalize rank-order mistakes, which may limit it from producing high rank-order correlations with the human perception. Fortunately, a wealth of studies about rank learning provide us with many useful methodologies to address this problem. In [25], Herbrich *et al.* discussed a margin based framework for learning the ranking function, where a pairwise rank constraint is utilized. In [26], Burges *et al.* introduced the probabilistic ranking function, which can be learned by a gradient descent method. In [27], Hüllermeier *et al.* further decomposed the ranking problem into an ensemble of binary classification problems to account for the pairwise preferences.

Following the pairwise comparison framework, some interesting attempts were explored for BIQA recently. In [28], Chen *et al.* used the rank SVM to evaluate the perceptual rank-order between two enhanced images. In [29], Yan *et al.* employed the multiple additive regression trees to predict the perceptual rank-order for the color images enhanced with different parameters. In [30], Reibman *et al.* proposed a probabilistic predictor to measure the confidence of the pairwise preceptual preference between two images. Although these rank learning methods cannot be directly used for estimating the absolute perceptual quality of a single image, their rank constraints inspire us to extend SVR to a more general rank-order regularized regression model.

III. Proposed Approach

A. Rank-Order Regularized Regression (R^3)

In the proposed regression model, we incorporate a pairwise rank constraint into (2), and rewrite the primal problem as

$$\begin{aligned}\min_{\mathbf{w},\xi_i,\xi_i^*,\eta_{ij}} \ & \frac{1}{2}\|\mathbf{w}\|^2 + C_1\sum_i \xi_i + C_2\sum_i \xi_i^* + C_3\sum_{ij}\eta_{ij} \\ s.t.\ & y_i - \mathbf{w}^T\Phi(x_i) - b \le \epsilon + \xi_i,\ \forall i \\ & \mathbf{w}^T\Phi(x_i) + b - y_i \le \epsilon + \xi_i^*,\ \forall i \\ & \mathbf{w}^T\Phi(x_i) - \mathbf{w}^T\Phi(x_j) \ge \varepsilon - \eta_{ij},\ \forall (i,j)\in\mathcal{P} \\ & \xi_i \ge 0,\ \xi_i^* \ge 0,\ \eta_{ij} \ge 0 \end{aligned} \tag{3}$$

where $\mathcal{P} = \{(i,j) : y_i > y_j + \Delta\}$ and Δ is a perceptual difference threshold which eliminates the pairwise training samples of indistinguishable differences in subjective quality scores. C_1, C_2 and C_3 are parameters to maintain a trade-off between minimizing the prediction error and preserving the pairwise rank-orders.

To illustrate the roles of different constraints, a simulated example is shown in Fig. 2, where five data samples are included in all plots, and a linear model is used to predict the vertical axis y from the horizontal axis x. Two drastically different prediction models are obtained (top and bottom), with model parameters $(\mathbf{w}_1, b_1)$ and $(\mathbf{w}_2, b_2)$, respectively. The left plots illustrate the penalties of prediction error, where the prediction error terms ξ and $\hat{\xi}$ are similar to each other. Therefore, the two linear models perform similarly. However, the predicted ranks are very different. For example, comparing the samples i and j in the right plots, since y_i is significantly higher than y_j, one would expect i to be ranked higher than j. Obviously, such an expected rank-order is well preserved by the top linear model, but not the bottom one, where $\mathbf{w}^T(x_i - x_j) < 0$. Since our pairwise rank-order constraint requires $\mathbf{w}^T(x_i - x_j) \ge \varepsilon - \eta_{ij}$, the mistaken ranks would increase η_{ij} and avoid selecting the parameters $\mathbf{w}_2$ and b_2. This demonstrates that an additional rank-order constraint can greatly help guide the regression training process to select a better model.

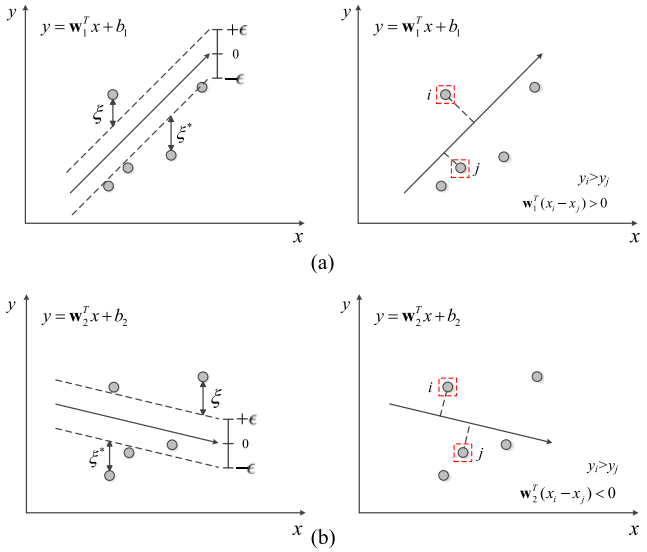


Fig. 2. Comparison of the regression performances under different model parameters $\mathbf{w}$ and b. The left plots show the prediction error with respect to the absolute opinion score. The right plots show the predicted pairwise rank-order. The x-coordinate represents the observation data and the y-coordinate indicates the target variable. The red bounding boxes highlight two samples whose ranks are correctly predicted in (a) and mistaken in (b).

To extend the linear prediction function, we further explore its kernel version, as in [19], [25]. Moreover, the constrained optimization problem of (3) can be converted to a unconstrained optimization problem by introducing Lagrangian loss functions

$$\begin{aligned}\mathcal{L} = & \frac{1}{2}\|\mathbf{w}\|^2 + C_1\sum_i \xi_i + C_2\sum_i \xi_i^* + C_3\sum_{ij}\eta_{ij} \\ & + \sum_i \alpha_i(y_i - \mathbf{w}^T\Phi(x_i) - b - \epsilon - \xi_i) \\ & + \sum_i \alpha_i'(\mathbf{w}^T\Phi(x_i) + b - y_i - \epsilon - \xi_i^*) \\ & + \sum_{ij}\alpha_{ij}''\left(\mathbf{w}^T(\Phi(x_i) - \Phi(x_j)) - \varepsilon + \eta_{ij}\right) \end{aligned} \tag{4}$$

where α_i, α_i' and α_{ij}'' are the Lagrange multipliers.

According to the Karush-Kuhn-Tucker (KKT) conditions [24], the solution that minimizes the primal problem of (4) is given by solving the saddle point, which corresponds to vanishing the partial derivatives of $\mathcal{L}$ w.r.t. the primal variables

$(\mathbf{w}, b, \xi_i, \xi_i^*, \eta_{ij})$. Meanwhile, the constraints and the Lagrange multipliers should satisfy the following requirements

1) Primal feasibility:

$$\begin{aligned} y_i - \mathbf{w}^T\Phi(x_i) - b &\le \epsilon + \xi_i,\ \forall i \\ \mathbf{w}^T\Phi(x_i) + b - y_i &\le \epsilon + \xi_i^*,\ \forall i \\ \mathbf{w}^T\Phi(x_i) - \mathbf{w}^T\Phi(x_j) &\ge \varepsilon - \eta_{ij},\ \forall (i,j) \in \mathcal{P}. \end{aligned} \tag{5}$$

2) Dual feasibility:

$$\alpha_i, \alpha_i', \alpha_{ij}'' \ge 0,\ \forall\ i \text{ and } (i,j) \in \mathcal{P}. \tag{6}$$

3) Complementary slackness:

$$\begin{aligned} \alpha_i(y_i - \mathbf{w}^T\Phi(x_i) - b - \epsilon - \xi_i) &= 0 \\ \alpha_i'(\mathbf{w}^T\Phi(x_i) + b - y_i - \epsilon - \xi_i^*) &= 0 \\ \alpha_{ij}''\big(\mathbf{w}^T(\Phi(x_i) - \Phi(x_j)) - \varepsilon + \eta_{ij}\big) &= 0. \end{aligned} \tag{7}$$

Particularly, the derivative on $\mathbf{w}$ is represented by

$$\partial\mathcal{L}/\partial\mathbf{w} = \mathbf{w} + \sum_i(\alpha_i' - \alpha_i)\Phi(x_i) + \sum_{ij}\alpha_{ij}''\Big(\Phi(x_i) - \Phi(x_j)\Big). \tag{8}$$

Given $\partial\mathcal{L}/\partial\mathbf{w} = 0$, we can deduce the $\mathbf{w}$ by

$$\mathbf{w} = \sum_i(\alpha_i - \alpha_i')\Phi(x_i) - \sum_{ij}\alpha_{ij}''\Big(\Phi(x_i) - \Phi(x_j)\Big). \tag{9}$$

Replacing $\alpha_i - \alpha_i'$ with β_i, we obtain the kernel version of the prediction function by plugging (9) into (1)

$$f(x) = \sum_i \beta_i\mathbb{K}(x_i, x) - \sum_{ij}\alpha_{ij}''\Big(\mathbb{K}(x_i, x) - \mathbb{K}(x_j, x)\Big) + b \tag{10}$$

where $\mathbb{K}(\cdot,\cdot)$ is the kernel function.

Let n denote the number of the training images, and m the number of all their pairwise combinations, where $m = \frac{n(n-1)}{2}$. From (10), we can find that the maximum number of the support vectors could be up to $n + m$. Since m rapidly increases with n, the complexities of both the training and testing processes of the regression model would greatly increase, making it intractable. To address this issue, simplification is necessary.

Inspired by the "Representer Theorem" in [31], we know that our loss function in (4) corresponds to a special case of the arbitrary cost function c which satisfies $c\,(\mathcal{X}\times\mathbb{R}^2)^n \to \mathbb{R}\cup\infty$. Let h denote a strictly monotonically increasing real-valued function on $[0,\infty)$, and $\mathcal{F}$ the function set

$$\mathcal{F} = \Big\{ f \in \mathcal{X} | f(x) = \sum_{i=1}^{\infty}\alpha_i\mathbb{K}(x_i, x), \quad \alpha_i \in \mathbb{R}, x_i \in \mathcal{X}, \|f\| < \infty \Big\}. \tag{11}$$

Then any $f \in \mathcal{F}$ minimizing the regularized risk function

$$c\big((x_1, y_1, f(x_1), \ldots, (x_n, y_n, f(x_n))\big) + h(\|f\|) \tag{12}$$

would admit a representation form of

$$f(x) = \sum_{i=1}^{n}\alpha_i\mathbb{K}(x_i, x). \tag{13}$$

By replacing the prediction function (10) using (13), we reduce the number of parameters from $n + m$ to n, which significantly simplifies the solution of the R^3 problem. In addition, as discussed in [32], the penalty for $\mathbf{w}$ could also be achieved by suppressing the coefficient vector $\alpha = [\alpha_1, \ldots, \alpha_n]^T$. Accordingly, the objective function in (3) can be rewritten as

$$\begin{aligned} \min_{\alpha,\xi_i,\xi_i^*,\eta_{ij}}\ & \|\alpha\|^2 + C_1\sum_i\xi_i + C_2\sum_i\xi_i^* + C_3\sum_{ij}\eta_{ij} \\ s.t.\ & y_i - \sum_{u=1}^{n}\alpha_u\mathbb{K}(x_u, x_i) \le \epsilon + \xi_i,\ \forall i \\ & \sum_{u=1}^{n}\alpha_u\mathbb{K}(x_u, x_i) - y_i \le \epsilon + \xi_i^*,\ \forall i \\ & \sum_{u=1}^{n}\alpha_u\big(\mathbb{K}(x_u, x_i) - \mathbb{K}(x_u, x_j)\big) \ge \varepsilon - \eta_{ij}, \\ & \forall (i,j) \in \mathcal{P} \\ & \xi_i \ge 0,\ \xi_i^* \ge 0,\ \eta_{ij} \ge 0. \end{aligned} \tag{14}$$

Here, the optimization for (14) is a quadratic programming (QP) problem, which can be solved by the interior-point algorithm [23]. Furthermore, inspired by the sample selection scheme in large-scale classification [33], [34], we introduce a *k*-nearest-neighbor (KNN) based method to speed up the solving of (14) by limiting the size of the pairwise training samples. Particularly, it follows three steps to train the regressor, i.e.,

1) Compute the chi-square distances [35] of the features between the test image and all training samples, based on which we find the KNN of the test image.
2) Use the KNN to construct the training set. Then, the parameter α of (13) is learned from the $2K$ prediction error constraints and $\frac{K(K-1)}{2} - l$ pairwise rank-order constraints as in (14), where l is the number of the pairwise training samples that do not pass the perceptual difference threshold condition in $\mathcal{P}$.
3) Use the learned regression function to predict the subjective quality score for the test image.

In this way, the number of the support vectors is limited to be less than or equal to K, which corresponds to the training samples located within the soft-margin in (14).

After building the R^3 model, we consider the differences of the deviations caused by different distortion types and estimate the image quality with two steps, which are similar to [8], [36]. Firstly, an SVM [37] classifier is used to identify the distortion type of a test image. Then, a distortion-specific regressor is used to predict its perceptual quality, where the training subset is identified to share the same distortion type with the test image by the SVM classifier.

B. Local Spatial Structure Feature

The structural information of an image is contained in both the spatial intensity variations and the dependencies between neighboring pixels. Existing NSS features work well in describing the statistical characteristic of the spatial intensity variations. The

property of the spatial neighboring interactions has also been exploited. For example, Mittal *et al.* [7] investigated the products of pairwise neighboring pixels in four directions. An asymmetric generalized Gaussian distribution (AGGD) model is used to describe their distribution. In [9], pairwise mutual information between 17 neighboring coefficients and one central coefficient is computed in the wavelet domain.

For local spatial structure (LSS), a complete description should contain the joint statistics between all neighboring pixels, which cannot be fully achieved from pairwise relationships only. To overcome this limitation, many local binary pattern (LBP) based methods [12], [38]–[40] were proposed. In comparison with pairwise neighboring statistics, LBP captures the relative relationships between the central pixel and all of its neighbors.

Inspired by the success of LBP, we focus on extracting image features based on LSS. Because LBP misses the magnitude information of the spatial contrast, we incorporate a more comprehensive joint statistic of LSS which is referred to as the completed LBP (CLBP) [35]. Particularly, CLBP utilizes three components to describe the local image structure, i.e., the central pixel's intensity, and the sign and the magnitude of the neighboring pixel's local difference. The three components are first encoded into binary patterns, which are denoted by $CLBP\_C$, $CLBP\_S$ and $CLBP\_M$, respectively. The joint statistics of these encoded patterns are used as the image features.

According to [35], the code of the central intensity's binary pattern is given by

$$CLBP\_C_{P,R} = t(g_c, c_I), \qquad t(x,c) = \begin{cases} 1, & x \geq c \\ 0, & x < c \end{cases} \quad (15)$$

where g_c denotes the gray level of the center pixel. The subscript P denotes the number of neighboring pixels and R is the radius between the center pixel and its neighbors. c_I is a threshold set as the average intensity of all the pixels in an image.

The encoding of the local difference's sign can be expressed as

$$CLBP\_S_{P,R} = \begin{cases} \sum_{p=0}^{P-1} t(g_p - g_c, 0), & \text{if } U(LBP_{P,R} \leq 2) \\ P+1, & \text{otherwise} \end{cases} \quad (16)$$

$$U(LBP_{P,R}) = |t(g_{P-1} - g_c, 0) - t(g_0 - g_c, 0)| + \sum_{i=1}^{P-1} |t(g_i - g_c, 0) - t(g_{i-1} - g_c, 0)| \quad (17)$$

where $U(LBP_{P,R})$ denotes the discontinuities of the binary presentation in a circular shape of neighboring pixels [41].

The code of the local difference's magnitude m_p is defined as

$$CLBP\_M_{P,R} = \sum_{p=0}^{P-1} t(m_p, c) 2^p \quad (18)$$

where c denotes the threshold, which is set as the average value of all m_p's in an image.

The image feature $s_{P,R}$ can be interpreted as the joint distribution of all the three binary patterns, i.e.,

$$s_{P,R} = hist3(CLBP\_C_{P,R}, CLBP\_S_{P,R}, CLBP\_M_{P,R}) \quad (19)$$

where $hist3(\cdot,\cdot,\cdot)$ is the 3D histogram operator.

Let $n_C(P)$, $n_S(P)$ and $n_M(P)$ denote the number of the binary patterns for the central intensity, sign and magnitude of the neighboring pixels' local differences, respectively. Let $D(s_{P,R})$ denote the dimension of the CLBP feature, which equals to $n_C(P) \times n_S(P) \times n_M(P)$. According to [35], the values of $D(s_{P,R})$ are given by

$$D(s_{P,R}) = \begin{cases} 2 \times 10 \times 10, & P = 8 \\ 2 \times 18 \times 18, & P = 16 \\ 2 \times 26 \times 26, & P = 24. \end{cases} \quad (20)$$

Inspired by the hierarchical property of visual perception [42]–[44], we extract the multi-scale CLBP feature under three parameter settings for (P, R), i.e., (8, 1), (16, 2) and (24, 3). In addition, we extract the CLBP features from all three color channels of a color image, which is implemented in the perceptually uniform Lab color space [45]. In Fig. 3, we show the diagram of the multi-scale feature extraction process for the L component. Let S^L, S^a and S^b denote the multi-scale features in the three color components, respectively, which are expressed as

$$\begin{aligned} S^L &= [s^L_{8,1}, s^L_{16,2}, s^L_{24,3}], \\ S^a &= [s^a_{8,1}, s^a_{16,2}, s^a_{24,3}], \\ S^b &= [s^b_{8,1}, s^b_{16,2}, s^b_{24,3}]. \end{aligned} \quad (21)$$

Finally, the proposed multi-scale and multi-channel LSS feature set $\mathcal{S}$ is given by

$$\mathcal{S} = [S^L, S^a, S^b]. \quad (22)$$

IV. Experimental Results

A. Procedures

To evaluate the performance of the proposed algorithms, we conduct the experiments on three publicly available benchmark databases, i.e., LIVE-II [46], TID2013 [47] and VCL@FER [48].

1) LIVE-II: The LIVE-II IQA database is composed of 29 original images and 779 distorted versions of them under 5 distortion types, which include 169 JPEG2000 (JP2K) compression images, 175 JPEG compression images, 145 additive Gaussian White Noise (WN) images, 145 Gaussian Blur images, and 145 Fast Fading (FF) images. Each distorted image is associated with a Differential Mean Opinion Score (DMOS), which ranges from 0 to 100 and a larger value indicates worse perceptual quality.
2) TID2013: The TID2013 IQA database includes 25 reference images and 3000 distorted versions of them under 17 distortion types, which are additive Gaussian WN, additive noise in color components (ANC), spatially correlated noise (SCN), masked noise (MN), high frequency noise (HFN), impulse noise (IN), quantization

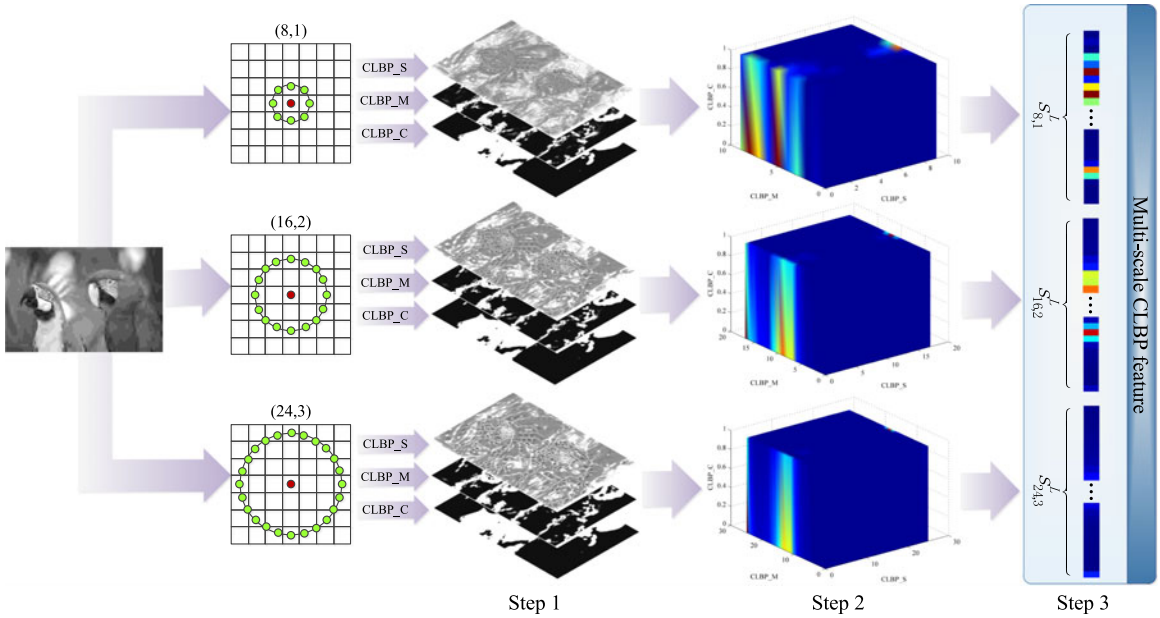


Fig. 3. Diagram of the multiscale CLBP feature extraction. Step 1: Generate three encoded maps, i.e., CLBP_S/M/C, for the local contrast's sign, magnitude and central intensity, respectively. Step 2: Compute the 3D joint histogram for CLBP_S/M/C, where the x/y/z-axis correspond to the values of three encoded maps and a warmer color denotes a higher frequency of occurrence. Step 3: Reshape the 3D histogram into the 1D vector $s^L_{P,R}$ and concatenate the multiscale vectors.

noise (QN), Gaussian Blur, image denoising (ID), JPEG compression, JP2K compression, JPEG transmission errors, JPEG2000 transmission errors (JPEGTE), Non eccentricity pattern noise (NEPN), Local block-wise distortions of different intensity (LBDDI), Mean shift (MS) and Contrast change (CC). Each distorted image is assigned a Mean Opinion Score (MOS), which ranges from 0 to 9 and a larger value indicates better subjective quality.

3) VCL@FER: The VCL@FER IQA database consists of 23 reference images and 552 distorted versions of them under 4 distortion types, which include additive Gaussian WN, Gaussian Blur, JP2K and JPEG. Each distorted image has a MOS value, which ranges from 0 to 100.

Following the same criterion in [7], [8], [36], [49], we divide the LIVE-II database into non-overlapped training and testing sets. Particularly, we randomly select 23 of the 29 reference images and their associated distorted images to construct the training set. The rest of the images are used for testing. We conduct the random splitting evaluation 100 times. The median values of the indices across 100 trials are reported.

Following the approach used by the video quality experts group (VQEG) [50], the objective quality scores are first mapped to the subjective scores via a four-parameter logistic function, whose coefficients are solved via the iterative least squares estimation [51]. Two common measures are then employed to evaluate the performance of different BIQA algorithms, which include the Pearson's linear correlation coefficient (PLCC) and the Spearman's rank-order correlation coefficient (SROCC) between the predicted quality metric Q and the ground-truth human perception (e.g, DMOS or MOS).

B. Implementation Details

In implementing the regression, the YALMIP toolbox [52] with the Gurobi solver [53] is used to model and solve the optimization problem in (14). The polynomial kernel is employed in this experiment. We set the perceptual difference threshold Δ to 10, which is obtained by rounding off the mean standard deviation of all DMOS in LIVE-II database. The model parameters $C_1 \sim C_3$ are determined by cross validation.

To better understand the impact of R^3, we implement two versions of the proposed regression method. The first is a partial version (2) that does not involve the rank-order constraint. The second is the full version (3) that includes the rank-order regularization.

C. Parameter Selection for K

In this section, we investigate how the prediction accuracy and complexity vary with the parameter K in the KNN method. Particularly, we implement the proposed BIQA algorithm with different K on the LIVE-II database, where K ranges from 5 to 40 with an interval of 5. The training and testing sets are obtained by randomly splitting the LIVE-II database 100 times, as discussed in Section IV-A. The median SROCC value is used to evaluate the prediction accuracy and the median running time is used to measure the complexity. The results are reported in Fig. 4.

As shown in Fig. 4(a), the SROCC performances of the proposed method with and without R^3 monotonically go up as K increases from 5 to 40, but the increase becomes slower when K is larger than 20. When K is larger than 30, SROCC starts to converge. By contrast, for both the proposed method with and without R^3, the running time is always monotonically increas-

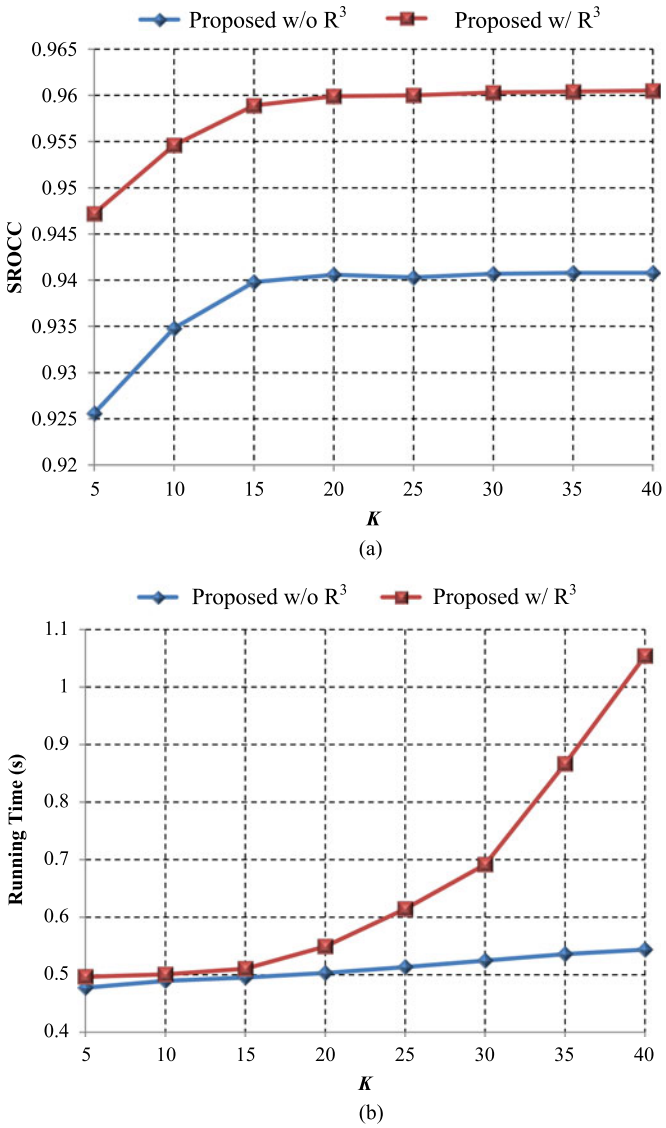


Fig. 4. Median SROCC and running time in seconds versus the parameter K as testing on the LIVE-II database. (a) SROCC variation. (b) Running time variation.

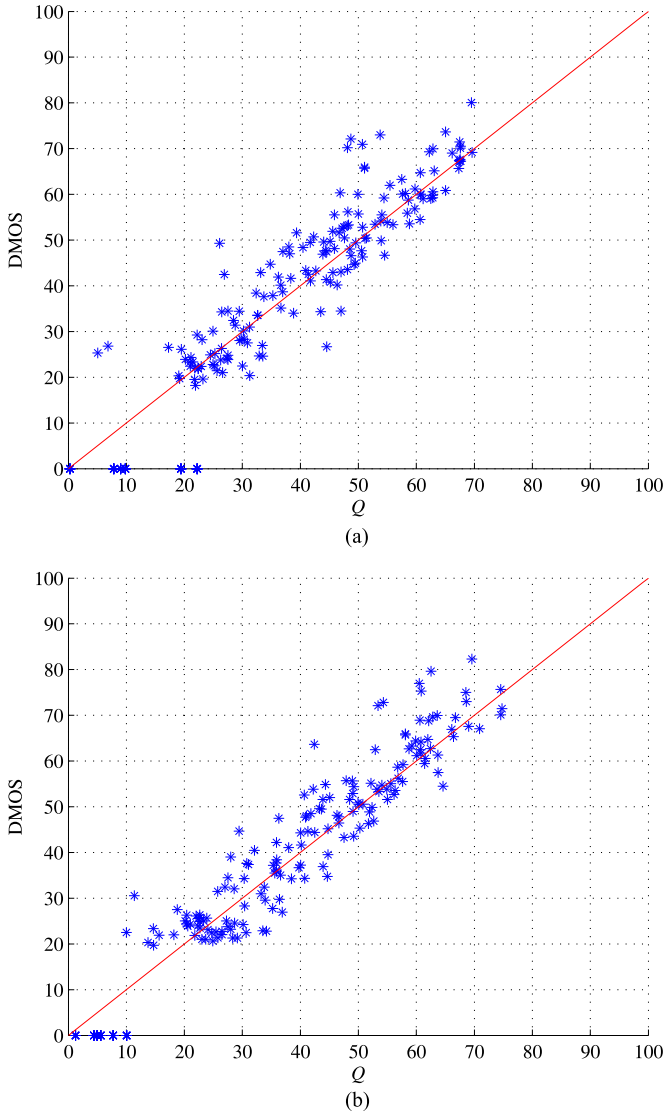


Fig. 5. Scatter plots of the predicted quality index Q versus DMOS for the test set with median SROCC. The x-axis is the predicted quality index Q and the y-axis is the DMOS value. The red line represents the ideal linear correlated line. (a) Proposed without R^3. (c) Proposed w/R^3.

ing. In particular, the running time of the proposed method with R^3 grows much faster due to the increased cost in computing the pairwise rank-order term. To achieve a balance between accuracy and complexity, we set the parameter K to 30 in the rest of the paper, and there are at most 30 support vectors needed for the quality regressor.

D. Prediction Accuracy and Monotonicity

A good BIQA metric should satisfy two requirements: 1) Predicts perceptual quality accurately with small prediction error, which can be measured by PLCC; 2) Predicts perceptual quality with high monotonicity, which can be evaluated by SROCC. In this section, we conduct such performance evaluation on the LIVE-II database. Both the proposed method with and without R^3 are compared with classical full-reference (FR) IQA metrics (PSNR, SSIM [54]), and state-of-the-art BIQA metrics (BIQI [36], DIIVINE [8], BLINDS-II [49], BRISQUE [7], NIQE [14] and QAC [55]).

In Fig. 5, we show the scatter plots of the proposed methods for the test set with median SROCC. It can be seen that the predicted image quality scores show a nearly linear relationship with respect to DMOS, where most samples compactly gather around the diagonal line, i.e., "$Q =$ DMOS". It demonstrates that the model predictions are highly consistent with human perception. By comparing Fig. 5(a) and 5(b), we can observe that the strongest impact of R^3 is on the "outlier" image samples, which create strong rank-order penalties that led to significant changes in the behavior of the regressor. As a result, these samples are more tightly clustered towards the diagonal line when the proposed R^3 approach is applied. To quantify the impact of R^3 in Fig. 5, we further compute the number of incorrectly ranked image pairs, which is represented by $P_{\text{outlier}} = \sum_{ij}^{N} \delta_{i,j}$. Par-

TABLE I
MEDIAN PLCC ACROSS 100 TRAIN-TEST TRIALS ON THE LIVE-II IQA DATABASE

Distortion		JP2K	JPEG	WN	Blur	FF	All
Metric	Type						
PSNR	FR	0 896	0 860	0 986	0 783	0 890	0 872
SSIM	FR	0.937	0.928	0.970	0.874	0.943	0.945
BIQI	Blind	0 750	0 630	0 968	0 800	0 722	0 740
DIIVINE	Blind	0 922	0 921	0 988	0 923	0 888	0 917
BLINDS-II	Blind	0 963	**0.979**	0 985	0 948	0 864	0 923
BRISQUE	Blind	0 923	0 974	0 985	0 951	0 903	0 942
NIQE	Blind	0.926	0.952	0.976	0.943	0.879	0.905
OAC	Blind	0 865	0 944	0 9 18	0 911	0 825	0 863
Proposed w/o R^3	Blind	0.938	0.945	0.982	0.957	0.917	0.943
Proposed w/ R^3	Blind	**0.965**	0 974	**0.989**	**0.972**	**0.948**	**0.966**

TABLE II
MEDIAN SROCC ACROSS 100 TRAIN-TEST TRIALS ON THE LIVE-II IQA DATABASE

Distortion		JP2K	JPEG	WN	Blur	FF	All
Metric	Type						
PSNR	FR	0 890	0 841	0 985	0 782	0 890	0 876
SSIM	FR	0.932	0.903	0.963	0.894	0.941	0.947
BIQI	Blind	0 736	0 591	0 958	0 778	0 700	0 726
DIIVINE	Blind	0 913	0 910	0 984	0 921	0 863	0 916
BLINDS-II	Blind	0.951	0.942	0.978	0.944	0.862	0.920
BRISQUE	Blind	0 914	**0.965**	0 979	0 951	0 877	0 940
NIOE	Blind	0.919	0.942	0.972	0.933	0.864	0.909
OAC	Blind	0 862	0 936	0 951	0 913	0 823	0 868
ProDosed w/o R^3	Blind	0.928	0.928	0.981	0.965	0.907	0.941
Proposed w/ R^3	Blind	**0.952**	0.954	**0.984**	**0.968**	**0.923**	**0.960**

ticularly, N is the total number of all possible combinations for image pairs in the test set, and $\delta_{i,j}$ denote the counting function, which is given by

$$\delta_{i,j} = \begin{cases} 1, \text{ if } (DMOS_i - DMOS_j) \cdot (Q_i - Q_j) < 0 \\ 0, \text{ otherwise} \end{cases} \tag{23}$$

where $DMOS_i$ and Q_i denote the ground-truth and the predicted quality for the ith image, respectively. Corresponding to Fig. 5, the P_{outlier} for the proposed method w/ and w/o R^3 are 1476 and 1789, respectively. That is, the proposed R^3 model could significantly reduce the "outlier" image samples, whose rank-orders are mistaken.

The quantitative comparisons of different IQA metrics are presented in Tables I and II, which list the median PLCC and SROCC results, respectively. For clarity, the optimal BIQA results in terms of PLCC and SROCC are highlighted in boldface for each distortion type. It can be seen that for each distortion type, the proposed methods with and without R^3 always achieve the top three PLCC or SROCC results. In particular, the proposed method with R^3 performs the best on almost all individual distortion types other than JPEG, for which its performance is very close to the best result obtained by the BRISQUE metric.

TABLE III
STANDARD DEVIATIONS IN PERFORMANCE OF THE PROPOSED METHOD WITHOUT AND WITH R^3 ACROSS 100 TRIALS ON THE LIVE-II DATABASE

	Proposed w/o R^3		Proposed w/R^3	
	PLCC	SROCC	PLCC	SROCC
JP2K	0.007	0.009	0.010	0.012
JPEG	0.008	0.015	0.008	0.014
WN	0.004	0.006	0.004	0.005
Blur	0.012	0.017	0.016	0.021
FF	0.013	0.025	0.022	0.027
All	0.005	0.005	0.008	0.007

TABLE IV
MEDIAN RMSE OF THE PROPOSED METHOD WITHOUT AND WITH R^3 ACROSS 100 TRIALS ON THE LIVE-II DATABASE

	JP2K	JPEG	WN	Blur	FF	All
Proposed w/o R^3	9.416	8.448	5.274	7.170	9.291	8.255
Proposed w/ R^3	6.924	5.924	4.183	5.377	7.514	6.202

A challenging task of general purpose BIQA is to accurately predict the subjective image quality across different distortion types. On this criterion, the proposed method performs the best on the entire database, which contains all test samples across all five distortion types. As shown in Tables I and II, both versions of the proposed methods outperform the other BIQA metrics. Specifically, the proposed method with R^3 achieves the highest PLCC and SROCC results. Since the rank-order constraint better preserves the perceptual preference, no matter for each distortion type or the entire database test, the proposed method with R^3 always achieves higher PLCC and SROCC results than that without R^3. In Table III, we show the standard deviations across 100 train-test trials. It can be seen that the standard deviations are all very small, which demonstrates that the performance of the proposed method is robust to variations of the training sets. In addition to the PLCC and SROCC indexes, the median root mean square error (RMSE) results of the proposed method w/ and w/o R^3 are also reported in Table IV. We can find that the RMSE of the regression model w/ R^3 is smaller than that w/o R^3 across all distortion types and the whole test set. It validates that the correct rank-order information is also beneficial for better approximating the human perception of image quality.

To verify that the superiority of the proposed method is statistically significant, we further perform the one-sided *t-test* [56] on the SROCC results which are generated from 100 train-test trials. The results are reported in Table V. It can be seen that the proposed method with R^3 is statistically better than all the other BIQA algorithms. This is consistent with the median SROCC comparison result in Tables I and II.

E. Classification Accuracy

In a two-step method, distortion-type classification plays an important role in assigning the correct distortion-specific regression function to the test image. To verify the superiority of our LSS feature in identifying different distortion types, we investigate the median classification accuracy across 100 trials

TABLE V
STATISTICAL SIGNIFICANCE ON RELATIVE PERFORMANCE OF BIQA METHODS. A VALUE OF "1"/"0"/"−1" INDICATES THE ROW ALGORITHM IS STATISTICALLY SUPERIOR/EQUIVALENT/INFERIOR TO THE COLUMN ALGORITHM

	BIQI	DIIVINE	BLINDS-II	BRISQUE	NIQE	QAC	Proposed w/o R^3	Proposed w/R^3
BIQI	0	−1	−1	−1	−1	−1	−1	−1
DIIVINE	1	0	1	−1	1	1	−1	−1
BLINDS-II	1	−1	0	−1	1	1	−1	−1
BRISQUE	1	1	1	0	1	1	−1	−1
NIQE	1	−1	−1	−1	0	1	−1	−1
QAC	1	−1	−1	−1	−1	0	−1	−1
Proposed w/o R^3	1	1	1	1	1	1	0	−1
Proposed w/ R^3	1	1	1	1	1	1	1	0

TABLE VI
MEDIAN CLASSIFICATION ACCURACY (%) ACROSS 100 TRAIN-TEST TRIALS ON THE LIVE-II IQA DATABASE

	JP2K	JPEG	WN	Blur	FF	All
DIIVINE	80.00	81.10	100.00	90.00	73.30	83.75
LSS	91.55	100.00	100.00	96.67	80.00	92.90

in Table VI, where DIIVINE based classification is also listed for comparison. It can be seen that the proposed LSS feature-based classification achieves higher accuracies for all distortion types. The mostly occurred classification errors are between the JP2K and FF datasets. This is likely because the FF images were created in two steps: JP2K compression followed by simulated random fast fading channel distortion. As a result, the artifacts in the FF set are very similar to those in JP2K, especially when the fast fading errors are minor. Consequently, such classification errors do not lead to strong degradations in the overall performance of the proposed algorithm.

F. Cross Database Verification

To verify the proposed methods across databases, we follow the approach in [7], [8], [49]. In particular, we use the entire LIVE-II database as the training set, and then test the proposed method on the "known" subset of TID2013 and VCL@FER databases which include four common distortion types, i.e., JP2K, JPEG, WN and Blur.

Tables VII and VIII show the detailed SROCC results between the predicted quality score and MOS on TID2013 and VCL@FER databases, respectively. For comparison, two classic FR-IQA metrics (i.e., PSNR, SSIM) and six state-of-the-art BIQA algorithms (i.e., BIQI, DIIVINE, BLINDS-II, BRISQUE, NIQE and QAC) are also listed. The best SROCC results among all BIQA metrics are highlighted by boldface.

In the TID2013 database, the proposed method with R^3 achieves the best SROCC result for the distortion type of Blur, and the second best for JP2K and JPEG. For the case of all test samples across four distortion types, it outperforms all the other BIQA metrics. Similarly, on the VCL@FER database, for JP2K, JPEG and Blur, the proposed method with R^3 obtains the best SROCC performance. When tested on the entire VCL@FER database, it is again superior to all the other BIQA algorithms.

TABLE VII
SROCC BETWEEN THE PREDICTED QUALITY SCORE AND MOS ON THE "KNOWN" SUBSETS OF THE TID2013 DATABASE

Metric	Type	JP2K	JPEG	WN	Blur	All
PSNR	FR	0.891	0.919	0.944	0.967	0.924
SSIM	FR	0.949	0.932	0.874	0.970	0.921
BIQI	Blind	0.852	**0.891**	0.807	0.839	0.839
DIIVINE	Blind	0.854	0.668	**0.872**	0.830	0.796
BLINDS-II	Blind	0.901	0.859	0.661	0.870	0.793
BRISQUE	Blind	**0.909**	0.889	0.851	0.881	0.882
NIQE	Blind	0.898	0.864	0.816	0.815	0.812
QAC	Blind	0.894	0.871	0.754	0.885	0.855
Proposed w/o R^3	Blind	0.870	0.872	0.809	0.869	0.857
Proposed w/ R^3	Blind	0.905	0.890	0.840	**0.919**	**0.886**

TABLE VIII
SROCC BETWEEN THE PREDICTED QUALITY SCORE AND MOS ON THE VCL@FER DATABASE

Metric	Type	JP2K	JPEG	WN	Blur	All
PSNR	FR	0.852	0.604	0.974	0.779	0.825
SSIM	FR	0.935	0.924	0.914	0.905	0.911
BIQI	Blind	0.573	0.583	0.703	0.640	0.617
DIIVINE	Blind	0.752	0.518	**0.913**	0.850	0.727
BLINDS-II	Blind	0.765	0.733	0.894	0.853	0.810
BRISQUE	Blind	0.744	0.711	0.823	0.901	0.812
NIQE	Blind	0.800	0.746	0.848	0.904	0.811
QAC	Blind	0.757	0.594	0.882	0.899	0.769
Proposed w/o R^3	Blind	0.798	0.785	0.891	0.913	0.805
Proposed w/ R^3	Blind	**0.810**	**0.838**	0.831	**0.915**	**0.827**

In addition, similar to the results in LIVE-II database, the proposed method with R^3 performs better than that without R^3 in this cross-database investigation. It verifies that our R^3 model is robust to the visual content variation across different IQA databases.

In order to show the superiority of the R^3 method in ranking the quality of images, we further build a challenging test set to compare different BIQA algorithms. Particularly, all human rated test samples are selected from the VLC@FER database, which has no visual content overlapping with the training set (i.e., LIVE-II database). For each reference image in the VLC@FER database, we implement all possible pairwise comparisons between its associated distorted images under 4 distortion types and 6 degradation levels, which produces $\binom{24}{2}$, i.e.,

TABLE IX
PLCC AND SROCC BETWEEN THE PREDICTED QUALITY SCORE AND MOS ON THE CHALLENGING SUBSET OF VCL@FER DATABASE

Metric	PLCC	SROCC
BIQI	0.375	0.363
DIIVINE	0.495	0.477
BLINDS-II	0.578	0.572
BRISQUE	0.554	0.566
NIQE	0.556	0.574
QAC	0.451	0.437
Proposed w/o R^3	0.699	0.700
Proposed w/ R^3	**0.717**	**0.721**

TABLE X
MEDIAN SROCC ACROSS 100 TRAIN-TEST TRIALS ON THE LIVEMD DATABASE

	Part 1	Part 2	All
BIQI	0.845	0.793	0.768
DIIVINE	0.828	0.853	0.822
BLIINDS-II	0.884	0.887	0.866
BRISQUE	0.915	0.876	0.874
NQIE	0.835	0.846	0.833
QAC	0.714	0.790	0.731
Proposed w/o R^3	**0.915**	0.895	0.890
Proposed w/ R^3	0.913	**0.929**	**0.911**

276 pairs of images. If more than half of the six representative BIQA metrics (i.e., BIQI, DIIVINE, BLINDS-II, BRISQUE, NIQE and QAC) mistake the perceptual rank-order of a pair of aforementioned images, this pair of images are collected into our challenging test set. Then, we compare the performance of different BIQA algorithms in ranking the quality of these selected images. The PLCC and SROCC results are reported in Table IX. It can be seen that the proposed methods outperform all the other methods in terms of both PLCC and SROCC by large margins.

G. Multiply Distortion Verification

In the real-world application, the natural images usually undergo multiply distortions during their acquisition and postprocessing process. To verify that the proposed method works well in this scenario, we further evaluate the performance of R^3 on the multiply distortion database–LIVEMD [57]. More specifically, the LIVEMD contains 15 pristine images and their 450 distorted versions, which are divided into two parts according to their distortion types. The Parts 1 and 2 are created by adding JPEG and noise to the blurred images, respectively.

Following the criteria in Section IV-A, we implement the verification based on 100 train-test trials on LIVEMD. The training set includes 360 distorted images that are generated from 12 randomly selected reference images, and the other 90 distorted images are used for testing. For comparison, the six state-of-the-art BIQA algorithms [7], [8], [14], [36], [49], [55] are also involved in this verification. The median SROCC results have been reported in Table X, where the highest value are highlighted by bold face in each column. It is seen that the quality predicted by R^3 model is highly consistent with the human perception on the multiply distortion images, whose SROCC is up to 0.911 on the entire test set of LIVEMD. In addition, similar to the results on the single distortion databases, the proposed method is also superior to all the other BIQA metrics in this investigation.

TABLE XI
MEDIAN SROCC ACROSS 100 TRAIN-TEST TRIALS ON THE CHALLENGEDB DATABASE

Metric	PLCC	SROCC
BIQI	0.540	0.519
DIIVINE	0.556	0.527
BLINDS-II	0.576	0.542
BRISQUE	0.625	0.597
NIQE	0.498	0.489
QAC	0.298	0.318
Proposed w/o R^3	0.625	0.612
Proposed w/ R^3	**0.642**	**0.631**

H. Authentic Distortions Verification

In addition to the graded simulated distortions discussed in previous sections, we further investigate the performance of the proposed method in coping with diverse authentic image distortions, which is implemented on the ChallengeDB database [58]. More specifically, there are totally 1162 authentically distorted images in [58], which are captured from diverse mobile devices.

The test criteria still follows the instruction in Section IV-A, and the 100 rounds of random train-test trials are implemented for all BIQA algorithms [7], [8], [14], [36], [49], [55] in comparison. In each train-test split, there are 930 images used for training and the rest 232 images are used for testing. The median SROCC results are reported in Table XI, and the highest value are highlighted by bold face for clarity. Similar to the results on the simulated distortions, the proposed method still achieves the best prediction performance with respect to all the other BIQA metrics. Meanwhile, the proposed method w/ R^3 is still consistently superior to the method w/o R^3.

I. Cross Feature Verification

To verify the proposed method across feature spaces, we compare the performance between the regression models w/o and w/ R^3 constraint by feeding them with six popular quality-aware image features [7], [8], [14], [36], [49], [55]. More specifically, we follow the train-test splitting criterion in Section IV-A and implement the test on LIVE-II database. Let S_{wo} and S_{w} denote the median SROCC produced from the regression models w/o and w/ R^3 constraint, respectively. The difference value d_{R^3}, which is equal to $\mathrm{S}_{w} - \mathrm{S}_{\mathrm{wo}}$, is used to evaluate the contribution of R^3 model. A positive d_{R^3} means that the R^3 constraint improves the quality prediction accuracy and vice versa. The experimental results have been shown in Fig. 6, where the x-axis indicates different features fed to two regression models and y-axis is the SROCC improvement brought by R^3 constraint.

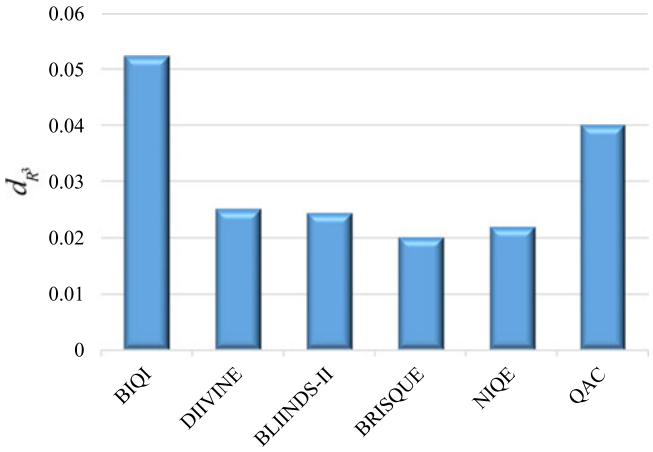


Fig. 6. SROCC improvement brought by R^3 across different features.

TABLE XII
PREDICTION PERFORMANCE OF THE PROPOSED METHOD WITH AND WITHOUT R^3 UNDER DIFFERENT KERNELS

Kemel	Proposed w/o R^3		Proposed w/R^3	
	PLCC	SROCC	PLCC	SROCC
Polynomial	0.943	0.941	0.966	0.960
RBF	0.943	0.940	0.964	0.957

It is seen that the proposed R^3 constraint could consistently improve the performance of regression model across all six image features, where d_{R^3} is always positive as shown in Fig. 6. In our investigation, the maximum SROCC improvement of R^3 is up to 0.052, which is achieved for the BIQI feature. When the state-of-the-art BRISQUE feature is fed to R^3 model, we could still increase the SROCC by 0.02 with respect to a classic SVR model without rank-order constraint.

J. Influence of Different Kernel

Through extensive experiments on five popular databases (i.e., LIVE-II, TID2013, VCLFER, LIVEMD and ChallengeDB), it is validated that the rank-order regularization plays a positive role in improving the prediction performance towards BIQA, where the proposed method w/ R^3 is consistently superior to the one w/o R^3. To further verify the necessary of R^3 in a quality regression model, we compare its influence with the popular radial basis function (RBF) kernel, which is widely used to better explore the training data. More specifically, we re-implement the consistency experiment in Section IV-D, and the RBF kernel is applied to the proposed regression models. The results are reported in Table XII. It is seen that the non-linear RBF kernel does not bring PLCC/SROCC improvements with respect to the polynomial kernel for both the regression models w/ and w/o R^3. As discussed in [59], [60], the kernel type and parameter selection is usually based on application-domain knowledge, which does not necessarily lead to a performance improvement by using one particular kernel. In contrast, when we add R^3 in the regression model, its PLCC/SROCC performance would consistently outperform the model w/o R^3, no matter which kernel is used.

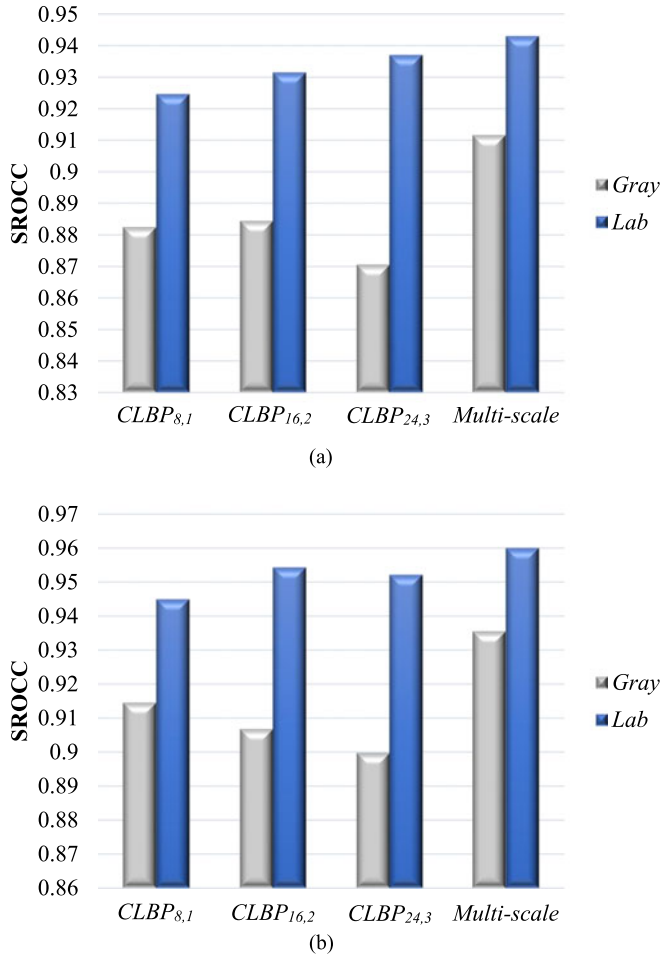


Fig. 7. SROCC performance comparison for the LSS features extracted from different scales and color spaces. (a) Regression model without R^3 constraint. (b) Regression model with R^3 constraint.

K. Contributions of LSS Feature

The proposed LSS feature is extended from classic texture descriptor CLBP [35] by introducing two visual perception clues, i.e., multi-scale and *Lab* color space convention. To verify that these two clues are beneficial in improving the quality prediction accuracy, we investigate the SROCC performance changes by combining each of the two clues with the CLBP. Following the train-test splitting criterion in Section IV-A, we conduct the investigation on the LIVE-II database. Both the regression models w/o and w/ R^3 constraint are tested on different combinations. The detailed results have been reported in Fig. 7, where the x-axis indicates different scales of CLBP and *Multi-scale* corresponds to the feature vector of concatenating $CLBP_{8,1}$, $CLBP_{16,2}$ and $CLBP_{24,3}$. In the legend, we show two color spaces which are used for extracting the image feature, where the *Gray* is initially used in [35] and *Lab* is employed in our LSS

TABLE XIII
MEDIAN SROCC COMPARISON BETWEEN THE PIXEL-WISE AND BLOCK-WISE LSS FEATURE

	JP2K	JPEG	WN	Blur	FF	All
Pixel-wise	0.952	0.954	0.984	0.968	0.923	0.960
Block-wise 3×3	0.937	0.939	0.976	0.946	0.892	0.939
Block-wise 5×5	0.894	0.897	0.956	0.911	0.776	0.891
Block-wise 7×7	0.827	0.867	0.930	0.850	0.733	0.841

feature. The y-axis is the median SROCC produced by different combinations.

It is clear that both the multi-scale and *Lab* color space convention boost the SROCC performance of BIQA task. As shown in Fig. 7(a), the *Multi-scale* feature outperforms all the other single-scale CLBP features as combing with the regression model w/o R^3 constraint. In the *Gray* color space, its maximum SROCC improvement is more than 0.05 with respect to $CLBP_{16,2}$. In the *Lab* color space, the maximum SROCC improvement of *Multi-scale* feature is close to 0.02 in comparison with $CLBP_{8,1}$. Similarly, when we test on the regression model w/ R^3 constraint, the *Multi-scale* feature still consistently outperforms the single-scale CLBP features on both of the color spaces as shown in Fig. 7(b). In addition, the *Lab* color space exhibits superior performance in estimating image quality with respect to the *Gray* in this investigation. As shown in Fig. 7(a) and (b), the median SROCC of the *Lab* features is higher than the *Gray* features across all scales. For the regression model w/o R^3 constraint, the maximum SROCC gain between *Lab* and *Gray* features is achieved by $CLBP_{16,2}$ which could be up to 0.059. As testing on the regression model w/ R^3 constraint, we can also achieve 0.05 SROCC improvement by extracting $CLBP_{24,3}$ from the *Lab* color space in comparison with the *Gray*.

L. Investigation on Block-Wise LSS Feature

The pixel-wise LSS feature describes the spatial complexity in a rather small scale, where only the central and neighboring pixels are involved in the comparison. It is interesting to investigate how a larger center and neighbor regions would affect the discriminating power of the proposed LSS feature. More specifically, we analyze the performance of a block-wise LSS feature, whose center and neighbors are represented by the mean intensity values of a $d \times d$ block whose locations are same with the settings of CLBP as shown in Fig. 3. Three scales of blocks are tested in this investigation, which include 3×3, 5×5, and 7×7. Similar to Section IV-D, we implement 100 rounds of random splitting trials on the LIVE-II database, where the sizes of training and testing sets are 80% and 20%, respectively. The median SROCC results are reported in Table XIII. It is seen that the block-wise LSS is inferior to the pixel-wise LSS feature for the BIQA task. Meanwhile, a larger block size would bring more performance degradation. The possible reason could be found from previous texture descriptor related literatures. As discussed in [61], [62], the block-wise or region based differential operation is widely used for developing robust texture descriptors, which aims to reduce its sensitivity to the noise or distortion. This intention is contrary to the purpose of designing quality or distortion degree relevant image feature, and could reduce its performance in the BIQA task.

TABLE XIV
RUNNING TIME (SECONDS) COMPARISON OF BIQA METRICS

Metric	Running Time
BIQI	1.198
DIIVINE	54.232
BLINDS-II	217.038
BRISQUE	0.565
NIQE	1.214
QAC	0.407
Proposed w/o R^3	8.694
Proposed w/ R	8.931

M. Computational Complexity

To evaluate computational complexity, we compare different BIQA metrics on their running time. The test is performed on a machine with Intel Core 2 processor of speed 2.0 GHz, 6 GB RAM and Windows 7 64-bit. The BIQA models are implemented using MATLAB2013a.

The running time is computed by evaluating a 512×768 color image selected from the LIVE-II database. The detailed results have been shown in Table XIV . It can be seen that the complexity of the proposed methods is moderate in comparison with DIIVINE and BLINDS-II. Meanwhile, the CLBP feature extraction takes up most computations in the proposed method, whose running time is 8.125 seconds for each image. It should be noted that the running time of the proposed methods is obtained from unoptimized MATLAB code. Since the computation of CLBP is completely independent across scales and color channels, a parallel operation can efficiently speed up the LSS feature extraction process.

In addition, as shown in Table XIV, the proposed method with R^3 is only a little slower than that without R^3. It indicates that the added rank-order constraint term does not significantly increase the computational cost of the regression model.

N. Discussion

In the proposed R^3 model, the parameter Δ provides additional flexibility to manage the pairwise samples which are involved in the rank-order constraint. More specifically, a larger threshold Δ would eliminate more pairwise training samples, whose perceptual differences are considered to be indistinguishable. Not surprisingly, the performance of R^3 model would converge to the same value with respect to the classic SVR as Δ removing all rank-order constraints. To investigate this impact, we repeat the train-test experiments on LIVE-II and change the parameter Δ from 0 to 100 at the interval of 10. The median SROCC results are shown in Fig. 8. In addition, the running time is also reported in Fig. 9.

We can find that the median SROCC would gradually decline as Δ increasing. Since more pairwise samples are eliminated

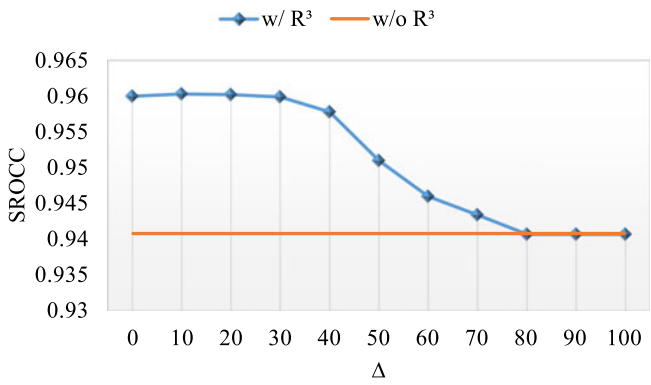


Fig. 8. Median SROCC variation with respect to the parameter Δ.

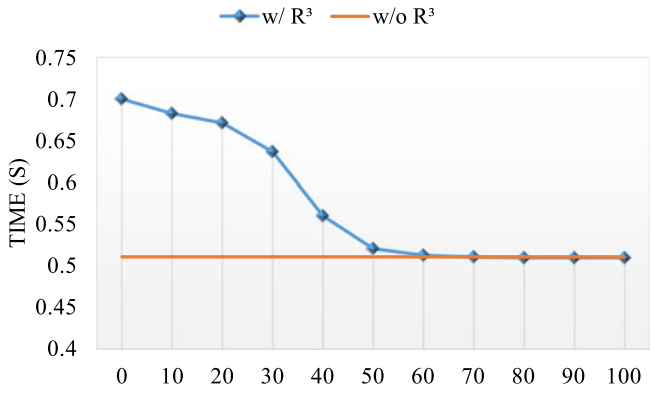


Fig. 9. Running time variation of R^3 with respect to the parameter Δ.

by higher Δ, the correcting effect from rank-order constraint is reduced, which results in the performance degradation of R^3 model. When Δ increases to 80, all rank-order constraints are removed and the performance of R^3 becomes converging to the value of regression model w/o R^3. Meanwhile, as shown in Fig. 9, the similar downtrend in terms of running time can be founded for R^3 model, where the decrease of pairwise samples reduces the complexity of solving the problem in (14).

In addition, another interesting extension of our R^3 model is to integrate it into a deep neural network. To adapt to the back-propagation computation, we can rewrite the constrained optimization task in (3) to a derivable unconstrained loss function, i.e.,

$$\begin{aligned}\mathcal{L} = &\frac{1}{N_a}\sum_i (y_i - \omega * \S_i)^2 + \\ &\lambda \cdot \frac{1}{N_p}\sum_{ij} \max\{0, \varepsilon + \omega * (\S_i - \S_j)\}, \\ &\forall (i,j) \in \mathcal{P}\end{aligned} \tag{24}$$

where ω is the kernel parameter for the last fully connected layer, $*$ denotes the convolution operator, λ is a weight parameter to balance the regression targets between the absolute opinion score estimation and pairwise rank-order preservation. $\S_i$ and $\S_j$ are the outputs from previous convolution layer for the ith and jth image, respectively. N_a and N_p are the total number of training samples for measuring the absolute opinion score and pairwise perceptual preference, respectively.

To facilitate the pairwise comparison, a modified Siamese network would be more suitable for optimizing the loss function in (24). In our future work, some more sophisticated network architecture and regularization terms will be further studied for interpreting the complex multi-target property of human perception towards image quality. Meanwhile, more perceptual properties, such as, masking effect and visual saliency would be studied for building quality-aware features. A more robust general regressor without regard to rank-order information would be explored as well.

V. Conclusion

In this paper, we propose a novel general purpose BIQA algorithm, where the main novelty is in the rank-order regularized regression (R^3) model and the local spatial structure (LSS) feature. Unlike the conventional methods which focus on minimizing the prediction error for labeled training samples only, our R^3 model also tries to preserve perceptual rank-orders. By computing the joint distribution of the encoded central intensity, and the sign and magnitude of contrast, we propose a CLBP based LSS feature that is able to capture the spatial intensity variations and the neighboring pixel dependencies simultaneously. Extensive experiments show that for the LIVE-II, TID2013, VCL@FER, LIVEMD and ChallengeDB databases, the proposed method with R^3 performs highly consistent with human perception of image quality and shows superior performance over state-of-the-art BIQA algorithms.

References

[1] Z. Wang, "Applications of objective image quality assessment methods [applications corner]," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 137–142, Nov. 2011.
[2] 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.
[3] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.
[4] W. Zhou and L. Yu, "Binocular responses for no-reference 3D image quality assessment," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1077–1084, Jun. 2016.
[5] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 339–343.
[6] Q. Wu, H. Li, F. Meng, K. N. Ngan, and S. Zhu, "No reference image quality assessment metric via multi-domain structural information and piecewise regression," *J. Vis. Commun. Image Represent.*, vol. 32, pp. 205–216, 2015.
[7] A. Mittal, A. Moorthy, and A. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[8] 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.
[9] 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.
[10] L. He, D. Tao, X. Li, and X. Gao, "Sparse representation for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 1146–1153.

[11] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, Jul. 2012.

[12] 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.

[13] A. Mittal, G. Muralidhar, J. Ghosh, and A. Bovik, "Blind image quality assessment without human training using latent quality factors," *IEEE Signal Process. Lett.*, vol. 19, no. 2, pp. 75–78, Feb. 2012.

[14] A. Mittal, R. Soundararajan, and A. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[15] 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 Recog.*, Jun. 2014, pp. 1733–1740.

[16] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "A deep neural network for image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 3773–3777.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[18] C. Li, A. 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.

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

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[21] M. Frank and P. Wolfe, "An algorithm for quadratic programming," *Nav. Res. Logist. Quart.*, vol. 3, no. 1–2, pp. 95–110, 1956.

[22] R. J. Vanderbei, "Loqo: An interior point code for quadratic programming," *Optim. Methods Softw.*, vol. 11, no. 1–4, pp. 451–484, 1999.

[23] Y. Ye, *Interior Point Algorithms: Theory and Analysis*. Hoboken, NJ, USA: Wiley, 2011, vol. 44.

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

[25] R. Herbrich, T. Graepel, and K. Obermayer, *Large Margin Rank Boundaries for Ordinal Regression*. Cambridge, MA, USA: MIT Press, Jan. 2000, ch. 7, pp. 115–132.

[26] C. Burges *et al*., "Learning to rank using gradient descent," in *Proc. Int. Conf. Mach. Learn.*, 2005, pp. 89–96.

[27] E. Hüllermeier, J. Fürnkranz, W. Cheng, and K. Brinker, "Label ranking by learning pairwise preferences," *Artif. Intell.*, vol. 172, no. 16, pp. 1897–1916, 2008.

[28] Z. Chen, T. Jiang, and Y. Tian, "Quality assessment for comparing image enhancement algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2014, pp. 3003–3010.

[29] J. Yan, S. Lin, S. B. Kang, and X. Tang, "A learning-to-rank approach for image color enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2987–2994.

[30] A. Reibman, K. Shirley, and C. Tian, "A probabilistic pairwise-preference predictor for image quality," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 413–417.

[31] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Proc. 14th Annu. Conf. Comput. Learn. Theory*, Jan. 2001, pp. 416–426.

[32] O. L. Mangasarian, "Generalized support vector machines," in *Advances in Large Margin Classifiers*. Cambridge, MA, USA: MIT Press, 1998, pp. 135–146.

[33] E.-H. Han, G. Karypis, and V. Kumar, "Text categorization using weight adjusted k-nearest neighbor classification," in *Proc. Adv. Knowl. Discovery Data Mining*, 2001, vol. 2035, pp. 53–65.

[34] H. Zhang, A. Berg, M. Maire, and J. Malik, "SVM-KNN: Discriminative nearest neighbor classification for visual category recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, vol. 2, pp. 2126–2136.

[35] Z. Guo, D. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jun. 2010.

[36] A. Moorthy and A. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.

[37] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.

[38] J. Wu, W. Lin, and G. Shi, "Image quality assessment with degradation on spatial structure," *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 437–440, Apr. 2014.

[39] M. Zhang, J. Xie, X. Zhou, and H. Fujita, "No reference image quality assessment based on local binary pattern statistics," in *Proc. Vis. Commun. Image Process.*, Nov. 2013, pp. 1–6.

[40] M. Panić, D. Ćulibrk, S. Sladojević, and V. Crnojević, "Local binary patterns and neural networks for no-reference image and video quality assessment," in *Proc. Eng. Appl. Neural Netw.*, 2013, pp. 388–395.

[41] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[42] V. Bruce, P. R. Green, and M. A. Georgeson, *Visual Perception: Physiology, Psychology, & Ecology*. New York, NY, USA: Psychology Press, 2003.

[43] S. Hochstein and M. Ahissar, "View from the top: Hierarchies and reverse hierarchies in the visual system," *Neuron*, vol. 36, no. 5, pp. 791–804, 2002.

[44] R. P. Rao and D. H. Ballard, "Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects," *Nature Neurosci.*, vol. 2, no. 1, pp. 79–87, 1999.

[45] A. Ford and A. Roberts, "Colour space conversions," *Westminster University,* London, U.K., vol. 1998, pp. 1–31, 1998.

[46] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE Image Quality Assessment Database release 2," Lab. Image Video Eng., Univ. Texas Austin, Austin, TX, USA, Accessed on: 2005. [Online]. Available: http://live.ece.utexas.edu/research/quality

[47] N. Ponomarenko *et al*., "A new color image database TID2013: Innovations and results," in *Proc. Adv. Concepts Intell. Vis. Syst.*, 2013, vol. 8192, pp. 402–413.

[48] A. Zaric *et al*., "VCL@FER image quality assessment database," in *Proc. ELMAR*, Sep. 2011, pp. 105–110.

[49] M. 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.

[50] Video Quality Experts Group, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase I (FR_TV)," 2000. [Online]. Available: http://www.vqeg.org/

[51] K. Ma *et al*., "Group mad competition—A new methodology to compare objective image quality models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2016, pp. 1664–1673.

[52] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. IEEE Int. Symp. Comput. Aided Control Syst. Des.*, Sep. 2004, pp. 284–289.

[53] *Gurobi Optimizer Reference Manual*, Gurobi Optimization, Inc., Houston, TX, USA, 2014. [Online]. Available: http://www.gurobi.com

[54] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

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

[56] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, FL, USA: CRC Press, 2003.

[57] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Proc. 46th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2012, pp. 1693–1697.

[58] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, Jan. 2016.

[59] O. Chapelle and V. Vapnik, "Model selection for support vector machines," in *Proc. Adv. Neural Inf. Process. Syst.*, 1999, pp. 230–236.

[60] B. Schölkopf, C. J. C. Burges, and A. J. Smola, *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA, USA: MIT Press, 1999.

[61] T. Song, H. Li, F. Meng, Q. Wu, B. Luo, B. Zeng, and M. Gabbouj, "Noise-robust texture description using local contrast patterns via global measures," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 93–96, Jan. 2014.

[62] F. M. Khellah, "Texture classification using dominant neighborhood structure," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3270–3279, Nov. 2011.

Qingbo Wu (S'12–M'15) received the B.E. degree in education of applied electronic technology from Hebei Normal University, Shijiazhuang, China, in 2009, and the Ph.D. degree in signal and information processing from the University of Electronic Science and Technology of China, Chengdu, China, in 2015.

From February 2014 to May 2014, he was a Research Assistant with the Image and Video Processing Laboratory, Chinese University of Hong Kong, Hong Kong, China. From October 2014 to October 2015, he was a Visiting Scholar with the Image and Vision Computing Laboratory, University of Waterloo, Waterloo, ON, Canada. 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 (M'06–SM'11) received the Ph.D. degree in electronics and information engineering from Xi'an Jiaotong University, Xi'an, China, in 2005.

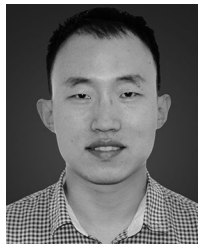
From 2005 to 2006, he was with the Visual Signal Processing and Communication Laboratory, Chinese University of Hong Kong (CUHK), Hong Kong, China, as a Research Associate, where he was a Postdoctoral Fellow from 2006 to 2008. He is currently a Professor with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China. He has authored or coauthored numerous technical articles in international journals and conferences. He is the co-editor of *Video Segmentation and Its Applications* (Springer). He was involved in many professional activities. His research interests include image segmentation, object detection, image and video coding, visual attention, and multimedia communication system.

Prof. Li is a Member of the Editorial Board of the *Journal on Visual Communications and Image Representation*, and the Area Editor of *Signal Processing: Image Communication*, Elsevier Science. He was a Technical Program Co-Chair of ISPACS 2009, General Co-Chair of the ISPACS 2010, Publicity Co-Chair of IEEE VCIP 2013, Local Chair of the IEEE ICME 2014, and TPC Member in a number of international conferences, e.g., ICME 2013, ICME 2012, ISCAS 2013, PCM 2007, PCM 2009, and VCIP 2010. He is currently a Technical Program Co-Chair of the IEEE VCIP 2016.

Zhou Wang (S'99–A'02–M'03–SM'12–F'14) received the Ph.D. degree from The University of Texas at Austin, Austin, TX, USA, in 2001.

He is currently a Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He has more than 100 publications with over 30 000 citations (Google Scholar). His research interests include image processing, coding, and quality assessment, computational vision and pattern analysis, multimedia communications, and biomedical signal processing.

Prof. Wang is a Fellow of the Canadian Academy of Engineering. He is a Senior Area Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2015–present), and an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2016–present). He was a Member of the IEEE Multimedia Signal Processing Technical Committee (2013–2015), an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING (2009–2014), Pattern Recognition (2006–present) and the IEEE SIGNAL PROCESSING LETTERS (2006–2010), and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (2013–2014 and 2007–2009). He was the recipient of the 2017 IEEE Signal Processing Society Sustained Impact Paper Award, 2015 Primetime Engineering Emmy Award, 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, 2013 IEEE Signal Processing Magazine Best Paper Award, 2009 IEEE Signal Processing Society Best Paper Award, 2009 Ontario Early Researcher Award, and ICIP 2008 IBM Best Student Paper Award (as a senior author).



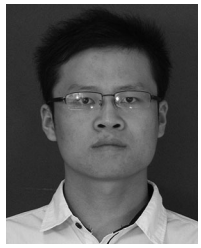
Fanman Meng (S'12–M'13) received the Ph.D. degree in signal and information processing from the University of Electronic Science and Technology of China, Chengdu, China, in 2014.

From July 2013 to July 2014, he joined the Division of Visual and Interactive Computing, Nanyang Technological University, Singapore, as a Research Assistant. He is currently an Associate Professor with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China. He has authored or coauthored numerous technical articles in well-known international journals and conferences. His research interests include image segmentation and object detection.

Prof. Meng is a Member of IEEE CAS society. He was the recipient of the "Best Student Paper Honorable Mention Award" for the 12th Asian Conference on Computer Vision (2014), Singapore, and the "Top 10% Paper Award" in the IEEE International Conference on Image Processing (2014), Paris, France.

Bing Luo received the B.Sc. degree in communication engineering from The Second Artillery Command College, Wuhan, China, in 2009, the M.Sc. degree in computer application technology from Xihua University, Chengdu, China, in 2012, and is currently working toward the Ph.D. degree in electronic engineering at the University of Electronic Science and Technology of China, Chengdu, China, supervised by Prof. Hongliang Li.

His research interests include image and video segmentation and machine learning.



Wei Li received the B.Sc. degree in electrical and information engineering from Henan Polytechnic University, Jiaozuo, China, in 2011, and is currently working toward the Ph.D. degree in the Intelligent Visual Information Processing and Communication Laboratory, University of Electronic Science and Technology of China, Chengdu, China.

His research interests include image recognition, object detection, and machine learning.

King N. Ngan (S'79–M'82–SM'91–F'00) received the Ph.D. degree in electrical engineering from the Loughborough University, Loughborough, U.K., in 1982.

He is currently a Chair Professor with the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong, China. He was previously a Full Professor with the Nanyang Technological University, Singapore, and the University of Western Australia, Perth, WA, Australia. Since 2012, he has been the Chair Professor with the University of Electronic Science and Technology, Chengdu, China, under the National Thousand Talents Program. He holds honorary and visiting professorships of numerous universities in China, Australia, and South East Asia. He has published extensively including three authored books, seven edited volumes, over 400 refereed technical papers, and edited nine special issues in journals. In addition, he holds 15 patents in the areas of image/video coding and communications.

Prof. Ngan is a Fellow of IET (U.K.), IEAust (Australia), and an IEEE Distinguished Lecturer (2006–2007). He was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *Journal on Visual Communications and Image Representation*, the EURASIP *Journal of Signal Processing: Image Communication*, and the *Journal of Applied Signal Processing*. He chaired and co-chaired a number of 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.