Multi-Task Rank Learning for Image Quality Assessment

Long Xu, *Member, IEEE*, Jia Li, *Senior Member, IEEE*, Weisi Lin, *Fellow, IEEE*, Yongbing Zhang, *Member, IEEE*, Lin Ma, *Member, IEEE*, Yuming Fang, *Member, IEEE*, and Yihua Yan

***Abstract*— In practice, images are distorted by more than one distortion. For image quality assessment (IQA), existing machine learning (ML)-based methods generally establish a unified model for all the distortion types, or each model is trained independently for each distortion type, which is therefore distortion aware. In distortion-aware methods, the common features among different distortions are not exploited. In addition, there are fewer training samples for each model training task, which may result in overfitting. To address these problems, we propose a multi-task learning framework to train multiple IQA models together, where each model is for each distortion type; however, all the training samples are associated with each model training task. Thus, the common features among different distortion types and the said underlying relatedness among all the learning tasks are exploited, which would benefit the generalization ability of trained models and prevent overfitting possibly. In addition, pairwise image quality ranking instead of image quality rating is optimized in our learning task, which is fundamentally departed from traditional ML-based IQA methods toward better performance. The experimental results confirm that the proposed multi-task rank-learning-based IQA metric is prominent against all state-of-the-art nonreference IQA approaches.**

***Index Terms*— Image quality assessment (IQA), machine learning (ML), mean opinion score (MOS), pairwise comparison, rank learning.**

Manuscript received March 10, 2015; revised June 26, 2015, September 26, 2015, November 18, 2015, and December 17, 2015; accepted January 27, 2016. Date of publication March 16, 2016; date of current version September 5, 2017. The work of L. Xu was supported by National Natural Science Foundation (NSFC) of China Grants 61202242 and 61572461 and CAS 100-Talents. The work of J. Li was supported by the National Natural Science Foundation of China under Grant 61370113. The work of Y. Zhang was supported by the National Natural Science Foundation of China within the Guangdong Joint Fund under Grant U1201255 and Grant U1301257. The work of Y. Fang was supported by the National Natural Science Foundation of China under Grant 61571212. The work of Y. Yan was supported by the National Natural Science Foundation of China under Grant 11433006. This paper was recommended by Associate Editor W. Zeng. *(Corresponding author: Yongbing Zhang.)*

L. Xu and Y. Yan are with the Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China (e-mail: lxu@nao.cas.cn; yyh@nao.cas.cn).

J. Li is with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China, and also with International Research Institute for Multidisciplinary Science, Beihang University, Beijing 100191, China (e-mail: jiali@buaa.edu.cn).

W. Lin is with the Department of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wslin@ntu.edu.sg).

Y. Zhang is with the Graduate School at Shenzhen, Tsinghua University, Shenzhen 518005, China (e-mail: zhang.yongbing@sz.tsinghua.edu.cn).

L. Ma is with the Huawei Noah's Ark Laboratory, Hong Kong (e-mail: forest.linma@gmail.com).

Y. Fang is with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: fa0001ng@e.ntu.edu.sg).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author.

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.2016.2543099

I. Introduction

THREE categories of nonreference (NR) IQA approaches were presented in the literature in terms of methodology. The *first* category takes the behavior of specific distortions into consideration. Sheikh *et al.* [1] employed a wavelet statistical model to capture the distortion introduced by JPEG 2000. Liang *et al.* [2] combined the sharpness, blurring, and ringing measurements together to evaluate the visual quality of the JPEG 2000 coded image. Brandão and Queluz [3] proposed an NR image quality assessment (NR-IQA) approach based on the discrete cosine transform (DCT) domain statistics to evaluate the quality of JPEG coded image. The *second* category uses quality-aware clustering. They group the image patches of training set into the given number of classes based on local image features, such as histogram of oriented gradients, difference of Gaussian, and Gabor filter. Each cluster center has a quality score that is derived from the qualities of image patches falling into this cluster. Associating cluster centers with their qualities, the researchers established a codebook. Patches of a test image look up codebook to search the most similar codewords and retrieve the associated quality values. In [4], a visual codebook associated Gabor-filter-based local appearance descriptors with the mean opinion score (MOS). Xue *et al.* [5] used FSIM [6] instead of MOS as image patch quality to establish codebook. The *third* category is to utilize the machine learning (ML) tool to map image features onto image qualities. Moorthy and Bovik [8] proposed to use support vector machine (SVM) and support vector regression (SVR) [9], [10] to learn a classifier and an ensemble of regressors for distortion-aware image quality assessment (IQA). It deploys summary statistics called natural scene statistics (NSS), which is derived from wavelet decomposition of an image. Tang *et al.* [7] proposed an approach similar to [8] but with more elaborate features, including distortion texture statistics, blur/noise statistics, and histogram of each sub-band of image decomposition. Ye *et al.* [11] used unsupervised learning to learn a dictionary over raw image patches for IQA. In addition, SVR was also used to train the IQA

1051-8215 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

model for image quality prediction. Xue *et al.* [12] proposed to use the joint statistics of gradient magnitude (GM) and Laplacian of Gaussian (LOG) as an image feature, and SVR was also employed to learn the model for image quality prediction. In [13], the multiple kernel learning (MKL) was investigated for distortion-specific IQA. Most recently, the deep neural network was successfully applied to IQA [14]–[16]. In [17], image quality preference in pairs was explored to lead to a rank learning optimization problem, and SVR with multiple kernels were adopted to solve this optimization problem. In [18], image quality ranking rather than rating was investigated for evaluating image-enhanced algorithms. Relative to conventional image quality rating models, image quality ranks employ rank learning tools [19], [20] for solving optimization.

The ML-based models can be arranged into two categories: distortion-aware and distortion-unaware. The former usually concerns distortion type identification and quality prediction to form a two-step scheme. It is associated with at least one of the problems as follows: 1) allocating training samples into several clusters, there would be fewer training samples for each training cluster, and this possibly results in overfitting and weak generalization capability of the trained model and 2) the common features shared by different distortion types are not exploited efficiently. Therefore, a combination of emphasize individuality and exploit commonness would create better performance. Toward this end, we propose multi-task rank-learning-based IQA (MRLIQ). In this approach, we construct multiple IQA models, each of which is responsible for one distortion type in order that each model can accurately describe the specific characteristics of each distortion type. Different from the single-task learning approach, the tasks for training these IQA models are combined to form a unified learning task; therefore, these models are trained together. Thus, the relatedness and information sharing across multiple training tasks are effectively exploited to improve the generalization ability of trained models. In addition, departing from the conventional ML-based IQAs, the proposed approach optimizes pairwise image quality rank instead of numerical image quality rating, as stated in [21].

The rest of this paper is arranged as follows. Section II describes the proposed MRLIQ in detail. Section III presents the experimental results. Section IV concludes this paper.

II. Multi-Task Rank Learning for Image Quality Assessment

As with single-task learning, each task is trained independently of the others. Only a subset of the whole training set related to that task (e.g., the distortion type) is used for training, which ignores the intrinsic relatedness among different tasks. For example, one can train multiple IQA models for multiple distortion types, each of which is responsible for one distortion type. Two problems are with such processing. One is that *only a small number of training samples are available for each task*, which possibly results in overfitting and harms the generalization ability of the trained model. The other is that the *common features among different distorted images are not exploited*. Multi-task learning is different from single-task learning in that the multiple tasks are trained together instead of independently, enabling all of the samples in the training set to be used for training each task, while each sample has a set of weights accounting for the importance of the sample to each task. Therefore, the multi-task scheme can exploit underlying intrinsic relatedness among multiple tasks and capture shared/common information of training data. It has attracted extensive attentions in many domains, such as information retrieval [23], [24] and visual saliency modeling [25], [26]. To the best of our knowledge, the multi-task learning has not been applied or discussed in IQA yet.

A. Single-Task Rank Learning

In [21], a single-task rank learning algorithm [19], [20] was proposed, where only one task and therefore one ranking model were learned. Given a subjective image database, we represent the image features $\{x_u \in R^L\}$ and the corresponding MOSs $\{y_u \in R\}$, $u = 1, 2, \ldots, M$. Thus, the goal of a rank-learning-based IQA can be described as identifying the ranks of $\{x_u\}$ with respect to $\{y_u\}$. Toward this end, we infer a ranking function $\varphi : x \to R$ trained on the basis of $\{x_u, y_u\}$ to assign rank order to each x_u. That is, $\varphi(x_u) > \varphi(x_v)$ indicates that x_u ranks higher than x_v with respect to image quality. In [21], only one task and, therefore, a distortion-unaware ranking model φ were trained on all the distorted images as

$$\min_{\omega} \left\{ \sum_{\boldsymbol{u} \neq \boldsymbol{v}}^{\boldsymbol{M}} [\boldsymbol{y_u} < \boldsymbol{y_v}]_I [\varphi(\boldsymbol{x_u}) \geq \varphi(\boldsymbol{x_v})]_I \right\} \tag{1}$$

for all types of distortions (e.g., five distortions for the LIVE image database), where φ is usually assumed to be a linear function, i.e., $\varphi(x) = \omega^T \cdot x$ for simplicity, and x and ω denote the extracted features and the parameter of the ranking function, respectively.

B. Multi-task Rank Learning

For distortion-specific purposes, we can directly apply the proposed single-task learning [21] to learn a specific model for each distortion. Such a model is specific to the distortion type and therefore has better prediction accuracy of quality assessment for that particular distortion type. However, single-task learning ignores the relatedness or commonness among different distortions and therefore results in efficiency loss.

We assume to group training samples into K clusters $\{S_j\}$, $j = 1, 2, \ldots, K$, with respect to their distortion types. An independent learning task is assigned to each cluster for training a specific model for each cluster. Meanwhile, these tasks are integrated into a unified learning framework by sharing the relatedness or commonness across different clusters. This is therefore named multi-task learning [22]. Let x_u^j ($u = 1, 2, \ldots, m_j$, $j = 1, 2, \ldots, K, \sum_{j=1}^{K} m_j = M$) be the uth image in the jth cluster, and y_u^j be the corresponding label (MOS or DMOS in IQA). The jth model φ_j is trained on the

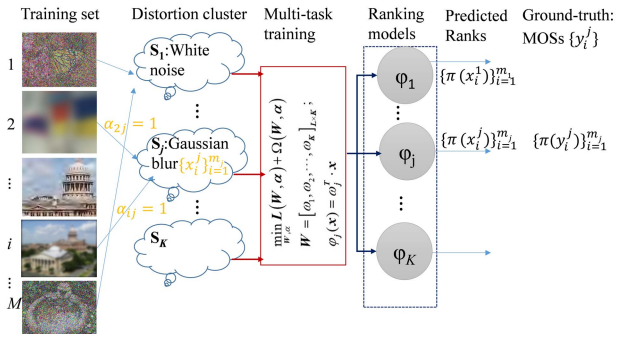

Fig. 1. Proposed multi-task rank learning framework.

cluster S_j consisting of m_j distorted images and associated labels $\{(x_u^j, y_u^j)\}$ as

$$\min_{\omega_j}\left\{\sum_{u\neq v}^{m_j}\left[y_u^j < y_v^j\right]_I\left[\varphi_j\left(x_u^j\right)\geq\varphi_j\left(x_v^j\right)\right]_I\right\} \tag{2}$$

where $\varphi_j(x)=\omega_j^T\cdot x$. Compared with (1), m_j training samples only concerning the jth task are used for training φ_j in (2). Applying (2) to K clusters, there would be K models, each of which is specific to one distortion type. Although the model is specific to distortion type, there are two problems as mentioned earlier in the beginning of this section. Therefore, we refer to a multi-task learning scheme to jointly train multiple models to avoid these two problems in this paper.

Comparing with the single-task case, multi-task learning means that: 1) all the models $\{\varphi_j, j=1,2,\ldots,K\}$ concerned are trained jointly instead of independently [see (1) and (2)] and 2) all the training samples in a training set participate in each model training task. Obviously, a sample distorted by a certain distortion type should contribute more than others to the model specific to that distortion type. To explore the contributions of a given sample to all the models, we define a matrix of distortion cluster labels $\boldsymbol{\alpha}=\{\alpha_{uj}\}$. Initially, we set $\alpha_{uj}\in\{0,1\}$, where $\alpha_{uj}=1$ indicates that a sample x belongs to the cluster S_j. After applying a weight α_{uj} to a sample, this sample could be used in all the training tasks. In practice, it is impractical and unnecessary to establish a different model for each distortion type since there are more than 30 commonly used distortions, and some of them are similar in characteristic. We usually cluster similar distortions into the same cluster in order that each cluster has sufficient samples for training models. According to distortion characteristic, the general distortions can be grouped into four to six classes, such as JP2K compression, JPEG compression, white noise, blurring, transmission error, color saturation, and so on. In addition, the clustering methods, such as K-means and C-means [27], [28], can be employed for image distortion clustering.

Fig. 1 describes the framework of multi-task learning. $\{\pi(x_u^j)\}_{u=1}^{m_j}$ represents the rank list of $\{x_u^j\}_{u=1}^{m_j}$ of the jth cluster using the corresponding φ_j. $\{\pi(y_u^j)\}_{u=1}^{m_j}$ is the rank list of MOSs by comparing their numerical values, which is the ground truth in the proposed learning algorithm. From Fig. 1, the images are grouped into K clusters. The models $\{\varphi_j\}$ are trained together instead of independently as in (2). As illustrated by the red rectangle in Fig. 1, all the samples are involved in a unified training task and the multi-task training outputs K models.

For the convenience of statement, we first formulate the whole optimization function and then detail each element. Let $\boldsymbol{W}$ be a $N\times K$ matrix with the jth column equal to ω_j, the bold $\boldsymbol{\alpha}$ be a $M\times K$ matrix with the uth row equal to $\{\alpha_{uj}\}_{j=1}^K$, which is a K-D vector. Each of the entries of this vector represents the weight of the uth sample as to the jth learning task. Therefore, the objective for jointly training multiple IQA models in a unified framework can be formulated as

$$\begin{aligned}&\min_{\boldsymbol{W},\boldsymbol{\alpha}}\ \boldsymbol{L}(\boldsymbol{W},\boldsymbol{\alpha})+\Omega(\boldsymbol{W},\boldsymbol{\alpha})\\&\text{s.t.}\ \sum_{j=1}^{K}\alpha_{uj}=1;\quad 1\leq u\leq M\\&\qquad\alpha_{uj}\in[0,1];\quad 1\leq j\leq K\end{aligned} \tag{3}$$

where $\boldsymbol{L}(\boldsymbol{W},\boldsymbol{\alpha})$ is the empirical loss and $\boldsymbol{\Omega}(\boldsymbol{W},\boldsymbol{\alpha})$ is the penalty term.

1) Empirical Loss: The empirical loss accounts for the falsely ranked image pairs with regard to their MOS ranks as

$$\boldsymbol{L}(\boldsymbol{W},\boldsymbol{\alpha})=\sum_{v=1}^{M}\sum_{j=1}^{K}\alpha_{vj}\sum_{u\neq v}^{M}[y_u<y_v]_I\left[\omega_j^T\boldsymbol{x_u}\geq\omega_j^T\boldsymbol{x_v}\right]_I \tag{4}$$

where $[p]_I=1$ if p holds; otherwise, $[p]_I=0$. In (4), each pair of images (x_u and x_v, $u\neq v$) in the whole training set (M samples) is compared. Their MOSs are given by y_u and y_v, respectively. The ranks of y_u and y_v give the ground truth of the comparison of x_u and x_v. If the ranks of $\omega_j^T\boldsymbol{x}_u$ and $\omega_j^T\boldsymbol{x}_v$ conflict with the ground truth, the empirical loss $L(\boldsymbol{W},\alpha)$ would increase. From (3), $\sum_j\alpha_{vj}=1$ for image v, so α_{vj} can be regarded as the probability of image v as to the jth task. In (4), α_{vj} is initialized to 1 or 0, indicating that x_v is in the jth cluster or not. For example, for an image x_v, $\alpha_{vj}=\{0,1,\ldots,0\}$ ($\alpha_{v2}=1$) implies that this image is in the second cluster. In the training process as described later, $\boldsymbol{\alpha}$ will be updated for each iteration, and thus $\boldsymbol{\alpha}$ becomes a weight vector. Each element of it represents the weight/probability of an image as to each learning task. By this way, each image actually participates in all learning tasks for training multiple IQA models. With regard to image clustering, the method in [26] is still applicable to our work. Besides, the clustering with regard to distortion can be referred to the prior knowledge about distortion characteristics.

2) Distortion Clustering: For image clustering with respect to distortion characteristics in this case, a penalty term is defined as

$$\Omega_1=\frac{1}{\boldsymbol{M}}\sum_{v=1}^{M}\sum_{u\neq v}^{M}\sum_{j=1}^{K}(\alpha_{uj}-\alpha_{vj})^2\cos\langle\boldsymbol{x_u},\boldsymbol{x_v}\rangle \tag{5}$$

where $\cos\langle\rangle$ denotes the similarity of the uth and the vth image ($u,\ v=1,2,\ldots,M$), which is computed as the cosine distance between two feature vectors x_u and x_v. If x_u and x_v belong to the same cluster and they are correctly clustered, there is no penalty increase since $\alpha_{uj}-\alpha_{vj}=0$ for any j. Otherwise, if two images in the same cluster were

mistakenly clustered into two different clusters, the penalty would increase. Thus, $\boldsymbol{\alpha}$ is responsible for image clustering and named image clustering parameter.

3) Model Correlation: To address the problem of lacking the generalization ability with training these clusters, respectively, an appropriate sharing of information across training tasks can avoid overfitting and improve the performance of each model. Then the penalty term can be defined as

$$\Omega_2 = \frac{1}{K}\sum_{i\neq j}^{K}\sum_{u\neq v}^{M}[y_u < y_v]_I \times \left[\omega_i^T x_u \geq \omega_i^T x_v\right]_I \left[\omega_j^T x_u \geq \omega_j^T x_v\right]_I. \quad (6)$$

The influence of this penalty is twofold. First, a sample [image pair (x_u, x_v)] mistakenly predicted by most models will be emphasized in training the model φ_i since the term $\sum_{j\neq i}[y_u < y_v]_I \times [\omega_j^T x_u \geq \omega_j^T x_v]_I$ increases with the number of samples that are wrongly predicted by most of the models $\{\varphi_i\}$, $j \neq i$. This ensures the diversity of training samples for φ_i, leading to improved generalization ability. Second, a sample successfully predicted by most models will be ignored in training φ_i since the term $\sum_{j\neq i}[y_u < y_v]_I \times [\omega_j^T x_u \geq \omega_j^T x_v]_I$ is small if the samples are successfully predicted by most of the models. This guarantees the diversity of different models. With this penalty term, each task is actually related to all the training samples, as stated in [26], leading to improved performance.

4) Model Complexity: To avoid optimizing complex models, the penalty term can be defined as

$$\Omega_3 = \sum_{j=1}^{K}\omega_j^T\omega_j \quad (7)$$

which amounts to imposing sparse constraints on trained models.

With these penalty terms, the overall penalty can be written as the weighted linear combination of them

$$\Omega(W, \alpha) = l_1\Omega_1 + l_2\Omega_2 + l_3\Omega_3 \quad (8)$$

where l_1, l_2, and l_3 are three nonnegative weights to combine these penalty terms. These three weights are set to (0.60, 0.25, and 0.15, respectively) empirically in this work.

C. Feature Extraction

In learning-based methods, an image is first represented by image features as input of the training process. For testing the proposed MRLIQ, NSS [8] and GM + LOG M3 [12] are employed as image features. We denote MRLIQ with the NSS by MRLIQ-I and MRLIQ with the GM + LOG M3 by MRLIQ-II. NSS measures the unnaturalness of distorted images with respect to natural ones [8]. GM + LOG M3 [12] gathers the statistics of two types of local contrast features: the GM map and the LOG response to form image features.

D. Training Process

By plugging (4)–(6) back to (3), we encounter a nonconvex optimization problem due to the function $[p]_I$. As in [21], the Boolean terms related to the variable ω are replaced by their upper bounds to facilitate the optimization as

$$\left[\omega_j^T x_u \geq \omega_j^T x_v\right]_I \leq e^{(\omega_j^T x_u - \omega_j^T x_v)} \triangleq \eta_{uv}^j \quad (9)$$

where the exponential upper bound is used since it is convex and can facilitate the optimization. After the terms containing the variable ω in (4) and (6) are replaced by those in (9), the optimization function (3) would turn out to be convex. In addition, two variables W and $\boldsymbol{\alpha}$ are correlated in (3). Thus, the expectation–maximization (EM) algorithm is employed to optimize W and $\boldsymbol{\alpha}$ alternately.

First, the training images are clustered into several clusters, usually 4–6 clusters. The clustering methods, such as K-means, can be referred. In this paper, each image clustering is simply defined over each distortion type. Second, the image clustering parameter $\boldsymbol{\alpha}$ is initialized for each image. After that, W is initialized by minimizing (4) without penalty consideration. After initialization, W and $\boldsymbol{\alpha}$ are optimized iteratively using the EM algorithm. The detail algorithm steps can be referred to in [26]. For the integration of the statement of this paper, we outline the algorithm as the following two steps.

Step 1: For $v = 1, \ldots, M$, update $\alpha_v = [\alpha_{v1}, \ldots, \alpha_{vK}]$ by solving the function that contains only the terms related to α_v

$$\min_{\alpha_v} \sum_{j=1}^{K}\alpha_{vj}\sum_{u\neq v}^{M}[y_u < y_v]_I\eta_{uv}^j + \frac{2l_1}{M}\sum_{u\neq v}^{M}\sum_{j=1}^{K}(\alpha_{uj} - \alpha_{vj})^2\cos\langle x_u, x_v\rangle$$
$$\text{s.t.} \sum_{k=1}^{K}\alpha_{vk} = 1; \quad \alpha_{vk} \in [0, 1] \quad \forall k \quad (10)$$

which is a quadratic formula about α and can be solved by quadratic programming. In (10), the first term optimizes the predicted image ranks in order that they are in accord with the ground truth. The second term tries to make sure that the similar images (high correlation between their feature vectors x_u and x_v) should be categorized into the same cluster.

Step 2: To optimize W, the optimization objective concerning only the variable W is

$$\min_{W} \sum_{v=1}^{M}\sum_{j=1}^{K}\alpha_{vj}\sum_{u\neq v}^{M}[y_u < y_v]_I\eta_{uv}^j + \frac{l_2}{K}\sum_{v=1}^{M}\sum_{u\neq v}^{M}[y_u < y_v]_I\sum_{i\neq j}^{K}\eta_{uv}^i\eta_{uv}^j + l_3\sum_{j=1}^{K}\omega_j^T\omega_j \quad (11)$$

which is convex and differentiable. Therefore, the gradient decedent method can be employed to solve (11). From (11), we can see that each model is actually optimized on the whole training set by using the parameter $\boldsymbol{\alpha}$, which weights each image to each task.

Given a new image, a soft SVM classifier [8] is first employed to compute the probability/weight of the image as to

each distortion type. Second, we compute the image qualities of the image using the multiple models $\{\varphi_j, j = 1, 2, \ldots, K\}$. Then, these qualities are weighted by the probabilities/weights to give the final quality of the image.

E. Converting Image Quality Rank to Physical Quantity of Image Quality

In pairwise rank learning framework, the optimization objective function (3) is established on pairwise comparison (PC) of image quality instead of numerical image quality rating. Thus, only image quality ranks can be given to the test images by the trained models. To give a physical quantity of image quality to an image, image quality ranks need to be further converted into image quality scores. In [17], a term gain was defined as the number of times of that an image is preferable against the others. The gain of an image is proportional to its perceived quality because the preference (better or worse image quality) essentially reflects the relative quality (ranking instead of numerical value) of this image relative to others. In addition, a linear mapping between the gain and the quality score was assumed and fitted by training data. Thus, after getting the ranks of images, their quality scores can be deduced by the two steps mentioned above. However, [17] needs to compute the differences between the test image and the training images. Thus, at least, the features of training images are needed, which is impractical in real applications. In [18], only the competition of image enhancement algorithms was concerned without the need of conversion from image quality rank to physical quantity of the image quality. The proposed MRLIQ can output a rank list of all the images instead of only binary preference of each of the two images as in [17] and [18]. Therefore, a mapping function between the rank list and image quality scores can be deduced from MRLIQ directly as follows. In the training stage of MRLIQ, a nonlinear fitting function is deduced from mapping the predicted rank list onto MOS rank list. In the test stage, the nonlinear fitting function can tell quality score of each test image without the need of any information from the training set.

The ith image is compared with other images in the training set. We count the times of failures of model prediction, i.e., predicted image ranks violate ground truth, and compute the accumulative image quality difference (AIQD) as

$$g(i) = \frac{\sum_{i \neq j}^{n} [\omega^T x_i - \omega^T x_j]_+}{\sum_{i \neq j}^{n} [\omega^T x_i > \omega^T x_j]_I}. \tag{12}$$

Since $[\omega^T x_i - \omega^T x_j]_+$ indicates the quality of the ith image relative to the jth image, (12) can represent the relative quality of the ith image against others. We draw the relationship between $g(i)$ and MOSs in Fig. 2. A nice fitting curve can be observed from Fig. 2. In addition, the shape of this mapping function is well fitted by an exponential function

$$q(i) = \beta_1 + \beta_2 \times e^{\beta_3 \times g(i)}. \tag{13}$$

The parameters can be easily deduced from nonlinear least squares regression (implemented by *nlinfit* function in MATLAB). One can derive image quality score for the given image quality rank from (13). The advantage of (13) lies in that accumulative summation reduces the interference of noise, so a good fitting curve is deduced as shown in Fig. 2. Observing (4)–(7), MRLIQ is based on accumulative summation, so AIQD is reasonable and fits for representing relative image quality. The experiments performed in Section IV prove its good performance.

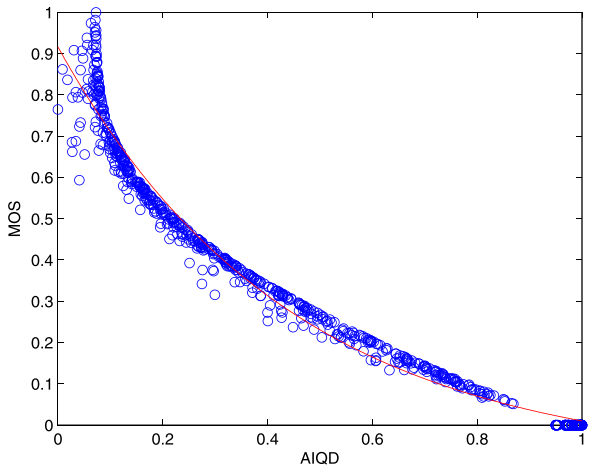

Fig. 2. Mapping between relative quality AIQD ($g(i)$) and MOSs: AIQD computed from (12) represents relative quality, the vertical axis represents image quality given by MOSs, and the red curve represents the fitting function (in this case, $\beta_1 = -0.1$, $\beta_2 = 1.0$, and $\beta_3 = 2.25$) using nonlinear least squares regression.

TABLE I
DESCRIPTIONS OF THE SUBJECTIVE DATABASES

Database	Description
LIVE [29]	29 reference images, each image has 5 distortion types (JPEG, JP2K, white noise (WN), Gaussian blur (Gblur) and fast fading (FF) channel distortions) and 5/6 distortion levels per type.
CSIQ [30]	30 reference images and their distorted ones of 6 distortion types (JPEG, JP2K, WN, Gblur, Gaussian pink noise (PN) and Global contrast (GC)) at 5 distortion levels.
TID2013 [32]	25 reference images with 24 distortion types at 5 distortion levels. It was updated from TID2008 [31] by adding 8 more distortion types and one more distortion level.
LIVEMD [33]	Two hybrid distortions: #1: "blurred and compression with JPEG" and #2: "blurred and white noise". For each of them: 15 reference images, each one is distorted by the first distortion at 3 distortion levels, and then distorted by the second one at 3 levels.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Databases and Evaluation Protocols

The performance of an IQA metric is evaluated by measuring the correlation between the model-predicted scores and the subjective scores provided by the subjective databases. The subjective databases: LIVE [29], CSIQ [30], TID2013 [32], and LIVE multiple distortion (LIVEMD) [33] are used in our experiments. Brief introductions of these databases are given in Table I and the distortion types are listed in Table II for simplicity of the following statement.

Considering the nonlinearity of the subjective scores introduced during the subjective tests, it is customary to perform

TABLE II
DISTORTION TYPES AND THEIR INDEXES OF TID2013

#1	Additive Gaussian noise
#2	Additive noise in color components is more intensive than additive noise in the luminance component
#3	Spatially correlated noise
#4	Masked noise
#5	High frequency noise
#6	Impulse noise
#7	Quantization noise
#8	Gaussian blur
#9	Image denoising
#10	JPEG compression
#11	JPEG2000 compression
#12	JPEG transmission errors
#13	JPEG2000 transmission errors
#14	Non eccentricity pattern noise
#15	Local block-wise distortions of different intensity
#16	Mean shift (intensity shift)
#17	Contrast change
#18	Change of color saturation
#19	Multiplicative Gaussian noise
#20	Comfort noise
#21	Lossy compression of noisy images
#22	Image color quantization with dither
#23	Chromatic aberrations
#24	Sparse sampling and reconstruction

a nonlinear mapping [34] on the objective scores before the correlation measurement. After the nonlinear mapping, Pearson's linear correlation coefficient (PLCC) is computed between the objective scores $\{x_i\}$ and the subject scores $\{y_i\}$ $(i = 1, \ldots, n)$ as

$$\text{PLCC} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{14}$$

where $\bar{x}$ and $\bar{y}$ represent the mean values of $\{x_i\}$ and $\{y_i\}$, respectively. It is a measure of the linear correlation between two variable sets. It can indicate the prediction accuracy of an IQA model for predicting the subjective scores. For measuring the prediction monotonicity, i.e., the degree of predictions agrees with the relative magnitudes of MOSs, Spearman's rank order correlation coefficient (SROCC) is defined as the PLCC between the ranked variables. Assuming that n raw scores $\{x_i\}$ and $\{y_i\}$ are converted to ranks $\{X_i\}$ and $\{Y_i\}$, the SROCC is defined as

$$\text{SROCC} = 1 - \frac{6\sum_{i=1}^{n}(X_i - Y_i)^2}{n(n^2 - 1)} \tag{15}$$

where $(X_i - Y_i)$ is the difference between the ith image's ranks in the subjective and objective evaluations. Larger PLCC and SROCC values indicate better correlations between the predicted scores and the subjective scores and therefore better performances.

B. Performance on Individual Databases

In this section, the training and testing are performed on the same database. The images in the whole database are divided into training set and testing set. A training set consists of 80% of the reference images and their associated distorted versions, and a testing set consists of the remaining 20% of the reference images and their associated distorted versions. In order to ensure that MRLIQ is robust across content and is not biased by the specific train–test split, a random split of 80% training and 20% test is repeated 1000 times on LIVE, CSIQ, and TID2013 databases, respectively. The median SROCCs across these 1000 times training processes are tabulated in Tables III–V for each distortion type. We adopt two kinds of image features: NSS and GM + LOG M3 in MRLIQ to form MRLIQ-I and MRLIQ-II. We also compared MRLIQ with the state-of-the-art approaches, including four full reference IQA (FR-IQA) metrics: PSNR, SSIM, MSSIM, FSIM, and ESSIM and ten NR-IQA metrics: DIIVINE [8], CORNIA [11], GM + LOG (M3) [12], NSS-TS [13], DLIQA [14], CNN [15], SDIQA [16], BIQA [17], BLIINDS-II [34], and BRISQUE [35].

The SROCC statistics are reported in Tables III–V. As can be seen from Tables III–V, MRLIQ outperforms most of the NR IQAs. Remarkably, MRLIQ-I is better than DIIVINE on most of the distortion types. Since these two algorithms use the same image features, it can be concluded that the achievement of MRLIQ-I just comes from its learning process. In MRLIQ, although each model is trained on a specific distortion type, all the training samples are associated with each model [refer to (3)–(5)]. Thus, the common features shared by different distortions are exploited, and overfitting problem may be prevented. DIIVINE takes a single-task learning framework. It trains models separately for each distortion type. Each model uses only a subset of training samples with the identical distortion type. Relative to DIIVINE, MRLIQ-I benefits from the proposed multi-task learning framework, and thus has a better prediction ability than DIIVINE. We also employ GM + LOG M3 image features in MRLIQ-II. The statistics indicate that MRLIQ also performs well on these image features. It can be observed that MRLIQ-II performs better than GM + LOG M3 on JPEG, JP2K, Gblur, FF, and all distortions on the LIVE database, and it is also better than GM + LOG M3 on JPEG and WN distortions on the CSIQ database. From Table III, it can be observed that deep neural network [14]–[16] can boost the performances of the NR-IQAs, especially for the CNN model [15]. From Table V, MRLIQ (I and II) ranks within the top two on most of the distortions of TID2013. MRLIQ-I is better than DIIVINE on most of the distortions: #1–#4, #7–#13, #20–#22, and #24. MRLIQ-II is better than GM+LOG M3 on distortions #2–#4, #6, #8, #12, #13, and #23.

For ML-based NR metrics, the success of them depends on the effective image features and the ML training process. From Tables III and IV, all the NR metrics work well on the considered distortions, which indicates that the image features used are effective on these distortions. However, for TID2013 in Table V, these NR metrics completely fail on distortions #14–#18, which is because the involved image features cannot

TABLE III
PERFORMANCE COMPARISONS (SROCC) AMONG MRLIQ AND THE BENCHMARKS ON THE LIVE DATABASE

	PSNR	SSIM	MSSIM	FSIM	ESSIM	BLIINDS -II	DIIVINE	BRISQUE	CORNIA	BIQA	GM+LOG (M3)	NSS -TS	DLIQA -R	CNN	SDIQA	MRLIQ -I	MRLIQ -II
JP2K	0.8762	0.9405	0.9746	0.9717	0.9809	0.9386	0.9233	0.9229	0.9139	0.9440	0.9283	0.9310	0.9330	**0.9520**	0.9210	**0.9483**	0.9369
JPEG	0.9029	0.9462	0.9793	0.9834	0.9819	0.9426	0.9347	**0.9734**	0.9647	0.9450	0.9659	0.9150	0.9140	**0.9770**	0.9010	0.9535	0.9701
WN	0.9173	0.9824	0.9883	0.9652	0.9764	0.9635	**0.9867**	0.9851	0.9786	0.9730	**0.9853**	0.9710	0.9680	0.9780	0.9620	0.9756	0.9700
Blur	0.7801	0.9004	0.9645	0.9708	0.9917	0.8994	0.9370	0.9506	0.9511	0.9530	0.9395	0.9390	0.9470	**0.9620**	0.9300	0.9523	**0.9562**
FF	0.8795	0.9514	0.9488	0.9499	0.9476	0.8790	0.8916	0.9030	0.8768	0.9080	0.9008	**0.9350**	0.8570	0.9080	0.8870	**0.9212**	0.9071
All	0.8636	0.9129	0.9535	0.9652	0.9611	0.9124	0.9250	0.9395	0.9350	0.9380	0.9511	0.9300	0.9290	**0.9560**	0.9230	0.9401	**0.9528**

TABLE IV
PERFORMANCE COMPARISONS (SROCC) AMONG MRLIQ AND THE BENCHMARKS ON THE CSIQ DATABASE

	PSNR	SSIM	MSSIM	FSIM	ESSIM	BLIINDS -II	DIIVINE	BRISQUE	CORNIA	BIQA	GM+LOG (M3)	MRLIQ -I	MRLIQ -II
JP2K	0.9362	0.9606	0.9683	0.9685	0.9676	0.8870	0.8692	0.8934	0.8950	0.8580	**0.9172**	**0.9221**	0.9011
JPEG	0.8881	0.9546	0.9634	0.9654	0.9649	0.9115	0.8843	0.9253	0.8849	0.8420	**0.9328**	0.9012	**0.9404**
WN	0.9363	0.8974	0.9471	0.9262	0.9494	0.8863	0.8131	0.9310	0.7980	0.8060	**0.9406**	0.8873	**0.9413**
Blur	0.9291	0.9609	0.9711	0.9729	0.9629	**0.9152**	0.8756	0.9143	0.9006	0.8380	0.9070	**0.9102**	0.8966
All	0.9216	0.8716	0.9535	0.9616	0.9438	0.9003	0.8697	0.9085	0.8845	0.8430	**0.9243**	0.8925	**0.9219**

TABLE V
PERFORMANCE COMPARISONS (SROCC) AMONG MRLIQ AND THE BENCHMARKS ON THE TID2013 DATABASE

	PSNR	SSIM	MSSIM	FSIM	ESSIM	BLIINDS -II	DIIVINE	BRISQUE	BIQA	CORNIA	GM+LOG (M3)	MRLIQ -I	MRLIQ -II
#1	0.9291	0.8528	0.8646	0.9216	0.8972	0.7985	0.8885	0.9069	0.7010	0.6660	**0.9527**	0.8922	**0.9441**
#2	0.8981	0.7737	0.7730	0.8211	0.8177	0.7127	0.7646	0.8465	0.5640	0.5110	**0.8715**	0.7861	**0.8762**
#3	0.9200	0.8616	0.8544	0.9401	0.8751	0.7281	**0.8423**	0.8228	0.6920	0.7720	0.8138	**0.8463**	0.8333
#4	0.8323	0.8090	0.8073	0.7353	0.7937	0.5985	**0.7862**	0.6464	0.4090	0.4240	0.7369	**0.7753**	0.7719
#5	0.9140	0.8461	0.8604	0.8959	0.8971	0.8364	0.9399	**0.9401**	0.8380	0.7890	**0.9403**	0.9215	0.9207
#6	0.8968	0.7995	0.7629	0.8398	0.8076	0.8022	0.9027	**0.9165**	0.7760	0.6240	0.8715	**0.8926**	0.8806
#7	0.8808	0.8055	0.8706	0.8747	0.8713	0.8258	0.6442	0.8227	0.6650	0.7470	**0.8997**	0.7821	**0.8522**
#8	0.9149	0.9629	0.9673	0.9701	0.9550	0.9176	**0.9192**	0.8912	0.8980	0.8760	0.9068	**0.9206**	0.9119
#9	0.9480	0.9102	0.9268	0.9435	0.9301	0.7708	0.7658	0.6865	0.7760	0.8080	**0.8458**	**0.8123**	0.8024
#10	0.9189	0.9096	0.9265	0.9295	0.9328	0.8639	0.7335	0.8346	0.8320	0.8410	**0.9104**	0.8311	**0.8906**
#11	0.8840	0.9049	0.9504	0.9602	0.9577	0.8781	0.8638	0.8847	0.9010	0.8240	**0.9300**	**0.9042**	0.8865
#12	0.7685	0.8183	0.8475	0.8415	0.8466	0.4992	0.5931	0.4281	–	–	0.5085	**0.6258**	**0.6961**
#13	0.8883	0.8696	0.8889	0.9104	0.8912	0.6450	0.7738	**0.8150**	–	–	0.7573	0.8036	**0.8125**
#14	*0.6863*	*0.7594*	*0.7968*	*0.7980*	*0.7917*	*-0.0755*	*0.2285*	*0.1994*	–	–	*0.1029*	*0.3501*	*0.1231*
#15	*0.1552*	*0.6169*	*0.4801*	*0.6694*	*0.5507*	*0.0657*	*0.0281*	*0.1508*	–	–	*0.1775*	*0.0196*	*0.1566*
#16	*0.7671*	*0.7767*	*0.7906*	*0.7572*	*0.7524*	*0.0380*	*NaN*	*NaN*	–	–	*NaN*	*NaN*	*NaN*
#17	*0.4400*	*0.3481*	*0.4634*	*0.4681*	*0.4675*	*0.0490*	*0.3962*	*0.2793*	–	–	*0.5072*	*0.3600*	*0.4699*
#18	*0.0766*	*-0.4055*	*-0.4099*	*-0.3567*	*-0.3790*	*-0.0050*	*NaN*	*0.3508*	–	–	*NaN*	*NaN*	*NaN*
#19	0.8905	0.7748	0.7786	0.8610	0.8468	0.7338	**0.8088**	0.8174	0.6200	0.6170	**0.8185**	0.7842	0.8034
#20	0.8411	0.8188	0.8528	0.9090	0.9118	0.5063	0.5316	0.3504	–	–	**0.6185**	0.6011	**0.6024**
#21	0.9145	0.9114	0.9068	0.9613	0.9470	**0.8358**	0.6589	0.6977	0.6150	0.7750	**0.8470**	0.6842	0.8335
#22	0.9269	0.7892	0.8555	0.8928	0.8757	0.7985	0.8038	**0.8885**	0.8150	0.7660	**0.9131**	0.8212	0.8701
#23	0.8872	0.8880	0.8784	0.8854	0.8713	0.5738	0.8071	**0.8315**	0.7540	0.7190	0.7341	**0.7823**	0.7569
#24	0.9042	0.9031	0.9483	0.9662	0.9563	0.8599	0.8388	0.8581	**0.8920**	**0.8920**	0.8873	0.8788	0.8862

represent these distortions. Most of the existing image features were designed for measuring the degradation of image quality by noise, compression, and transmission error. For distortion #18 in Table V, even the full reference (FR) metrics fail since it concerns only the changes of color components. In Table V, the bad statistical results are italicized, and the reader can ignore them. NaN represents the bad results leading to overflow. The statistical results of CORNIA and BIQA are referred to in [11] and [17], where no statistics are provided for distortions 12–18 and 20.

In multi-task learning, each image contributes to all tasks with a certain weight described by the parameter $\boldsymbol{\alpha}$ in (3), where $\boldsymbol{\alpha}$ is responsible for clustering images with respective to their distortion types. For analyzing $\boldsymbol{\alpha}$, the LIVE image database is used. It has five distortion types with a distinct distortion characteristic, so five image clusters are considered, each of which consists of images of the same distortion type. Thus, $\boldsymbol{\alpha}$ is a 5D vector ($\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 4$, and $\alpha 5$), and α_j represents the weight/probability of an image as to the jth task. Fig. 3 shows the average $\boldsymbol{\alpha}$ for images of each cluster.

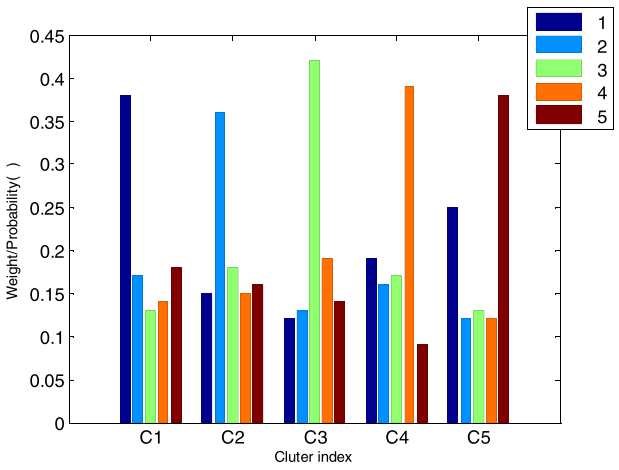

Fig. 3. Average $\boldsymbol{\alpha}$ for images of different clusters [in this experiment, five image clusters: C1, C2, C3, C4, and C5 are given by the distortion of JP2K, JPEG, WN, Gblur, and FF, respectively. Therefore, five bar sets are drawn in the figure. For each image, $\boldsymbol{\alpha}$ is a 5D vector ($\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 4$, and $\alpha 5$) and the jth element of α represents the weight/probability of this image as to the jth task. For each cluster (Cj, $j = 1, 2, \ldots, 5$), the average $\boldsymbol{\alpha}$ for images of that cluster is drawn in a bar set and the height of each bar represents the weight/probability of images of that cluster as to each task].

TABLE VI

PERFORMANCE COMPARISONS (SROCC): TRAINED ON THE LIVE DATABASE AND TESTED ON THE CSIQ DATABASE

	WN	JPEG	JP2k	Gblur	PN	GCD	AllSub
DIIVINE	0.8662	0.8689	0.8692	0.8667	0.3840	**0.4130**	0.8621
BLIINDS-II	0.8680	0.8930	0.8330	0.8420	0.4040	0.1710	0.8610
BRISQUE	0.8160	0.3740	0.6840	0.7290	0.1450	0.1460	0.4310
CORNIA	0.7490	0.8910	**0.9010**	0.8840	**0.4140**	0.3020	0.8810
GM+LOG M3	**0.9241**	**0.9092**	0.8764	0.8621	0.3477	0.0258	0.8948
BIQA	0.8240	0.8570	0.8850	0.8450	0.2380	0.1040	0.8480
MRLIQ-I	0.8865	**0.9018**	0.8910	**0.8846**	**0.4652**	**0.3583**	**0.9088**
MRLIQ-II	**0.9215**	0.8959	**0.9011**	**0.8965**	0.4002	0.3117	**0.9115**

In addition, the statistics of $\boldsymbol{\alpha}$ is also averaged over the 1000 times of training processes mentioned above. There are five clusters labeled C1, C2, C3, C4, and C5 in Fig. 3. For each cluster, the bar αj ($j = 1, 2, \ldots, 5$) represent the weight/probability of images of that cluster as to the jth task. From Fig. 3, the jth element of $\boldsymbol{\alpha}$ is biggest for the jth training task, which indicates that the images of cluster j contribute most to the jth task. This phenomenon coincides with the functionality of the parameter $\boldsymbol{\alpha}$ in (3). By this clustering parameter $\boldsymbol{\alpha}$, each image is related to all tasks with certain weights on the one hand. On the other hand, the common features among different distortions are explored for jointly training all IQA models in the proposed multi-task learning scheme.

C. Performance Across Databases

To verify that MRLIQ is independent of database, it is trained on the whole LIVE database and tested on the CSIQ and TID2013 databases. The SROCC statistics are reported in Tables VI and VII. For CSIQ database, JPEG, JP2k, WN, and Gblur are four common distortions in both LIVE and CSIQ, and the additive Gaussian pink noise (PN) and global contrast decrements (GCD) distortion types are outside of the LIVE database. Thus, the image quality prediction fails on these two distortions. AllSub represents the statistics across JPEG, JP2k, WN, and Gblur without PN and GCD. The competitions are performed among several closely related ML-based NR-IQA metrics. The bold fonts highlight the best metrics. It can be observed that MRLIQ (I and II) ranks within the top two on all the distortion types. MRLIQ-II performs best on JP2K and Gblur among all the competitors. MRLIQ-I is better than DIIVINE on WN, JPEG, JP2K, Gblur, and PN. MRLIQ-II is better than GM + LOG M3 on JP2K, Gblur, PN, and GCD and a little inferior to the latter on WN and JPEG. All in all, MRLIQ is competitive among all compared metrics. TID2013 has 24 distortion types, and the SROCC statistics on each distortion type are reported in Table VII for MRLIQ and the benchmarks. Since distortions #12–#18 and #20 are excluded from the training set, all of the NR-IQA models fail on these distortions. In addition, it also implies that these distortion types are much different from the distortions in the training set. Thus, the correlation coefficients of SROCC are very bad on these distortions. For the distortion, which is although excluded from the training set, but with the similar property as the one in the training set, the trained models still work well. For example, for #19: multiplicative Gaussian noise, we can still obtain good results, i.e., high SROCC by the considered models.

D. Performance on Hybrid Distortions

In practice, the images are usually distorted by more than one distortion. We test the MRLIQ and several NR-IQA metrics on the LIVEMD database [33]. It contains two hybrid distortions (please refer to Table I). The same train–test split as Section III-B is employed and the performance comparisons are tabulated in Table VIII.

Since LIVEMD is a new database, there is no comprehensive performance report on it. We implement DIIVINE, BLIINDS-II, BRISQUE, and GM + LOG M3 and report the SROCC statistics of them on LIVEMD in Table VIII. We use the absolute same SVR for training models in these algorithms. The standard LIBSVM package [36] is employed to implement the SVR. In addition, the genetic algorithm is used to get optimized parameters. The MATLAB code can be accessed via our homepage [37]. Since these algorithms use the same SVR training process, the only difference of them lies in the different image features. DIIVINE uses the statistics Wavelet pyramid decomposition on a whole image, BLIINDS-II uses an NSS model of blocked DCT coefficients, BRISQUE uses an NSS in the spatial domain, and GM + LOG M3 uses the statistical distributions of GM and LOG coefficients.

From Table VIII, the proposed model performs best among all the competitors. Among these five IQA models, DIIVINE and MRLIQ-I, GM + LOG M3, and MRLIQ-II use the same image features, respectively, so it can be concluded that the achievement is due to the proposed multi-task learning method.

E. Performance of Distortion Classification

To predict the image quality of an image with an unknown distortion type, a classifier is first employed to identify its

TABLE VII

PERFORMANCE COMPARISONS (SROCC): TRAINED ON THE LIVE DATABASE AND TESTED ON THE TID2013 DATABASE

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24
DIIVINE	0.855	0.712	0.463	0.675	0.878	0.806	0.165	0.834	0.723	0.629	0.853	*0.239*	*0.061*	*0.060*	*0.093*	*0.010*	*0.460*	*0.068*	0.787	*0.116*	0.633	0.436	0.661	0.833
BRISQUE	0.889	0.784	0.380	0.674	0.888	0.686	0.757	0.769	0.559	0.842	0.855	*0.045*	*0.390*	*0.183*	*0.215*	*0.097*	*0.189*	*0.183*	0.781	*0.194*	0.738	**0.787**	0.693	0.892
BLIINDS-II	0.788	0.562	0.369	0.668	0.802	0.632	0.518	0.828	0.703	0.645	0.721	*0.112*	*0.305*	*0.119*	*0.251*	*0.083*	*0.052*	*0.296*	0.731	*0.095*	0.574	0.616	0.670	0.829
CORNIA	0.761	0.679	**0.615**	0.686	0.828	0.741	0.399	**0.915**	**0.834**	**0.885**	0.899	*0.622*	*0.655*	*0.371*	*0.168*	*0.123*	*0.173*	*0.071*	0.659	*0.483*	**0.874**	0.530	**0.749**	0.714
BIQA	0.764	0.727	0.505	0.664	0.736	0.732	**0.768**	0.818	0.742	0.873	**0.908**	*0.105*	*0.408*	*0.082*	*0.358*	*0.208*	*0.099*	*0.332*	0.657	*0.096*	0.636	**0.840**	0.636	**0.895**
GM+LOG (M3)	0.876	0.796	**0.616**	0.684	**0.915**	0.801	**0.820**	0.836	0.615	0.848	0.801	*0.091*	*0.504*	*0.017*	*0.371*	*0.112*	*0.193*	*0.291*	0.784	*0.127*	0.676	0.479	0.557	0.546
MRLIQ-I	**0.898**	**0.882**	0.391	**0.725**	**0.901**	**0.826**	0.537	**0.886**	**0.852**	0.875	**0.919**	*0.336*	*0.258*	*0.188*	*0.089*	*0.158*	*0.267*	*0.153*	**0.820**	*0.139*	0.782	0.687	**0.765**	**0.916**
MRLIQ-II	**0.877**	**0.826**	0.561	**0.695**	0.890	**0.816**	0.717	0.869	0.822	**0.895**	0.892	*0.286*	*0.458*	*0.108*	*0.259*	*0.195*	*0.231*	*0.218*	**0.790**	*0.151*	**0.792**	0.627	0.685	0.809

TABLE VIII

PERFORMANCE COMPARISONS (SROCC) ON THE LIVEMD DATABASE

	BLIINDS-II	DIIVINE	BRISQUE	GM+LOG (M3)	MRLIQ-I	MRLIQ-II
#1	0.9025	0.9227	0.9109	0.9055	**0.9504**	0.9233
#2	0.9015	0.8663	0. 8946	0.8376	0.9158	**0.9201**

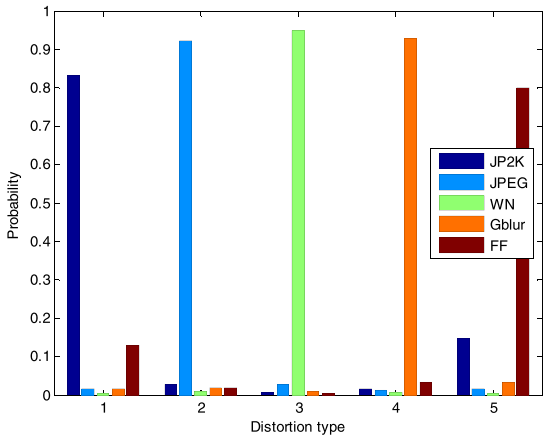

Fig. 4. Probability of each distortion is mistakenly judged to other distortions (1, 2, 3, 4, and 5 represent the distortion type of JP2K, JPEG, WN, Gblur, and FF, respectively).

distortion type. Referring to [8], an SVM classifier with five distortion types is trained on the LIVE database in this work. Here, the NSS image feature as in [8] is employed for training the classifier.

To test the classification accuracy of the trained classifier, the same train–test split as Section III-B is employed for training and testing. The classification accuracy is reported in Table IX for individual distortion and across distortion. It can be observed that this classifier can identify white noise very well. For other distortion, the accuracy is acceptable. In addition, the soft classification strategy is employed, which computes the probability of each distortion type of an image. And then, a quality score is computed from each model specific for each distortion type. The average of these scores is the final quality score of the image. Fig. 4 illustrates that the probabilities of each distortion type is judged to other distortion types. For example, distortion 1 is mistakenly judged to distortions 2, 3, 4, and 5 with the probabilities of about 0.02, 0.01, 0.02, and 0.13, respectively. Table IX gives the accuracy of distortion classification using the trained SVM classifier on the LIVE image database, where the accuracy

TABLE IX

MEDIAN CLASSIFICATION ACCURACY OF CLASSIFIER ACROSS 1000 TRAIN–TEST TRIALS ON THE LIVE IMAGE DATABASE

	JP2K	JPEG	WN	Gblur	FF	All
MRLIQ	83.08	91.85	95.36	90.93	80.27	84.53
NSS-TS [13]	94.44	94.37	96.67	93.33	93.33	93.39

of more than 80% can be achieved for all distortion types. Referring to [13], the distortion classification accuracy can be significantly improved using a multiclass MKL. More than 90% accuracy was reported in [13] for both individual distortion and all distortions, as shown in Table IX. In the proposed MRLIQ, soft classification [8] is employed so that an image is associated with multiple models by a set of weighting factors instead of the single one as in [13], which is particularly beneficial for hybrid distortion.

IV. CONCLUSION

In this paper, we have investigated pairwise rank learning and multi-task learning for IQA. First, since it is less confusing to give preference for the two compared images than to rate them for the subjects, PC is more reliable than image quality rating for small image quality difference in subjective evaluation. Based on this point, the proposed MRLIQ is superior to the traditional image-quality-rating-based approaches. Second, due to the significant difference in the statistics of different distorted images, each model for each distortion type is more accurate than the methods without distortion type discrimination. In addition, multiple models are trained simultaneously on all the training samples with the consideration of sharing a common feature and highlighting a specific feature of each task in MRLIQ, and this enhances the generalization ability of trained models. All in all, MRLIQ takes a fundamental and interesting departure from the traditional learning framework optimized on numerical rating systems and trained models separately, so a meaningful exploration on the new perspectives of IQA research is presented.

REFERENCES

[1] 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. 11, pp. 1918–1927, Nov. 2005.

[2] L. Liang, S. Wang, J. Chen, S. Ma, D. Zhao, and W. Gao, "No-reference perceptual image quality metric using gradient profiles for JPEG2000," *Signal Process., Image Commun.*, vol. 25, no. 7, pp. 502–516, Aug. 2010.

[3] T. Brandão and M. P. Queluz, "No-reference image quality assessment based on DCT domain statistics," *Signal Process.*, vol. 88, no. 4, pp. 822–833, Apr. 2008.
[4] 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.
[5] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 995–1002.
[6] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[7] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Colorado Springs, CO, USA, Jun. 2011, pp. 305–312.
[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] B. Gu, V. S. Sheng, K. Y. Tay, W. Romano, and S. Li, "Incremental support vector learning for ordinal regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 7, pp. 1403–1416, Jul. 2015.
[10] B. Gu, V. S. Sheng, Z. Wang, D. Ho, S. Osman, and S. Li, "Incremental learning for ν-support vector regression," *Neural Netw.*, vol. 67, pp. 140–150, Jul. 2015.
[11] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE CVPR*, Jun. 2012, pp. 1098–1105.
[12] 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.
[13] 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.
[14] 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.
[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 Recognit.*, Jun. 2014, pp. 1733–1740.
[16] W. L. Hou and X. B. Gao, "Saliency-guided deep framework for image quality assessment," *IEEE Multimedia Mag.*, vol. 22, no. 2, pp. 46–55, Apr./Jun. 2015.
[17] F. Gao, D. Tao, X. Gao, and X. Li, "Learning to rank for blind image quality assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 10, pp. 2275–2290, Oct. 2015.
[18] Z. Chen, T. Jiang, and Y. Tian, "Quality assessment for comparing image enhancement algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 3003–3010.
[19] H. Li, "Learning to rank for information retrieval and natural language processing," *Synth. Lectures Human Lang. Technol.*, vol. 4, no. 1, pp. 1–113, 2011.
[20] T.-Y. Liu, "Learning to rank for information retrieval," *Found. Trends Inf. Retr.*, vol. 3, no. 3, pp. 225–331, 2009.
[21] L. Xu, W. Lin, J. Li, X. Wang, Y. Yan, and Y. Fang, "Rank learning on training set selection and image quality assessment," in *Proc. IEEE Conf. ICME*, Chengdu, China, Jul. 2014, pp. 1–6.
[22] L. Xu *et al.*, "Multi-task rank learning for image quality assessment," in *Proc. IEEE Conf. ICASSP*, Brisbane, QLD, Australia, Apr. 2015, pp. 1339–1343.
[23] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Reading, MA, USA: Addison-Wesley, May 1999.
[24] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–446, Oct. 2002.
[25] J. Li, Y. Tian, T. Huang, and W. Gao, "Cost-sensitive rank learning from positive and unlabeled data for visual saliency estimation," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 591–594, Jun. 2010.
[26] J. Li, Y. Tian, T. Huang, and W. Gao, "Multi-task rank learning for visual saliency estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 5, pp. 623–636, May 2011.
[27] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, 1967, pp. 281–297.
[28] Y. Zheng, J. Byeungwoo, D. Xu, Q. M. J. Wu, and H. Zhang, "Image segmentation by generalized hierarchical fuzzy C-means algorithm," *J. Intell. Fuzzy Syst.*, vol. 28, no. 2, pp. 961–973, 2015.
[29] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "*Live Image Quality Assessment Database Release 2*," accessed on Sep. 1, 2011. [Online]. Available: http://live.ece.utexas.edu/research/quality
[30] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, p. 011006, Jan. 2010. [Online]. Available: http://vision.okstate.edu/?loc=csiq
[31] N. Ponomarenko *et al.*, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv. Modern Radioelectron.*, vol. 10, no. 4, pp. 30–45, 2009. [Online]. Available: http://www.ponomarenko.info/tid2008.htm
[32] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process., Image Commun.*, vol. 30, pp. 57–77, Jan. 2015. [Online]. Available: http://ponomarenko.info/tid2013.htm
[33] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Proc. Conf. Rec. 46th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2012, pp. 1693–1697. [Online]. Available: http://live.ece.utexas.edu/research/quality/live_multidistortedimage.html
[34] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[35] 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.
[36] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011, Art. no. 27.
[37] "Long Xu's homepage—PhD in computer science," Assessed on Jun. 30, 2009. [Online]. Available: http://sites.google.com/site/songjun0629/Home

Long Xu (M'12) received the M.S. degree in applied mathematics from Xidian University, Xi'an, China, in 2002, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China.

He held a post-doctoral position with the Department of Computer Science, City University of Hong Kong, Hong Kong, and with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, from 2009 to 2012. From 2013 to 2014, he held a post-doctoral position with the School of Computer Engineering, Nanyang Technological University, Singapore. He is currently with Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences. His research interests include image/video processing, wavelet, machine learning, and computer vision.

Dr. Xu was selected into the 100-Talents Plan, Chinese Academy of Sciences, in 2014.

Jia Li (M'12–SM'15) received the B.E. degree from Tsinghua University, Beijing, China, in 2005 and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2011.

He is currently an Associate Professor with State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing. He has authored over 30 peer-reviewed papers, including some high-quality ones in IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, *International Journal of Computer Vision*, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, ICCV, and CVPR. His research interests include computer vision and image/video processing.

Weisi Lin (M'92–SM'98–F'16) received the Ph.D. degree from King's College London, London, U.K.

He served as the Laboratory Head of Visual Processing with the Institute for Infocomm Research, Singapore. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, perceptual signal modeling, video compression, and multimedia communication, in which he has authored over 130 journal papers and over 200 conference papers, filed seven patents, and has authored two books.

Dr. Lin is a fellow of The Institution of Engineering and Technology. He was a Technical Program Chair of the IEEE International Conference on Multimedia and Expo in 2013, the Pacific-Rim Conference on Multimedia in 2012, and the International Workshop on Quality of Multimedia Experience in 2014. He was the Chair of the IEEE MMTC Special Interest Group on QoE from 2012 to 2014. He is an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and *Journal of Visual Communication and Image Representation*, and was an Associate Editor of IEEE TRANSACTIONS ON MULTIMEDIA. He has also served as a Guest Editor of eight special issues in international journals. He has been an Invited/Panelist/Keynote/Tutorial Speaker in more than ten international conferences, and was a Distinguished Lecturer of the Asia-Pacific Signal and Information Processing Association from 2012 to 2013.

Yongbing Zhang (M'13) received the B.A. degree in English and the M.S. and Ph.D. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 2004, 2006, and 2010, respectively.

He joined the Graduate School at Shenzhen, Tsinghua University, Shenzhen, China, in 2010, where he is currently an Associate Professor. His research interests include video processing, image and video coding, video streaming, and transmission.

Dr. Zhang was a recipient of the Best Student Paper Award at the IEEE International Conference on Visual Communication and Image Processing in 2015.

Lin Ma (M'13) received the B.E. and M.E. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 2006 and 2008, respectively, and the Ph.D. degree from the Department of Electronic Engineering, The Chinese University of Hong Kong (CUHK), Hong Kong, in 2013.

He was a Research Intern with Microsoft Research Asia, Beijing, China, from 2007 to 2008. He was a Research Assistant with the Department of Electronic Engineering, CUHK, from 2008 to 2009. He was a Visiting Student with the School of Computer Engineering, Nanyang Technological University, Singapore, in 2011. He is currently a Researcher with the Huawei Noah's Ark Laboratory, Hong Kong. His research interests include deep learning and multimodal learning, specifically for image and language, image/video processing, and quality assessment.

Dr. Ma received the Microsoft Research Asia Fellowship in 2011. He received the best paper award at the Pacific-Rim Conference on Multimedia in 2008. He was a finalist to the HKIS Young Scientist Award in engineering science in 2012.

Yuming Fang (M'12) received the B.E. degree from Sichuan University, Chengdu, China; the M.S. degree from Beijing University of Technology, Beijing, China; and the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore.

He was a (Visiting) Post-Doctoral Research Fellow with the IRCCyN Laboratory, Polytech Nantes/Université de Nantes, Nantes, France; University of Waterloo, Waterloo, ON, Canada; and Nanyang Technological University, Singapore. He is currently an Associate Professor with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, and 3D image/video processing.

Dr. Fang was a Secretary of the Ninth Joint Conference on Harmonious Human Machine Environment in 2013. He was also a Special Session Organizer in the Visual Communications and Image Processing Conference in 2013 and the International Workshop on Quality of Multimedia Experience in 2014.

Yihua Yan received the B.E. and M.E. degrees from Northwestern Polytechnical University, Xi'an, China, in 1982 and 1985, respectively, and the Ph.D. degree from Dalian University of Technology, Dalian, China, in 1990.

He was a Foreign Research Fellow with the National Astronomical Observatory of Japan, Tokyo, Japan, from 1995 to 1996, and an Alexander von Humboldt Fellow with the Astronomical Institute, Würzburg University, Würzburg, Germany, from 1996 to 1997. He is currently a Professor and the Chief Scientist of Solar Radio Group, and the Director of Key Laboratory of Solar Activity and the Solar Physics Division with National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China. He is the President of the International Astronomical Union Division E: Sun and Heliosphere from 2015 to 2018. His research interests include solar magnetic fields, solar radio astronomy, space solar physics, and radio astronomical methods.