QL-IQA: Learning Distance Distribution from Quality Levels for Blind Image Quality Assessment

Rui Gao[a,1], Ziqing Huang[a,2] and Shiguang Liu[a,b,*,3]

[a]School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
[b]Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin 300350, China

ARTICLE INFO

Keywords:
No-reference image quality assessment
Siamese network
Clustering
Convolutional neural network

ABSTRACT

Recently, blind image quality assessment (BIQA) has been intensively studied with deep learning, in which the limited quality-annotated datasets restrict its further development. Although patch-based methods have been leveraged to acquire more training data, they usually assign the image quality score to all patches in an image. Consequently, much noise would be introduced. To avoid using image patches, we hence propose a method learning distance distribution from quality levels to expand the amount of training data, which is a Siamese network-based no-reference image quality assessment (NR-IQA). Specifically, we firstly use the K-means clustering algorithm to classify the quality scores of distorted images into five levels. Subsequently, the Siamese network is adopted to learn the distance distribution from these quality levels. We mainly decrease distances between the quality scores of images from the same quality level and increase distances between the quality scores of images from different quality levels, in which the distorted images can be compared across different distortion types. Finally, we introduce a fusion layer to average the quality scores learnt by the two branches of trained Siamese network, and take fine-tuning on this basis to learn image quality score from one image. To demonstrate the effectiveness of the proposed approach, we evaluate it on benchmark databases (LIVE, TID2013 and LIVE MD). The result shows superior performance to some widely used NR-IQA methods and full-reference IQA (FR-IQA) methods. Cross-database evaluation verifies high generalization ability and high effectiveness of our model.

1. Introduction

Image quality assessment (IQA) is a process of predicting the quality of distorted images with respect to human perception automatically, which does not resort to the human subjective judgement. Therefore, it is convenient to estimate the quality of a large number of images simultaneously. IQA plays an important role in terms of computer vision and has been widely applied. For example, images always go through various processing stages (e.g., digitization, compression and transmission) before arriving at the terminal receiver, which will introduce distortions into images and lead to poor perceptual experience. Hence, predicting image quality helps to guide the optimization and evaluation of processing systems [24, 2]. What's more, with the development and maturation of stereoscopic technology, stereoscopic image and video quality assessment have been quite active[56, 57], which also need IQA technology to guide them to improve.

Generally, IQA algorithms are classified into three categories based on the available information from original images: full-reference IQA (FR-IQA), reduced-reference IQA (RR-IQA), and no-reference IQA (NR-IQA). Owing to the information of reference images, FR-IQA and RR-IQA have achieved a high correlation with human perception [20, 59, 12, 47, 37, 50, 14, 8, 6]. However, in most scenarios, it's hard to get original images or partial information from them. Accordingly, it's important to explore NR-IQA, which only takes distorted images as input without any information of reference images. In the field of NR-IQA, most traditional methods focus on using hand-crafted statistic features [31, 13, 27]. Typically, Natural Scene Statistics (NSS) based features are extracted in image transformation domains [35, 42, 29] or spatial domains [33], which limits the diversity and flexibility for modeling the multiple complex distorted types. With the rise of deep learning, hand-crafted features have been substituted by feature learning increasingly. In computer vision, Convolutional Neural Networks (CNNs) allow end-to-end learning of features and regression based on the raw image pixels without any hand-engineering. They have made great achievements in image classification and object detection. Promising results of CNNs encourage researchers to investigate their formidable application to the NR-IQA task. They have shown improved advantages compared to these hand-crafted methods [35, 42, 29, 33]. Unfortunately, as data-driven model, CNNs require a huge number of annotated training data for avoiding overfitting. Nevertheless, extremely limited quality-annotated samples in public databases greatly limit the power of CNNs, which has been a performance bottleneck for most NR-IQA methods based on CNNs. To alleviate the absence of large datasets for IQA, previous works [18, 19, 5, 4] usually utilize data augmentation strategy by randomly sampling small patches from the quality annotated images or directly dividing images into small patches. They have achieved some results in the field of NR-IQA. However, it is worth noting that small patches from an image are assigned to the quality labels of the corresponding annotated image, which are imprecise in most practical cases because in some distortion types, the

*Corresponding author
15735170462@163.com (R. Gao); skyhuangzq@163.com (Z. Huang); lsg@tju.edu.cn (S. Liu)
ORCID(s):

quality of patches in one image varies much and the patches' quality score can't be simply assigned as the image quality score. In other words, the quality of the whole image is insufficient to represent the local quality, which will lead to a high amount of label noise in the augmented datasets. To avoid introducing much noise, there is a method expected to estimate quality score for each patch based on the real quality score of one image [16]. However, it only improves the performance of existing learning-based IQA methods and its effect is less obvious.

In this work, we adopt the method learning distance distribution from quality levels to avoid sampling small patches from original images. The main idea is that according to the FIVE-points MOS scale, which is that one prefers to conduct evaluations qualitatively rather than numerically, thus it's easy to classify the distorted features into five grades, corresponding to five explicitly mental concepts, i.e., excellent, good, fair, poor, and bad. Hence, it makes sense to employ clustering algorithm and classify these quality scores of distorted images into five levels (excellent, good, fair, poor, bad). Following this, considering the advantages that Siamese network can evaluate the similarity of any pair of images, we randomly select image pairs as the input of Siamese network to learn the distance distribution from five quality levels. Through pairing, the amount of training datasets will be greatly increased and be effective for training CNNs. We call this learning from quality levels approach QL-IQA. After training, we introduce a quality score fusion stage in order to meld the quality scores learnt from two branches of the Siamese network, and then we transfer knowledge learnt from quality levels to the new Siamese network structure and fine tune it on IQA data to improve the accuracy of IQA. Based on the existing distorted images, we improve the performance to some extent by learning the distance distribution between the quality scores of them without introducing or generating any other additional distorted data.

The second contribution is that the distorted images can be compared across distortion types based on our approach. That is, the two distorted images in an image pair can be corrupted with different types of distortion or even hybrid distortion. As pointed out by Ghadiyaram and Bovik [9], "images captured using typical real-world mobile camera devices are usually afflicted by complex mixtures of multiple distortions, which are not necessarily well modeled by the synthetic distortions found in existing database." In consideration of the quality scores that are our main attention in the training process, we simply care about the quality scores of distorted images. The limitation from distortion types is broken in our method. Hence, we can simultaneously process images of any type of complex or simple distortion, which is more practical. That is, even though there are other distortion types or more complicated hybrid distortion in the future, our method is still valid and reliable. Experimental results also demonstrate that our approach is more applicable to all kinds of distorted images, compared to other methods with similar accuracy in singly synthetic distortions.

Another contribution of our method is that we qualitatively verify the operability of the pseudo Siamese network through experiment. The input pair belongs to the same model, rather than cross-model. In view of the two input objects which are of the same format or appearance and the two branch architectures which are uniform, the two sets of parameters (e.g., weights arguments, bias arguments) learnt by the two branches of the Siamese network won't be much different. Thus, we call this network weakly-pseudo Siamese network. We conduct experiments with popular datasets to evaluate the efficacy of our method. The results show that our method is effective and feasible. It can improve the accuracy of IQA to some extent than other existing FR-IQA and NR-IQA methods.

2. Previous work

As mentioned above, the IQA approaches can be divided into three groups, depending on the additional information needed: FR-IQA [7, 38, 51, 15, 1], NR-IQA [36, 33, 34, 32, 25, 26]. Since NR-IQA is more generally applicable and practical, below we will briefly introduce several classic FR-IQA approaches and keep focusing on reviewing some NR-IQA approaches.

2.1. Full-reference image quality assessment

FR-IQA performs a direct comparison between the distorted images and original images defined in a proper image space. The two simplest and most widely used FR-IQA metrics are the mean squared error (MSE) computed by averaging the squared intensity differences of distorted and reference image pixels, and the related quantity of peak signal-to-noise ratio (PSNR). They are appealing because they are easy to be calculated with clear physical meanings, and they are mathematically convenient in the context of optimization. However, they are not very well matched to the perceived visual quality [11, 48]. Later, Wang et al. [51] proposed a prominent approach based on the structural similarity (SSIM), which considers the sensitivity of human visual system (HVS) to structural information by pooling luminance similarity, contrast similarity and structural similarity. Inspired by this method [51], Zhang et al. [61] proposed a feature similarity index for image quality assessment (FSIM), which employs two features to compute the local similarity map, the phase congruency and the gradient magnitude. In the quality score pooling stage of FSIM, phase congruency map is utilized again as a weighting function since it can roughly reflect how perceptually important a local patch is to the HVS. In this work, we propose an approach called QL-IQA to improve the accuracy of IQA, which shows better performance without any reference information than these FR-IQA methods just mentioned.

2.2. No-reference image quality assessment

NR-IQA assumes that image quality can be determined without a direct comparison between the original images and the distorted images, thus they can be applied whenever the original image is unavailable. According to the effect of the amount of quality-annotated images on the algorithm

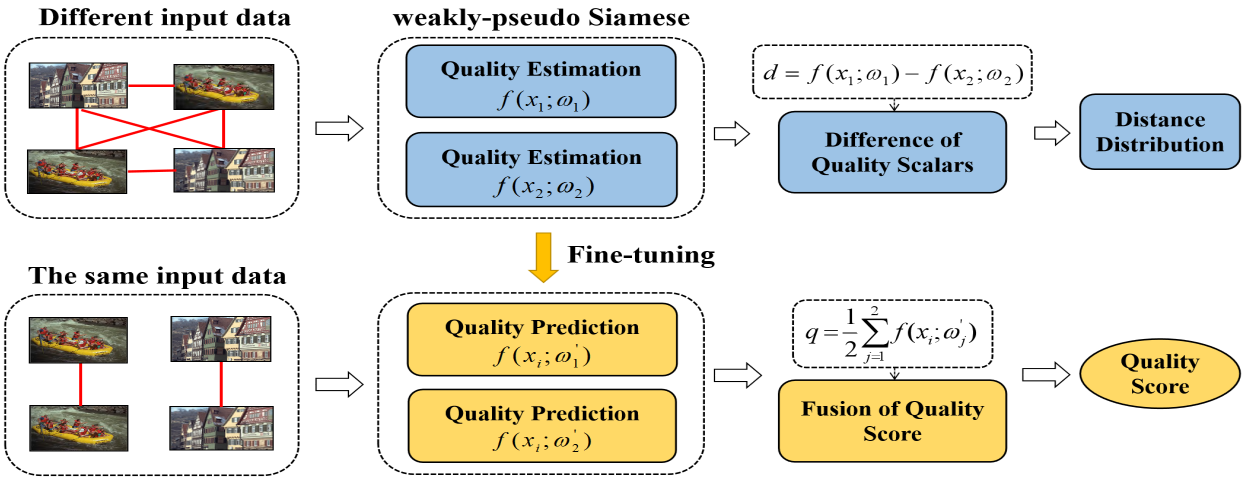

Figure 1: An illustration of our proposed QL-IQA framework. Features are extracted from the distorted image pairs by two network branches without sharing weights, which is named as weakly-pseudo Siamese network. "Different input data" indicates that arbitrary two input objects have different quality scores, and the quality score of the former is higher than the latter. These image pairs can be generated according to the quality levels obtained from the given quality scores. Then the trained network model is added with a fusion layer to average quality scores and the original network parameters are transferred to the new model for fine-tuning. Besides, "The same input data" indicates that the two input objects have the same appearance and quality score.

performance, we classified the NR-IQA methods into two groups: non-data-driven methods and data-driven methods. Most traditional NR-IQA methods belong to the non-data-driven methods. Most of them mainly estimate the image quality by measuring deviations from Natural Scene Statistic (NSS) models that characterize the distributions of certain filter responses. For example, Moorthy and Bovik [35] designed a two-step framework named BIQI for constructing blind image quality indices, which is used to extract the features in a sub-band based on Wavelets. Besieds, they further proposed the DIIVINE framework [36] and achieved good results, Saad et al. [42] leveraged a generalized Gaussian density function to model block DCT coefficients of images. This method concludes that distortion will affect the contrast, clarity and anisotropy of the image. There is a great coincidence that human eyes are good at extracting structural information from natural scene images and are sensitive to contrast. However, these above methods are time-consuming. To obtain significant speed-ups, Mittal et al. [33] managed to extract NSS features in the spatial domain. They proposed BRISQUE approach, which quantifies the distortion of images by using the local normalized brightness factor and gets the efficiency. Since then, Mittal et al. [34] came up with NIQE algorithm, which extracts features based on a multivariate Gaussian model and relates them to perceived quality in an unsupervised manner. It's not difficult to find out that these methods tend to tackle IQA in two steps: extracting handcrafted features and building regression. Both of them are completely independent with each other. If there is a method which allows joint end-to-end learning of features and regression based on the raw input data without any hand-engineering or other types of prior domain knowledge about the human visual system or image statistics, it would be better.

Therefore, with the significant progress of Deep Neural Networks (DNNs) on computer vision, several methods have utilized deep learning for NR-IQA [3, 18, 19, 4, 30]. They are end-to-end methods that integrate feature learning and regression into the same network framework and rely on large labeled datasets, which indicates they are data-driven methods just mentioned in the previous paragraph. Yet the existing IQA training datasets are not enough to support deep learning. Accordingly, Kang et al. [18] considered 32×32 patches rather than the entire images, thereby greatly augmenting the number of training datasets. The authors of [5, 19, 4] followed the same pipeline. Kang et al. [19] designed a multi-task CNN to learn the distortion type and image quality simultaneously. These above methods might introduce much noise in patches' labels because in some distortion types the quality of patches in one image varies much and the image quality score can't be simply assigned to all patches in this image. Binaco et al. [3] used a pre-trained network to relieve the lack of training data. Liu et al. [30] generated a huge of ranked images to train a Siamese network to learn the rankings for NR-IQA according to the distortion types and grades in particular datasets. The generation of our approach is largely enlightened by Liu et al. In contrast, we don't have to synthesize masses of ranked images. We directly train on existing IQA database by pairing images to learn the distance distribution between image quality scores so as to simplify method and achieve good results. Besides, compared to [30], the two images in any image pair are not restricted to be corrupted with the same type of distortion. We can train any image pairs, which are afflicted by two arbitrary distortion types, respectively.

Table 1
The Definitions of Main Notations

Notation	Definition
L	Loss function
d	Distance between the prediction quality score of the two branches
ω	Weights parameters learnt by the weakly-pseudo Siamese network
x	Input image of the network
y	Practical quality score of the image x
$\hat{y}$	Prediction quality score of the image x

3. Methodology

In this section, we will introduce NR-IQA in detail. An overview of our training model is illustrated in Figure 1, which is separated into two parts from top to bottom. The top part of the training model is employed to learn the distance distribution from the same or different quality levels. The bottom part of the training model is utilized to estimate quality scores. The overall pipeline of our approach is depicted into three stages. Firstly, we represent the generation process of the pairs of distorted images, and then describe how we use a weakly-pseudo Siamese network to learn distance distribution between the quality scores of the distorted images. Finally, we show the fine-tuning stage on IQA data to estimate quality by adding a fusion layer.

3.1. Generating image pairs

3.1.1. Image preprocessing

In order to speed up the convergence of the network to the optimal solution, we apply local contrast normalization to each channel of a color image by referring to [18]. Then we assign the three normalized color channels to the original corresponding channels respectively and get the new image with three channels. The specific computing details on each channel are as follows:

$$\begin{aligned} \hat{I}(i,j) &= \frac{I(i,j)-\mu(i,j)}{\sigma(i,j)+C} \\ \mu(i,j) &= \sum_{p=-3}^{3}\sum_{q=3}^{-3} I(i+p,j+q) \\ \sigma(i,j) &= \sqrt{\sum_{p=-3}^{3}\sum_{q=3}^{-3}(I(i+p,j+q)-\mu(i,j))^2} \end{aligned} \tag{1}$$

where $I(i,j)$ is the pixel value at location (i,j). C is a positive constant that prevents dividing by zero. p and q are the normalized window sizes, and $\mu(i,j)$, $\sigma(i,j)$ represent the local mean and variance at location (i,j). The $\hat{I}(i,j)$ is the final pixel value after local normalization at location (i,j). We take local contrast normalization on all of training and testing image data used later.

3.1.2. Image Pairing

Facing the scarcity of IQA data, patch-based approaches introduce much noise. As a contrast, the Siamese network taking image pairs as input enlarges the amount of the training datasets from another angle, which motivates us to propose a Siamese network-based approach to learn distance distribution from image quality levels. With this approach, we circumvent the patch techniques radically. Our approach depends on the existing IQA datasets, and it is easy to collect all of distorted images of them together and extract corresponding quality scores. We ignore specific distortion types and concentrate on the ground-truth IQA scores called Differential Mean Opinion Scores (DMOS) or Mean Opinion Scores (MOS) [45]. The DMOS value is taken to reflect the differences between the human subjective evaluation scores of undistorted images and distorted images, and the MOS value is leveraged to mirror human subjective evaluation scores of distorted images. Based on these quality scores, we exploit K-means clustering algorithm to complete the clustering. According to the FIVE-points MOS Scale, They are divided into five clusters, namely five levels, which are qualitatively expressed as "excellent", "good", "fair", "poor" and "bad". Afterwards, we trace back these quality scores to the corresponding images and match them according to the quality levels. As an example, given an image we can pair it with another image from the same level with it or we can choose one image from another level which is away from its level one or two levels to match it.

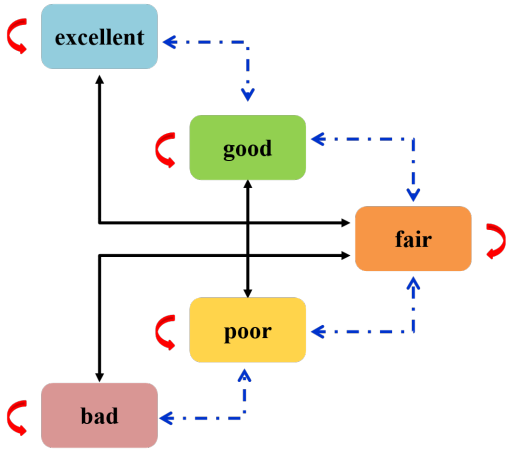

Figure 2: An illustration of different quality levels. After pairing, these image pairs will be utilized to train weakly-pseudo Siamese network.

Table 2
Variable Values of the Loss Function in Eq. (3)

level	1	2	3	4	5
range	$[r_{11}, r_{12}]$	$[r_{21}, r_{22}]$	$[r_{31}, r_{32}]$	$[r_{41}, r_{42}]$	$[r_{51}, r_{52}]$
g	$\frac{(r_{12}-r_{11})+(r_{22}-r_{21})+(r_{32}-r_{31})+(r_{42}-r_{41})+(r_{52}-r_{51})}{5}$				
σ	$\frac{(r_{31}-r_{12})+(r_{41}-r_{22})+(r_{51}-r_{32})}{3}$				
β	$\frac{(r_{32}-r_{11})+(r_{42}-r_{21})+(r_{52}-r_{31})}{3}$				

Figure 2 shows the pairing condition between different quality levels. In detail, the arcuate single arrows indicate that images from the same quality level will be paired. The two-way dashed arrows indicate that the images at one quality level interval from each other will be paired. The two-way solid arrows represent that the images at two quality levels interval from each other will be paired. In order to keep the quantity balance between each kind of labeled image pairs, we only select image pairs with zero, one or two quality levels interval, excluding image pairs with three and four quality levels interval, so as to achieve better training effects. Note that these pairs of images belong to the training data of the Siamese network. Next, labels are required for these image pairs. Specifically, the images within the same level are paired and labeled as '0'. Images with gap of one level are paired and labeled as '1'. Similarly, images with gap of two levels are paired and labeled as '2'. What's more, we overlook distortion types and distortion degrees, and simply pay attention to the values of their quality scores, thus the one with a higher quality score is in the front of the one with a lower score value in an image pair. At this moment, the labeled image pairs are generated through the above process.

3.2. Training weakly-pseudo Siamese network

Taking image pairs generated in the previous subsection, we feed them into the Siamese network to learn the distance distribution from the same or different quality levels by combining with our designed loss module. After that, we get a weakly-pseudo Siamese network. The top row of Figure 1 shows the framework of this training stage. It's noteworthy that we train our network on randomly sampling sub-images from the distorted images instead of scaling to avoid introducing noise caused by interpolation or filtering. Different from small patches of images, the size of the sub-images is at least 1/3 of the original distorted images, which is large enough to capture context information and distortion to represent the information and quality of the entire images.

The objective of training is to learn the distance distribution from five quality levels. Based on this, we minimize the following loss function:

$$L(x_1, x_2; \omega_1, \omega_2) = l_0 + l_1 + l_2 + l_3 \tag{2}$$

where l_0, l_1, l_2 represent the distance loss between the prediction quality of image pairs that are labeled by '0', '1', and '2', respectively. l_3 denotes the range control of prediction score. By minimizing the loss L, we shorten the distance between the quality scores of images from the same quality level and increase the distance between the quality score of images from different quality levels. The specific form of these four parts l_0, l_1, l_2, l_3 are as follows.

$$\begin{aligned} l_0 &= 0^{|0-y|}[max(0-d,0) - max(d-g,0)]^2 \\ l_1 &= 0^{|1-y|}[max(0-d,0) - max(d-2g,0)]^2 \\ l_2 &= 0^{|2-y|}[max(\sigma-d,0) - max(d-\beta,0)]^2 \\ l_3 &= [max(-f(x_1;\omega_1),0) - max(f(x_1;\omega_1)-r,0)]^2 \end{aligned} \tag{3}$$

$$d = f(x_1;\omega_1) - f(x_2;\omega_2) \tag{4}$$

where y represents the input label for image pair, valued '0', '1', '2', respectively. d is the distance between the prediction quality for x_1 and x_2, which is listed in Eq. (4) where $f(x_1;\omega_1)$ and $f(x_2;\omega_2)$ denote the prediction quality of the two network branches in Siamese network and ω_1 and ω_2 represent weights parameters learnt by the two branches. Moreover, g, σ and β are all determined by analysing the distribution rule of quality scores from five levels in IQA datasets. Their computing details are summarized in Table 1, where "*level*" represents the tag number of each of the five levels and "*range*" expresses the range of quality scores within each level. Due to different ranges, sizes, and distribution laws of quality scores in different IQA databases, the values of the three parameters are certainly flexible instead of fixed, which depends on the selected training data.

Specifically, the destination of l_0 is to control the prediction quality distance in the range of 0 to g for the image pair from the same quality level. The form of l_1 and l_2 are consistent with l_0. The difference is l_1 focus on keeping the quality distance within 0 and $2g$ for the image pair differing by one quality levels, and l_2 controls the quality distance in the range of σ to β for the image pair differing by two levels.

The destination of l_3 is to control the prediction quality score within the range of r that represents the maximum value in the range of human IQA measurements referred as DMOS or MOS, which certainly relies on the selected IQA database. For instance, the LIVE database measure image quality with DMOS whose values ranges from 0 to 100. Accordingly, the value of r in Eq. (3) is 100.

Following this, the gradients of the loss $L(x_1, x_2; \omega_1, \omega_2)$ with respect to all model parameters are computed by back-propagation and these parameter values are updated with the Stochastic Gradient Decent (SGD) method. Below is the

gradient of the loss:

$$\bigtriangledown_\omega l = \begin{cases} \bigtriangledown_\omega l_0 = \begin{cases} 2d\bigtriangledown_\omega d & d \le 0 \\ 0 & 0 < d \le g \\ 2(d-g)\bigtriangledown_\omega d & d > g \end{cases} & y = 0 \\ \bigtriangledown_\omega l_1 = \begin{cases} 2d\bigtriangledown_\omega d & d \le 0 \\ 0 & 0 < d \le 2g \\ 2(d-2g)\bigtriangledown_\omega d & d > 2g \end{cases} & y = 1 \\ \bigtriangledown_\omega l_2 = \begin{cases} 2(\sigma-d)\bigtriangledown_\omega d & d \le \sigma \\ 0 & \sigma < d \le \beta \\ 2(d-\beta)\bigtriangledown_\omega d & d > \beta \end{cases} & y = 2 \end{cases} \tag{5}$$

Note that when the value of d is within a certain range, the gradient decreases by 0, indicating that the weakly-pseudo Siamese network has learnt correct distance distribution for the quality scores from quality levels in IQA databases. Otherwise, we need to continue to train and make the loss to fall more as far as possible.

Certainly, we cannot ignore the range of two output values. In the process of training Siamese network, we are required to control it within the range of MOS or DMOS values determined by IQA data sets. It is expressed as l_3 in the loss function appearing in Eq. (3). Besides, we are also obliged to take that into account when we calculate the gradient of loss L. Eq. (6) is the final gradient of the loss, where it can be found that when the value of $f(x_1;\omega)$ is between 0 and r, the gradient of loss L is the smallest. Correspondingly, the probability of L meeting the smallest value is the largest, which is in accordance with our expectations. Otherwise, we need to continue to train until all the conditions are met as much as possible.

$$\bigtriangledown_\omega L = \begin{cases} \bigtriangledown_\omega l + 2f(x_1;\omega_1) & f(x_1,\omega_1) \le 0 \\ \bigtriangledown_\omega l & 0 < f(x_1,\omega_1) \le r \\ \bigtriangledown_\omega l + 2[f(x_1;\omega_1) - r] & f(x_1;\omega_1) > r \end{cases} \tag{6}$$

3.3. Fine-tuning

Considering that we are training weakly-pseudo Siamese network, the parameters of the two branches of this network are different. Whereas the quality score of image is a single scalar, we transform the output of our weakly-pseudo Siamese network to a single scalar that indicates the quality score of one image. Hence we add a fusion layer as the last layer of the Siamese network, whose principal function is to average the outputs of the two network branches, so as to indicate image quality and meet the demand for data format eventually. Through fusion, we conserve the diversity of features extracted from the same image by using two branches with distinct parameters, and make the prediction score to approximate the practical image quality score preferably. $\hat{y}_i$ represents the prediction quality score of the image x_i after fusion in Eq. (7).

$$\hat{y}_i = \frac{f(x_i;\omega_1{}') + f(x_i;\omega_2{}')}{2} \tag{7}$$

We begin to fine-tune the weakly-pseudo Siamese network with slight changes in structure for domain adaptation to the exact image quality score. The bottom row of Figure 1 shows the training framework in fine-tuning stage. Given N images in mini-batch with MOS or DMOS values, we use l_1-norm as loss function, which is expressed in Eq. (8).

$$\begin{aligned} L(y_i,\hat{y}_i) &= \frac{1}{N}\sum_{i=1}^{N} \parallel \hat{y}_i - y_i \parallel_{l_1} \\ \omega' &= min_\omega L(y_i,\hat{y}_i) \end{aligned} \tag{8}$$

where y_i is the practical quality score of image x_i. We train the weakly-pseudo Siamese network with single scalar output by minimizing the loss module, $L(y_i,\hat{y}_i)$. Accordingly, the update of weights ω also lies on the gradient of L by using SGD mentioned in the previous subsection.

4. Experimental Results and Analysis

We design various experiments to evaluate the performance of our approach. We try some minor modifications on the existing foundation for further analyzing and discussing the details of our method. Moreover, we take our QL-IQA compared with some other representative IQA approaches to measure the availability and effectiveness.

4.1. Experimental Preparation

4.1.1. Datasets.

We employ three widely used singly synthetic databases, i.e., LIVE [44] and TID2013 [40] along with a multiply distorted synthetic dataset LIVE MD [17]. The LIVE database [44] consists of 779 quality annotated images that are generated from 29 source reference images by distorting them with 5 different types of distortions - JPEG compression (JPEG), JPEG2000 compression (JP2K), Gaussian blur (GB), white Gaussian noise (WN) and fast fading error (FF) at different distortion grades. DMOS are provided for each image, which is in the range of [0,100]. Higher DMOS indicates lower quality. The TID2013 database [40] is an extension of the earlier published TID2008 database [41], which is derived from 25 source reference images by distorting them with 24 different types of distortions at 5 distortion grades each. The total number of distorted images is 3000. Each image is associated with a MOS in the range of [0,9]. Contrary to DMOS, a higher MOS value indicates higher quality. LIVE MD [40] contains 450 images generated from 15 pristine images with corruption under two multiple distortion scenarios, i.e., blur followed by JPEG compression and blur followed by white Gaussian noise. DMOS values lie in the range of [0,100], and a higher value indicates lower quality.

4.1.2. Network Architecture

We take the VGG-16 network [46] as pre-trained model, and the two branches of the weakly-pseudo Siamese network

both adopt VGG-16. In addition, we need to change the original final layer with 1000 output nodes into fully connected layer with only one output node. Since our ultimate aim is to predict the quality score which is a single scalar for each image, we need to change the number of nodes in the output layer of every branch appropriately in order to adapt to the requirement of the output form. Certainly, we sample 224×224 pixel images as the input images, depending on the input requirement of VGG-16 network.

4.1.3. Evaluation Metrics

Two measures are used to evaluate the performance of model: the Linear Correlation Coefficient (LCC) and Spearman Rank Order Correlation Coefficient (SROCC). LCC measures the linear dependence between the ground truth and the prediction quality scores and SROCC measures how well the prediction quality scores can be described as a monotonic function of the ground truth. Given N distorted images, the ground truth of the i-th image is denoted by y_i, and the prediction score is $\hat{y}_i$. The LCC and SROCC are computed as:

$$LCC = \frac{\sum_{i=1}^{N}(y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}})}{\sqrt{\sum_{i}^{N}(y_i - \overline{y})^2 \sum_{i}^{N}(\hat{y}_i - \overline{\hat{y}})^2}} \tag{9}$$

$$SROCC = 1 - \frac{6\sum_{i=1}^{N}(v_i - p_i)^2}{N(N^2 - 1)} \tag{10}$$

where $\overline{y}$ and $\overline{\hat{y}}$ are the means of the ground truth and prediction quality scores, respectively. In Eq. (10), v_i is the rank of the ground-truth score y_i in the ground-truth IQA scores, and p_i is the rank of $\hat{y}_i$ in the prediction scores for all of the images.

As suggested in [49], a nonlinear regression formula involving five parameters is employed in this work to map the objective quality predictions to the subjective scores before computing LCC:

$$\tilde{s} = \beta_1(\frac{1}{2} - \frac{1}{\exp(\beta_2(\hat{s} - \beta_3))} + \beta_4\hat{s} + \beta_5) \tag{11}$$

where $\hat{s}$ belongs to the predicted quality scores computed by the IQA model, and $\tilde{s}$ is the corresponding mapped scores. $\{\beta_i; i = 1, 2, 3, 4, 5\}$ are the fitting parameters, which are determined by minimizing the sum of squared errors between the subjective scores (MOS or DMOS values) and the mapped objective scores $\tilde{s}$.

4.2. Analysis and Study on the Details

For sake of thorough analysis and exploration on the efficacy of our weakly-pseudo Siamese network model, we perform a quantity of experiments to test our method on several details. We evaluate the assessment result on different quality levels. Moreover, there are a number of variable factors in the process of training and fine-tuning weakly-pseudo Siamese network, which can affect the ultimate performance. In this section, we examine how these variable

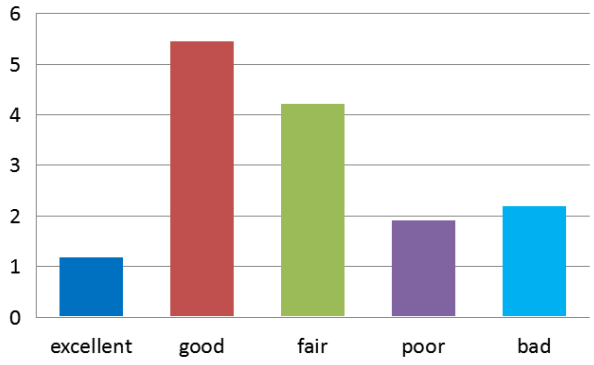

Figure 3: Loss values on different quality levels.

factors affect the performance on the LIVE database. We follow the protocol used in HOSA [52], RankIQA [30] and Hallucinated-IQA [28], where the entire distorted images are divided into 80% training images and 20% testing images according to the reference images. Thus, there is not any overlap image content between training and testing sets. It should be noticed that 80% training data are utilized in both the training weakly-pseudo Siamese network stage and fine-tuning stage. The same is valid for the 20% testing data. All the experiments are performed ten times with train-test splitting operation randomly, and the average SROCC and LCC values are reported as final results.

4.2.1. Evaluation on Different Quality Levels

During the training process, we neglect image distortion types as well as distortion degrees under the same distortion type, and only focus on the specific quality scores. Based on this, we utilize weakly-pseudo Siamese network to learn distance distribution of distorted images quality from quality levels. Namely, we shorten the distance between the quality scores of images from the same quality level and increase the distance between the quality scores of images from different quality levels. Therefore, it's necessary to evaluate the model prediction capacity in each quality level. Figure 3 illustrates the distribution of loss values, which is calculated over different quality levels containing the same number of distorted images. Note that the ordinate indicates the loss value, which is the average of all the differences between the predicted and real quality scores under the same quality level, while different colors represent five quality levels, respectively. The height of the cylinder and prediction ability of the model are in inverse proportion. Consequently, it's easy to find out that our model has a higher capacity to predict quality scores of distorted images under "excellent", "poor", and "bad" quality levels than the remainder two levels. This observation gives information that in terms of extremely distorted images, it's more easy to predict their quality scores. On the contrary, for images with medium distortion level, our weakly-pseudo Siamese network has weaker ability to predict their quality. That is consistent with human subjective evaluation of the images, because people show better perception in ultra-distorted im-

Table 3
LCC and SROCC on the Networks of Different Depths

	Kang	AlexNet	VGG-16
LCC	0.928	0.948	**0.965**
SROCC	0.942	0.951	**0.965**

ages than moderately distorted images. Furthermore, this phenomenon also proves from the side that it's feasible and effective to classify image quality scores into five levels via clustering. In other word, even if our method does not focus on specific distortion type and distortion degree, the experimental results indicate that learning distance distribution from quality levels is in accordance with human subjective judgement, which is significant and important.

4.2.2. The Depth of Network

Table 2 illustrates that deeper network is required by a large amount of training data generated through pairing. We take three different depths of networks (Kang et al. [18], AlexNet [23], VGG-16 [46]) respectively as the baseline for our approach. We refer to the network architecture in [18], which is a 32×32 - $26 \times 26 \times 50$-2×50 - 800 - 800 - 1 structure. In order to adapt the network to a larger input image size, three convolutional layers are added in front of it and the original input layer with one channel is changed to fifty channels. We call this new network as Kang, which is the first author's name of [18]. Besides, we select AlexNet that is famous for ImageNet contest as one of the comparative objects. The contrast result explains the significance of the deeper network. Obviously, under the same initialization from ImageNet, it achieves about 2% LCC and 4% LCC higher than Kang and AlexNet when using VGG-16 as baseline for our method. Similar conclusions are obtained for SROCC. This also indicates that it's useful and effective by pairing images for generating a great amount of training data and learning distance distribution from quality levels, so that a deeper network can be employed to learn more image presentations without overfitting.

4.2.3. Baseline performance

The objective of this experiment is to validate the necessity and effectiveness of learning distance distribution from quality levels. We compare with two methods, one of which is to input IQA data directly into VGG-16 network initialized from ImageNet in training phase, and then obtain the predicted quality scores in testing phase. We call this method baseline. Another method is to input image pairs with different distorted images firstly, and then train Siamese network based on VGG-16 initialized from ImageNet and learn distance distribution between the distorted image quality scores. After that, we fine tune this weakly-pseudo Siamese network to obtain the mapping from images to their prediction scores. We call this method QL. Table 4 exhibits the final result, where it's apparent that the QL method has a higher prediction precision for image quality scores. This method improves the baseline by over 1% for SROCC and LCC, which implies that our QL method provides a higher correlation with the IQA scores and more accurate IQA scores. Consequently, it is not difficult to draw the following conclusions. First, owing to the scarcity of training data, it's hard to obtain reliable results by training a deep network directly on existing IQA data. Second, it's obvious to discover that by learning distance distribution from quality levels, we can get more useful image representations on large image pair datasets. Therefore, we can fine-tune the deep network with these representations to get better results.

4.2.4. Weight Allocation

Table 5 shows the advantage of non-sharing weights during training Siamese network, where non-sharing weights achieves 0.7% LCC and 0.5% SROCC improvements than sharing weights. Although the increase is slight, it's still valuable and significant. Therefore, we qualitatively verify the two branches of the same structure in Siamese network can also learn different parameters, even though the two inputs are of the same format or the same appearance.

4.3. Comparisons with SOTA

4.3.1. Evaluation on LIVE

Table 3 presents the performance evaluation on the entire LIVE database, in which the first-ranked performance figures of each measurement criterion (i.e., LCC or SROCC) in each row are indicated in bold. The proposed method QL-IQA model is compared with multiple IQA models, including PSNR, SSIM [51], FSIM [61], GMSD [55], DOG-SSIM [39], DIVINE [36], BLIINDS-II [42], BRISQUE [33], CORNIA [58], NIQE [34], FRIQUEE [10], BIECON [21], CNN [18], SOM [62], where the last nine IQA models are specifically designed for the NR-IQA (separated by a vertical line in the Table 3), while the rest are all for the FR-IQA. The row indicated with ALL implies we put all five distortion types of images together to train and test the weakly-pseudo Siamese network. Among the IQA methods based on deep learning which are compared in the experiments, BIECON applies FR-IQA methods to generated proxy quality scores and CNN adopts a shallow network to estimate the quality score for each patch and averages the patch scores to obtain a quality estimation for the image. Our method severally achieves about 0.3% and 1.2% improvements than BIECON and CNN reported on ALL distortions for LCC and similar results are obtained for SROCC. Compared with SOM, our method gets slight improvement. Nevertheless, it is more advantageous on almost all of individual distortions than SOM method. Additionally, The experimental results indicate that our approach works well on each of the five distortions, especially on WN, GB and FF. There is a mild difference between the proposed method and the other methods on JPEG distortion type, but it's still approximate and optimistic to predict JPEG distortion images with our method. For the overall evaluation, QL-IQA method outperforms most of the existing work including the FR-IQA methods and NR-IQA methods. These performances imply that our approach using clustering on image quality scores is meaningful while

Table 4
LCC and SROCC on LIVE Database

	Distortions	PSNR	SSIM [51]	FSIM [61]	GMSD [55]	DOG [39]	DIVINE [36]	BLIINDS-II [42]	BRISQUE [33]	CORNIA [58]	NIQE [34]	FRIQUEE [10]	BIECON [21]	CNN [18]	SOM [62]	QL-IQA
LCC	JP2K	0.873	0.921	0.910	-	-	0.922	0.935	0.923	0.951	0.937	-	**0.965**	0.953	0.952	0.950
	JPEG	0.876	0.955	0.985	-	-	0.921	0.968	0.973	0.965	0.956	-	**0.987**	0.981	0.961	0.972
	WN	0.926	0.982	0.976	-	-	0.988	0.980	0.985	0.987	0.977	-	0.970	0.984	**0.991**	0.983
	GB	0.779	0.893	0.978	-	-	0.923	0.938	0.951	0.968	0.953	-	0.945	0.953	0.974	**0.994**
	FF	0.870	0.939	0.912	-	-	0.888	0.896	0.903	0.917	0.913	-	0.931	0.933	0.954	**0.963**
	ALL	0.856	0.906	0.960	0.960	0.963	0.917	0.930	0.942	0.935	0.915	0.930	0.962	0.953	0.962	**0.965**
SROCC	JP2K	0.870	0.939	0.970	0.971	-	0.913	0.929	0.914	0.943	0.917	-	0.952	0.952	0.947	**0.960**
	JPEG	0.885	0.946	**0.981**	0.978	-	0.910	0.942	0.965	0.955	0.938	-	0.974	0.977	0.952	0.965
	WN	0.942	0.964	0.967	0.974	-	0.984	0.969	0.979	0.976	0.966	-	0.980	0.978	0.984	**0.985**
	GB	0.763	0.907	0.972	0.957	-	0.921	0.923	0.951	0.969	0.934	-	0.956	0.962	0.976	**0.988**
	FF	0.874	0.941	0.949	0.942	-	0.863	0.889	0.887	0.906	0.859	-	0.923	0.908	0.937	**0.943**
	ALL	0.866	0.913	0.964	0.960	0.961	0.916	0.931	0.940	0.942	0.914	0.950	0.961	0.956	0.964	**0.965**

Table 5
Performance Evaluation Based on Baseline and QL on LIVE and TID2013 Databases

	LIVE		TID2013	
	LCC	SROCC	LCC	SROCC
baseline	0.949	0.952	0.688	0.636
QL	0.965	0.965	0.701	0.653

Table 6
Performance of Networks with Shared and Non-shared Weights on LIVE and TID2013 Databases

	LIVE		TID2013	
	LCC	SROCC	LCC	SROCC
shared weights	0.958	0.960	0.695	0.647
non-shared weights	0.965	0.965	0.701	0.653

ignoring specific distortion type and degree during training Siamese network.

4.3.2. Evaluation on TID2013

A comparison of performances for the TID2013 data is shown in Table 6. Our proposed method performs superior in terms of LCC and SROCC, which is greatly large than those of compared methods. QL-IQA acquires about 4.7% higher than the compared DIVINE method and it is the best for LCC among all other evaluated methods. As to BRISQUE method, which performs slightly better than other contrasted NR-IQA methods for SROCC, our method still obtains superior capacity. This result is consistent with our expectations, and also indicates that our weakly-pseudo Siamese network can be applied on arbitrary IQA database, not just on LIVE database.

4.3.3. Evaluation on Multiple Databases

In order to present the comprehensive performance comparisons over multiple databases, as suggest in [55, 53], two commonly-used average measurements are adopted to evaluate the average performance of different IQA models over the LIVE and TID2013 databases in this work. The average values are computed in two cases: Direct Average and Weighted Average. The Direct Average is about setting the weights of each database equal, while the weights in the Weighted Average are determined by the number of distorted images in each database.They can be defined as:

$$\bar{I} = \frac{\sum_{i=1}^{M} I_i \cdot \omega_i}{\sum_{i=1}^{M} \omega_i} \tag{12}$$

where M means the total number of database ($M = 2$ in this work), I_i indicates the value of the performance index (i.e., LCC and SROCC) on i-th database (i=1,2 correspond to LIVE and TID2013, respectively), and ω_i is the corresponding weight i-th database. For the Direct Average, ω_1=ω_2=1. For the Weighted Average, ω_i relies on the size of database and the number of distorted images in i-th database (i.e., 779 for LIVE and 3000 for TID2013) is assigned to it. Table 7 exhibits the average performance of

Table 7
LCC and SROCC on TID2013 Database

IQA methods	LCC	SROCC
DIVINE [36]	0.654	0.549
BLIINDS-II [42]	0.628	0.536
BRISQUE [33]	0.651	0.573
CORNIA [58]	0.613	0.549
NIQE [34]	0.426	0.317
QL-IQA	**0.701**	**0.653**

Table 8
LCC and SROCC on Multiple Databases

IQA methods	Direct Average		Weighted Average	
	LCC	SROCC	LCC	SROCC
DIVINE [36]	0.786	0.733	0.708	0.625
BLIINDS-II [42]	0.779	0.734	0.690	0.617
BRISQUE [33]	0.797	0.757	0.711	0.649
CORNIA [58]	0.774	0.746	0.679	0.630
NIQE [34]	0.671	0.616	0.527	0.440
QL-IQA	**0.833**	**0.809**	**0.755**	**0.717**

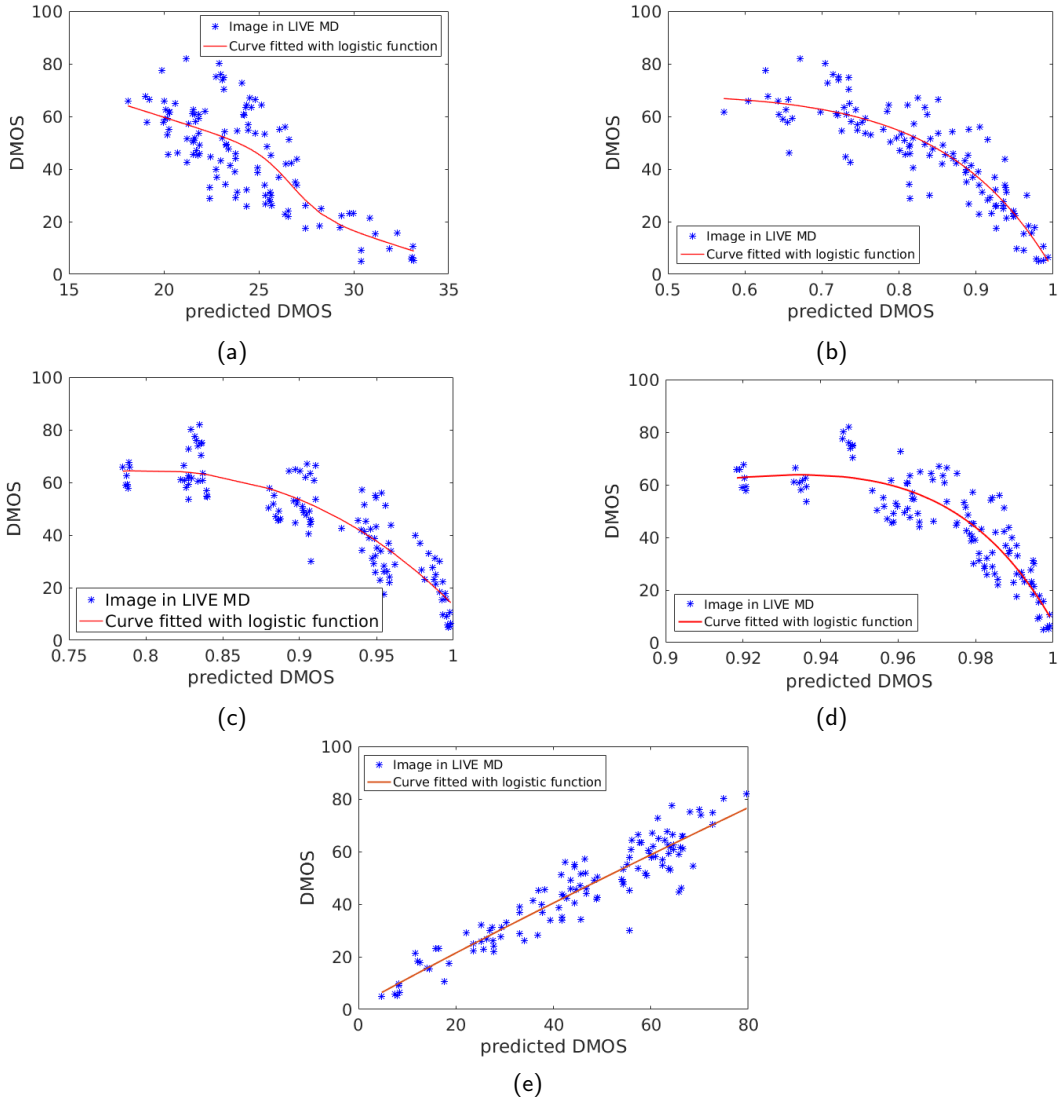

Figure 4: Scatter plots of actual DMOS with (a) PSNR, (b) SSIM, (c) FSIM, (d) VSI and (e) proposed method QL-IQA on the LIVE MS database.

different IQA methods on these two databases. It can be obviously observed that the proposed QL-IQA yields the highest LCC and SROCC in both Direct Average and Weighted Average performance comparison. Seen another way, there is approximate 8.18% and 10.83% improvements on average for LCC and SROCC respectively than other IQA models based on these two average measurements. It's evident that QL-IQA offers more significantly comprehensive performance over multiple databases than a single database, meaning that the proposed QL-IQA performs superiorly to other state-of-the-art models under comparison.

4.3.4. Evaluation on LIVE MD

In order to test the practicability on more complex distortions, we conduct experiments on LIVE MD dataset. The evaluation results are listed in Table 8. As shown in this table, our method delivers better performance against other methods. Although there is a slight difference between our approach and ResNet-ft for SROCC, contrary to LCC, the 0.5% difference proves the two are comparable. Besides, as for the singly synthetic distorted image datasets LIVE and TID2013, QL-IQA method outperforms other methods in SROCC value and LCC value comparisons. Certainly, it still keep superior evaluation capacity for multiply distorted im-

Table 9
LCC and SROCC on LIVE MD Database

IQA methods	LCC	SROCC
PSNR	0.777	0.689
SSIM [51]	0.886	0.867
FSIM [61]	0.883	0.864
VSI [60]	0.880	0.852
BRISQUE [33]	0.917	0.886
CORNIA [58]	**0.921**	0.899
M3 [54]	0.919	0.892
ResNet-ft [22]	0.920	**0.909**
QL-IQA	**0.925**	**0.904**

age, i.e., LIVE MD database. This suggests that QL-IQA generalizes well to different distortion scenarios. In view of the complexity of distortion types in real life, our method is more viable and practical.

Some scatter plots between ground truth DMOS and predicted scores are shown in Figure 4 to highlight the advantages of our proposed method more intuitively. We illustrate the DMOS (subjective scores) versus the objective scores computed by the conventional FR-IQA models (i.e., PSNR, SSIM, FSIM and VSI), which are classical and widely-used and our proposed NR-IQA model. Note that there is a blue line presented in each sub-plot, and it is obtained by exploiting a nonlinear curve fitting process according to Eq. (11). In a way, this line shows the "mean" value of the performance points. That is, for each DMOS value (along the vertical axis), hopefully all the IQA scores (along the horizontal axis) are as close to this DMOS value as possible. Equivalently, the closer the performance points gather around the blue line, more accurate the IQA model's prediction compared with the DMOS scores (i.e., ground truth). One can see that the proposed QL-IQA has a 'tighter' curve fitting when comparing with other IQA models, meaning that the proposed QL-IQA achieves higher linearity with quality scores than conventional methods and has better overall performance.

4.3.5. Evaluation across Different Databases

In order to test the robustness and generalization ability of our method, we carry out experiments on the LIVE and LIVE MD databases. Firstly, in consideration of the three parameters g, σ, β mentioned in Section 3.2, which are flexible rather than fixed, depending on the selected training data, robustness is expected to demonstrate that QL-IQA performs well even though the values of these three parameters fluctuate over a small range. Secondly, the generic QL-IQA model is expected to not only perform well on the training database, but also generalize well to other IQA databases. Hence, we train weakly-pseudo Siamese network on the whole LIVE database, conduct fine-tuning on the LIVE MD database, and proceed to take cross dataset test on it. Next, we exchange the role of the two databases and further conduct experiment again to justify the effectiveness of our proposed method. LCC and SROCC results are reported in Table 9. It's expected that when taking g, σ, β calculated from LIVE MD database as the parameter values in training weakly-pseudo Siamese network stage, the model generalizes well to LIVE database, especially in individual distortions and vice versa. Note that compared to training on LIVE database directly, there is little improvement for LCC in individual distortions, i.e., JP2K, WN, FF, which is unexpected. Therefore, when we feed image pairs into Siamese network to learn the distance distribution from quality levels, the relationship between different distortion degrees is also learnt simultaneously, which is less affected by parameters. This guarantees the robustness of our method. Definitely, in order to get a more accurate evaluation, the changes of these three parameter values need to be controlled within a reasonable range.

5. Conclusions and Future Work

In this paper, to avoid sampling small patches, we proposed a novel method which expanded the amount of datasets from a new perspective by learning the distance distribution from quality levels. Consequently, we can reduce noise caused by small patches' labels and improve the performance further. Besides, we broke through the restriction of distortion types and multifarious distorted images can be trained at the same time, which comes closer to the real-world use. Furthermore, we proposed a weakly-pseudo Siamese network, whose effectiveness was verified through comparing with sharing weights. Various experiments were conducted and the results displayed that our method improved the accuracy of IQA compared with other widely used IQA methods.

Considering the difference of statistical features between natural images and screen content images (SCIs), for the future work, we plan to extend this framework to evaluate SCIs [43], in which we intend to keep employing Siamese network and utilize one branch of it for extracting the text features.

References

[1] Alaei, A., Raveaux, R., Conte, D., 2017. Image quality assessment based on regions of interest. Signal, Image and Video Processing 11, 673–680.

[2] Bhattacharya, S., Sukthankar, R., Shah, M., 2010. A framework for photo-quality assessment and enhancement based on visual aesthetics, in: Proceedings of the 18th ACM International Conference on Multimedia, pp. 271–280.

[3] Bianco, S., Celona, L., Napoletano, P., Schettini, R., 2018. On the use of deep learning for blind image quality assessment. Signal, Image and Video Processing 12, 355–362.

[4] Bosse, S., Maniry, D., Müller, K.R., Wiegand, T., Samek, W., 2018. Deep neural networks for no-reference and full-reference image quality assessment. IEEE Transactions on Image Processing 27, 206–219.

[5] Bosse, S., Maniry, D., Wiegand, T., Samek, W., 2016. A deep neural network for image quality assessment, in: Proceedings of the IEEE International Conference on Image Processing, pp. 3773–3777.

[6] Decherchi, S., Gastaldo, P., Zunino, R., Cambria, E., Redi, J., 2013. Circular-elm for the reduced-reference assessment of perceived image quality. Neurocomputing 102, 78–89.

[7] Eckert, M.P., Bradley, A.P., 1998. Perceptual quality metrics applied to still image compression. Signal Processing 70, 177–200.

[8] Gao, F., Wang, Y., Li, P., Tan, M., Yu, J., Zhu, Y., 2017. Deepsim: Deep similarity for image quality assessment. Neurocomputing 257, 104–114.

[9] Ghadiyaram, D., Bovik, A.C., 2014. Crowdsourced study of subjective image quality, in: Proceedings of Asilomar Conference on Signals, Systems and Computers, pp. 84–88.

[10] Ghadiyaram, D., Bovik, A.C., 2015. Feature maps driven no-reference image quality prediction of authentically distorted images, in: Proceedings of the Human Vision and Electronic Imaging, p. 93940J.

[11] Girod, B., Watson, A.B., 1993. What's wrong with mean-squared error?, digital images and human vision. MIT Press. 207–220.

[12] Golestaneh, S., Karam, L.J., 2016. Reduced-reference quality assessment based on the entropy of dwt coefficients of locally weighted gradient magnitudes. IEEE Transactions on Image Processing 25, 5293–5303.

[13] Gu, K., Lin, W., Zhai, G., Yang, X., Zhang, W., Chen, C.W., 2016. No-reference quality metric of contrast-distorted images based on information maximization. IEEE transactions on cybernetics 47, 4559–4565.

[14] Gu, K., Zhai, G., Lin, W., Liu, M., 2015. The analysis of image

Table 10
LCC and SROCC in a Cross Database Setting

Training	LIVE MD						LIVE
FT+Testing	JP2k (LIVE)	JPEG (LIVE)	WN (LIVE)	GB (LIVE)	FF (LIVE)	ALL (LIVE)	ALL (LIVE MD)
LCC	0.952	0.948	0.989	0.970	0.964	0.963	0.918
SROCC	0.951	0.911	0.973	0.975	0.931	0.959	0.889

contrast: From quality assessment to automatic enhancement. IEEE transactions on cybernetics 46, 284–297.

[15] He, L., Gao, X., Lu, W., Li, X., Tao, D., 2011. Image quality assessment based on s-cielab model. Signal, Image and Video Processing 5, 283–290.

[16] Heng, W., Jiang, T., 2017. From image quality to patch quality: An image-patch model for no-reference image quality assessment, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1238–1242.

[17] Jayaraman, D., Mittal, A., Moorthy, A.K., Bovik, A.C., 2012. Objective quality assessment of multiply distorted images, in: Proceedings of Signals, Systems and Computers, pp. 1693–1697.

[18] Kang, L., Ye, P., Li, Y., Doermann, D., 2014. Convolutional neural networks for no-reference image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1733–1740.

[19] Kang, L., Ye, P., Li, Y., Doermann, D., 2015. Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks, in: Proceedings of the IEEE International Conference on Image Processing, pp. 2791–2795.

[20] Kim, J., Lee, S., 2017a. Deep learning of human visual sensitivity in image quality assessment framework, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1676–1684.

[21] Kim, J., Lee, S., 2017b. Fully deep blind image quality predictor. IEEE Journal of Selected Topics in Signal Processing 11, 206–220.

[22] Kim, J., Zeng, H., Ghadiyaram, D., Lee, S., Zhang, L., Bovik, A.C., 2017. Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment. IEEE Signal Processing Magazine 34, 130–141.

[23] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Proceedings of Advances in Neural Information Processing Systems, pp. 1097–1105.

[24] Li, C., Loui, A.C., Chen, T., 2010. Towards aesthetics: A photo quality assessment and photo selection system, in: Proceedings of the 18th ACM International Conference on Multimedia, ACM. pp. 827–830.

[25] Li, J., Yan, J., Deng, D., Shi, W., Deng, S., 2017. No-reference image quality assessment based on hybrid model. Signal, Image and Video Processing 11, 985–992.

[26] Li, J., Zou, L., Yan, J., Deng, D., Qu, T., Xie, G., 2016. No-reference image quality assessment using prewitt magnitude based on convolutional neural networks. Signal, Image and Video Processing 10, 609–616.

[27] Li, L., Lin, W., Wang, X., Yang, G., Bahrami, K., Kot, A.C., 2015. No-reference image blur assessment based on discrete orthogonal moments. IEEE transactions on cybernetics 46, 39–50.

[28] Lin, K.Y., Wang, G., 2018. Hallucinated-iqa: No-reference image quality assessment via adversarial learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 732–741.

[29] Liu, L., Dong, H., Huang, H., Bovik, A.C., 2014. No-reference image quality assessment in curvelet domain. Signal Processing: Image Communication 29, 494–505.

[30] Liu, X., van de Weijer, J., Bagdanov, A.D., 2017. Rankiqa: Learning from rankings for no-reference image quality assessment, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1040–1049.

[31] Lu, W., Zeng, K., Tao, D., Yuan, Y., Gao, X., 2010. No-reference image quality assessment in contourlet domain. Neurocomputing 73, 784–794.

[32] Mahmoudpour, S., Kim, M., 2016. No-reference image quality assessment in complex-shearlet domain. Signal, Image and Video Processing 10, 1465–1472.

[33] Mittal, A., Moorthy, A.K., Bovik, A.C., 2012. No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing 21, 4695–4708.

[34] Mittal, A., Soundararajan, R., Bovik, A.C., 2013. Making a "completely blind" image quality analyzer. IEEE Signal Processing Letters 20, 209–212.

[35] Moorthy, A.K., Bovik, A.C., 2010. A two-step framework for constructing blind image quality indices. IEEE Signal Processing Letters 17, 513–516.

[36] Moorthy, A.K., Bovik, A.C., 2011. Blind image quality assessment: From natural scene statistics to perceptual quality. IEEE Transactions on Image Processing 20, 3350–3364.

[37] Panetta, K., Gao, C., Agaian, S., Nercessian, S., 2016. A new reference-based edge map quality measure. IEEE Transactions on Systems, Man, and Cybernetics: Systems 46, 1505–1517.

[38] Pappas, T.N., Safranek, R.J., Chen, J., 2000. Perceptual criteria for image quality evaluation. Handbook of Image and Video Processing , 669–684.

[39] Pei, S.C., Chen, L.H., 2015. Image quality assessment using human visual dog model fused with random forest. IEEE Transactions on Image Processing 24, 3282–3292.

[40] Ponomarenko, N., Ieremeiev, O., Lukin, V., et al., 2013. Color image database tid2013: Peculiarities and preliminary results, in: Proceedings of European Workshop on Visual Information Processing, pp. 106–111.

[41] Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., Battisti, F., 2009. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. Advances of Modern Radioelectronics 10, 30–45.

[42] Saad, M.A., Bovik, A.C., Charrier, C., 2012. Blind image quality assessment: A natural scene statistics approach in the dct domain. IEEE Transactions on Image Processing 21, 3339–3352.

[43] Shao, F., Gao, Y., Li, F., Jiang, G., 2017. Toward a blind quality predictor for screen content images. IEEE Transactions on Systems, Man, and Cybernetics: Systems 48, 1521–1530.

[44] Sheikh, H., 2005. Live image quality assessment database release 2. http://live. ece. utexas. edu/research/quality .

[45] Sheikh, H.R., Sabir, M.F., Bovik, A.C., 2006. A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Transactions on Image Processing 15, 3440–3451.

[46] Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

[47] Tao, D., Li, X., Lu, W., Gao, X., 2009. Reduced-reference iqa in contourlet domain. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 39, 1623–1627.

[48] Teo, P.C., Heeger, D.J., 1994. Perceptual image distortion, in: Proceedings of Human Vision, Visual Processing, and Digital Display V, pp. 127–142.

[49] VQEG, 2000. Final report from the video quality experts group on the validation of objective models of video quality assessment. Online: http://www.vqeg.org .

[50] Wang, S., Deng, C., Lin, W., Huang, G.B., Zhao, B., 2016. Nmf-based

image quality assessment using extreme learning machine. IEEE transactions on cybernetics 47, 232–243.

[51] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13, 600–612.

[52] Xu, J., Ye, P., Li, Q., Du, H., Liu, Y., Doermann, D., 2016. Blind image quality assessment based on high order statistics aggregation. IEEE Transactions on Image Processing 25, 4444–4457.

[53] Xu, Y., Liu, D., Quan, Y., Le Callet, P., 2015. Fractal analysis for reduced reference image quality assessment. IEEE Transactions on Image Processing 24, 2098–2109.

[54] Xue, W., Mou, X., Zhang, L., Bovik, A.C., Feng, X., 2014a. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. IEEE Transactions on Image Processing 23, 4850–4862.

[55] Xue, W., Zhang, L., Mou, X., Bovik, A.C., 2014b. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE Transactions on Image Processing 23, 684–695.

[56] Yang, J., Sim, K., Gao, X., Lu, W., Meng, Q., Li, B., 2019. A blind stereoscopic image quality evaluator with segmented stacked autoencoders considering the whole visual perception route. IEEE Transactions on Image Processing 28, 1314–1328.

[57] Yang, J., Zhao, Y., Jiang, B., Lu, W., Gao, X., 2020. No-reference quality evaluation of stereoscopic video based on spatio-temporal texture. IEEE Transactions on Multimedia 22, 2635–2644.

[58] Ye, P., Kumar, J., Kang, L., Doermann, D., 2012. Unsupervised feature learning framework for no-reference image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1098–1105.

[59] Zhang, L., Li, H., 2012. Sr-sim: A fast and high performance iqa index based on spectral residual, in: Proceedings of the IEEE International Conference on Image Processing, pp. 1473–1476.

[60] Zhang, L., Shen, Y., Li, H., 2014. Vsi: A visual saliency-induced index for perceptual image quality assessment. IEEE Transactions on Image Processing 23, 4270–4281.

[61] Zhang, L., Zhang, L., Mou, X., Zhang, D., 2011. Fsim: A feature similarity index for image quality assessment. IEEE Transactions on Image Processing 20, 2378–2386.

[62] Zhang, P., Zhou, W., Wu, L., Li, H., 2015. Som: Semantic obviousness metric for image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2394–2402.