

# Learning with Noisy Low-Cost MOS for Image Quality Assessment via Dual-Bias Calibration

Lei Wang, Qingbo Wu, *Member, IEEE*, Desen Yuan, *Member, IEEE*, King Ngi Ngan, *Life Fellow, IEEE*, Hongliang Li, *Senior Member, IEEE*, Fanman Meng, *Member, IEEE* and Linfeng Xu, *Member, IEEE*

**Abstract**—Learning-based Image Quality Assessment (IQA) models have obtained impressive performance with the help of reliable subjective quality labels, where Mean Opinion Score (MOS) is the most popular choice. However, in view of the subjective bias of individual annotators, the Labor-Abundant MOS (LA-MOS) typically requires large collections of opinion scores from multiple annotators for each image, which significantly increases the learning cost. In this paper, we aim to learn robust IQA models from Low-Cost MOS (LC-MOS), which only requires very few opinion scores or even a single opinion score for each image. More specifically, we consider the LC-MOS as the noisy observation of LA-MOS and enforce the IQA model learned from LC-MOS to approach the unbiased estimation of LA-MOS. Thus, we represent the subjective bias between LC-MOS and LA-MOS, and the model bias between IQA predictions learned from LC-MOS and LA-MOS (i.e., dual-bias) as two latent variables with unknown parameters. By means of the expectation-maximization-based alternating optimization, we can jointly estimate the parameters of the dual-bias, which suppresses the misleading of LC-MOS via a gated dual-bias calibration (GDBC) module. To the best of our knowledge, this is the first exploration of robust IQA model learning from noisy low-cost labels. Theoretical analysis and extensive experiments on four popular IQA datasets show that the proposed method is robust toward different bias rates and annotation numbers and significantly outperforms the other Learning-based IQA models when only LC-MOS is available. Furthermore, we also achieve comparable performance with respect to the other models learned with LA-MOS.

**Index Terms**—image quality assessment, low-cost MOS, labor-abundant MOS, subjective bias, noisy label learning.

## I. INTRODUCTION

Image quality assessment (IQA) is an active research area in multimedia technology, which is critical for evaluating and developing various perceptual-friendly image/video applications [1], [2], [3], [4], [5], [6], [7]. With the rapid development of the deep neural network (DNN), Learning-based IQA models have gained increasing attention. Various advanced networks have been developed for IQA and achieved impressive performance recently. It is important to note that the success of Learning-based IQA models highly relies on reliable subjective quality labels, which determine the direction of optimizing these advanced networks. Unfortunately, due to the subjective bias of individual annotators, it is greatly challenging to collect

The authors are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China. (Corresponding author: Qingbo Wu) This work was supported in part by the National Natural Science Foundation of China (No. U23A20286), the Independent Research Project of Civil Aviation Flight Technology and Flight Safety Key Laboratory (FZ2022ZZ06), and the Natural Science Foundation of Sichuan Province (2023NSFSC1972).

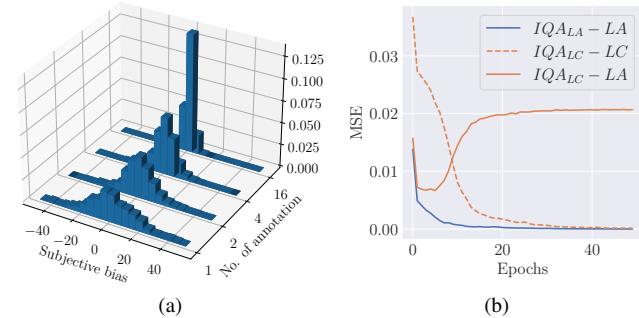


Fig. 1. (a) The distributions of the subjective bias under different numbers of human annotations on the VCL dataset; (b) The loss curves of training the DNN model with different labels on the KONIQ dataset.

reliable subjective quality labels, especially on a large-scale image database.

Subjective experiments are further complicated by many factors including viewing distance, display device, lighting condition, subjects' vision ability, and subjects' mood [8]. By averaging individual scores, MOS reduces the impact of subjective bias, which has long been regarded as the best method of image quality measurement [9]. However, the MOS method is labor-abundant (LA-MOS), which requires large collections of opinion scores from multiple annotators (usually more than 15) for each image.

Under a strictly standardized subjective experiment, the mean value of all annotators' opinion scores is finally used to represent their majority decision, which attempts to eliminate the subjective bias from individual annotators. As shown in Fig. 1 (a), subjective bias is presented as a Gaussian distribution, and the variance of the distribution diminishes as the annotation number increases. Meanwhile, different methods for pre/post-screening of the subject are also developed in BT500 [10], P910 [11], and P913 [12], to further refine the regular LA-MOS by rejecting or weakening the votes of a subject, whose behavior is biased and inconsistent with the others. Despite efficient subjective bias reduction, the aforementioned methods are only applicable to LA-MOS, whose annotators are large enough for each image. On the one hand, this label collection process is time-consuming and expensive, which significantly increases the learning cost. On the other hand, the model bias caused by subjective bias and the corresponding calibration strategy have rarely been investigated in previous works.

In this paper, we aim to learn robust IQA models from Low-Cost MOS (LC-MOS), which only requires very few opinion

scores or even a single opinion score for each image. LC-MOS exists in practical scenes and its corresponding noise form is subjective bias presented as a Gaussian distribution, which is different from the uniform distribution of the classification noisy label scene. Then we explore the negative impact of LC-MOS on learning-based IQA models. As shown in Fig. 1 (b), “ $IQA_{LA} - LA$ ” represents the MSE curve between the output of the model trained by LA-MOS and its fit target LA-MOS; “ $IQA_{LC} - LC$ ” represents the MSE curve between the output of the model trained with LC-MOS and its fit target LC-MOS; “ $IQA_{LC} - LA$ ” represents the MSE curve between the output of the model trained with LC-MOS and its potential fit target LA-MOS. Although “ $IQA_{LC} - LC$ ” (the orange dotted line) can reach 0 as well as “ $IQA_{LA} - LA$ ” (the blue line), “ $IQA_{LC} - LA$ ” deviates from 0 (the orange solid line), which means that the IQA model prediction trained with LC-MOS is far from its real potential fitting target LA-MOS.

To address this issue, we consider the LC-MOS as the noisy observation of LA-MOS and enforce the IQA model learned from LC-MOS to approach the unbiased estimation of LA-MOS. In this way, we represent the dual-bias, including the subjective bias between LC-MOS and LA-MOS and model bias between IQA models learned from LC-MOS and LA-MOS, as two latent variables with unknown parameters. Since the joint parameter estimation of the dual-bias is numerically difficult for deep Learning-based IQA model, we adapt the computationally efficient alternating optimization algorithm, i.e., Expectation Maximization (EM) algorithm, to our dual-bias calibration task, which consists of iterative Expectation step (E-step) and Maximization step (M-Step). The E-step computes the expectation of the log likelihood function of given parameters w.r.t. all possible subjective bias values and observed LC-MOS. Then, the M-step updates the parameters of subjective bias by differentiating the quantity computed in the E-step w.r.t. the subjective bias. Thus, we can jointly estimate the parameters of the dual-bias and adaptively suppress the misleading of LC-MOS via a Gated Dual-Bias Calibration (GDBC) module. Meanwhile, GDBC also achieved comparable performance in terms of four metrics using LC-MOS to the IQA model learned from LA-MOS. To the best of our knowledge, this is the first exploration of robust IQA model learning from noisy LC-MOS. For clarity, the main contributions of this paper are summarized in the following:

- We propose an alternative optimization-based dual-bias (including subjective bias and model bias) calibration method for robust IQA model learning from noisy LC-MOS, which significantly reduces the learning cost.
- We develop a GDBC module to adaptively update the estimated subjective bias by measuring the stability of IQA model learning in neighboring iterations, which reduces the risk of overadjustment.
- We verify the effectiveness of the proposed GDBC method from both the theoretical and experimental analysis, which achieves state-of-the-art performance when very few opinion scores are available for each image.

## II. RELATED WORK

In this section, we first briefly review Learning-based IQA methods. Then we introduce related work on robust classification models and subjective annotation of practical conditions.

### A. Learning-based Image Quality Assessment

As an important research area in image processing, IQA is essential in many applications such as image compression, image restoration, medical imaging, etc., where the quality of the visual information is critical for accurate analysis and interpretation. Recently, Learning-based IQA has received considerable attention due to the powerful ability of DNNs. For example, Ma et al. [13] proposed a multi-task DNN for IQA with Learning-based end-to-end optimization utilizing auxiliary distortion information. Zhang et al. [14] designed a deep bilinear convolutional neural network (DNN) for IQA with both synthetic and authentic distortions. Talebi et al. [15] proposed a convolutional neural network predicting score distribution for neural image assessment (NIA). Su et al. [16] proposed a self-adaptive hyper-network architecture for IQA (HET) in the wild. Zhang et al. proposed a unified IQA model UNIQUE [17] that tends to preserve the pairwise rank and prediction difference rather than the MOS of each single image. Golestaneh et al. proposed a IQA model TReS [18] via transformers, relative ranking, and self-consistency. Sun et al. [19] developed a distortion graph representation learning framework for IQA. In addition, new paradigms for Learning-based IQA have emerged in complex scenarios. For example, the meta-learning IQA for fast adaptation [20], the generative adversarial network (GAN) for active inference [21], the evolvable predictive head for continuous learning [22], the disentangled representation based on variational auto-encoders (VAE) for image generation [7], the vision-language correspondence [23] for multimodal scenarios, and the perceptual attack for security scenarios [24].

Models like UNIQUE [17], and TReS [18] do not rely on single image-based training losses. Instead, these models preprocess the data and use rank loss terms, affecting the prediction MSE w.r.t. MOS. Since deep learning is an end-to-end optimization that relies on quality score regression, the superior performance of the above Learning-based IQA models heavily depends on reliable subjective quality labels, especially on a large-scale image database. Moreover, there is still a strong demand to construct new datasets for many emerging scenarios and tasks, including distorted images [3], virtual reality (VR) [25], [26], light field [27], hazy images [28], [29], smartphone photography [30] 3D CG image [31], [32] and etc. Therefore, it is necessary and urgent to explore feasible IQA models in the subjective bias scenario.

### B. Robust Model Learning and Subjective Label Screening

Completely clean labels are difficult to obtain in practical conditions. Researchers have found the over-parameterized network can learn any complex function from corrupted labels [33], [34], [35], [36], [37]. Zhang et al. [38] demonstrated that DNNs can easily fit the entire training dataset with any

corrupted label ratio, ultimately leading to less generality on the test dataset. To train efficient DNNs in noisy cases, many methods have been proposed including robust loss functions [34], [35], [39], regularization [40], robust network architecture [36], sample selection [37], training strategy [41], [42], and etc. These methods focus on robust classification problems and the noise of perturbed labels is assumed to obey a uniform distribution. The design of most methods is based on noise tolerance and one-hot label properties such as sparsity regularization [43], [40]. These methods cannot be directly transferred to LC-MOS for IQA regression tasks due to the inconsistency of data properties.

In the practical condition for IQA, the opinion score of each annotator is biased against the ideal objective label, which is different from artificial perturbation in classification problems [44], [45], [46]. There are different acquisition and processing methods for IQA datasets [3], such as ensuring the consistency of the subjective evaluation environment [4], adding post-processing to the collected data [47], and discarding outliers [5]. These datasets are collected through crowd sourcing and require multiple annotators' opinion scores, which is very time-consuming and expensive. Furthermore, the International Telecommunication Union (ITU) and researchers have proposed a number of standards [10], [11], [12] for crowd-sourced data processing to eliminate the subjective bias in MOS, containing the model based on subject rejection in BT500 [10], the model based on subject bias/inconsistency modeling and maximum likelihood estimation in P910 [11], [48], and the model based on subject bias removal in P913 [12]. However, these methods rely on sufficient annotation information and cannot be directly transferred to IQA models under LC-MOS scenarios.

### III. METHODS

In this section, we first introduce the preliminaries of subjective bias problem formulation. Next, we describe the misleading effect of LC-MOS and propose an expectation-maximization-based dual-bias calibration scheme.

#### A. Preliminaries

Let  $x \in \mathcal{X} \subset \mathbb{R}^d$  denote an  $d$ -dimensional image, and  $y/y^* \in \mathcal{Y}$  denote its corresponding LC-MOS/LA-MOS, i.e.,

$$\begin{aligned} y &= \frac{1}{M} \sum_{m=1}^M R_m \\ y^* &= \frac{1}{S} \sum_{s=1}^S R_s \end{aligned}, \quad (1)$$

where  $R_m/R_s$  denote the  $m$ -th /  $s$ -th manual annotation for  $x$ ,  $\{R_m\}_{m=1}^M \subset \{R_s\}_{s=1}^S$ , and  $M \ll S$ . To simplify the discussion, we represent the LC-MOS and LA-MOS with a normalized label space, i.e.  $\mathcal{Y} \subset [0, 1]$ , and a higher  $y/y^*$  means better subjective quality in terms of very few/abundant manual annotations. In this context, the subjective bias  $z$  is defined as the difference between  $y$  and  $y^*$ , i.e.,

$$z = y - y^*, \quad (2)$$

which is assumed to follow a Gaussian distribution with unknown parameters, i.e.,  $z \sim \mathcal{N}(\mu_z, \sigma_z^2)$ .

Learning-based IQA aims to obtain a parametric model  $f_\theta : \mathcal{X} \rightarrow \mathcal{Y}$  that maps the image space to the subjective quality-based label space, where  $\theta$  is learned from paired image and label samples. Typically, the LA-MOS serves as the label, and we derive the optimal parameter  $\theta^*$  by minimizing the risk  $R$  defined in the following,

$$\theta^* = \arg \min_{\theta \in \Theta} R(\theta), \quad (3)$$

where  $\Theta$  is the available parameter set for  $f_\theta$ , and  $R(\theta)$  is measured with the expectation of the loss between  $f_\theta(x)$  and  $y^*$  on the training set, i.e.,

$$R(\theta) = \mathbb{E}_{x,y^*} [\mathcal{L}(f_\theta(x), y^*)], \quad (4)$$

where  $\mathcal{L}(\cdot, \cdot)$  is the loss of the IQA model with respect to the label for each training sample. To save annotation costs, this paper tries to replace parts of LA-MOS with LC-MOS, which may mislead the IQA model learning. Let  $\theta^*$  denote the optimal parameters learned from  $y^*$ . We define the model bias  $b$  by

$$b = f_\theta(x) - f_{\theta^*}(x). \quad (5)$$

In the following, a theoretical analysis of the misleading effect of LC-MOS on the popular square-error loss function is conducted. Then, we put forward our Gated Dual-Bias Calibration (GDBC) module to efficiently suppress the aforementioned misleading effect, which enforces  $f_\theta(x)$  to approach  $y^*$  rather than  $y$ .

#### B. Misleading Effect of LC-MOS

Following the discussion in [34], we denote the noisy labels with bias rate  $\eta$  by  $y^\eta$ , i.e.,

$$y^\eta = \begin{cases} y, & \text{with probability } \eta \\ y^*, & \text{with probability } 1 - \eta \end{cases}. \quad (6)$$

Then, given a collection of training samples with previous noisy labels  $y^\eta$ , IQA model learning usually employs square error to measure the loss of  $f_\theta(x)$  with respect to  $y^\eta$ , i.e.,

$$\mathcal{L}(f_\theta(x), y^\eta) = \|f_\theta(x) - y^\eta\|_2^2. \quad (7)$$

Let  $R^\eta(\theta) = \mathbb{E}_{x,y^\eta} [\mathcal{L}(f_\theta(x), y^\eta)]$  denote the risk with bias rate  $\eta$ , and  $\theta^{*,\eta}$  denote the parameter for the global minimum of risk  $R^\eta(\cdot)$ . To facilitate the analysis, we rewrite the risk  $R^\eta(\theta)$  with bias rate  $\eta$  in the following expanded form, i.e.,

$$\begin{aligned} R^\eta(\theta) &= \mathbb{E}_{x,y^\eta} [\mathcal{L}(f_\theta(x), y^\eta)] \\ &= \mathbb{E}_x \mathbb{E}_{y^*|x} \mathbb{E}_{y^\eta|x,y^*} [\mathcal{L}(f_\theta(x), y^\eta)] \\ &= \mathbb{E}_x \mathbb{E}_{y^*|x} \{(1 - \eta) \mathcal{L}(f_\theta(x), y^*) + \\ &\quad \eta \mathbb{E}_{z|x,y^*} [\mathcal{L}(f_\theta(x), y^* + z)]\} \\ &= \mathbb{E}_x \mathbb{E}_{y^*|x} \{(1 - \eta) \mathcal{L}(f_\theta(x), y^*) + \\ &\quad \eta \mathbb{E}_{z|x,y^*} [\mathcal{L}(f_\theta(x), y^*) - 2z(f_\theta(x) - y^*) + z^2]\} \\ &= \mathbb{E}_x \mathbb{E}_{y^*|x} \{\mathcal{L}(f_\theta(x), y^*) + \\ &\quad \eta \mathbb{E}_{z|x,y^*} [z^2 - 2z(f_\theta(x) - y^*)]\} \\ &= R(\theta) - \eta \mathbb{E}_x \mathbb{E}_{y^*|x} [2\mu_z(f_\theta(x) - y^*) + \mu_z^2 + \sigma_z^2]. \end{aligned} \quad (8)$$

Then, given the optimal and random parameters of  $R(\cdot)$ , i.e.,  $\theta^*$  and  $\theta$ , we can represent their risk difference  $D$  under  $R^\eta(\cdot)$  by

$$\begin{aligned} D &= R^\eta(\theta^*) - R^\eta(\theta) \\ &= R(\theta^*) - R(\theta) - \\ &\quad \eta \mathbb{E}_x \mathbb{E}_{y^*|x} [2\mu_z(f_{\theta^*}(x) - y^*) + \mu_z^2 + \sigma_z^2] + \quad (9) \\ &\quad \eta \mathbb{E}_x \mathbb{E}_{y^*|x} [2\mu_z(f_\theta(x) - y^*) + \mu_z^2 + \sigma_z^2] \\ &= R(\theta^*) - R(\theta) + \eta \mathbb{E}_x \mathbb{E}_{y^*|x} [2\mu_z b]. \end{aligned}$$

Although  $R(\theta^*) - R(\theta) \leq 0$ , we can not guarantee that  $R^\eta(\theta^*) - R^\eta(\theta) \leq 0$  due to the uncertainty of  $\mu_z$  and  $b$ , which are both probably nonzero and with the same signs. That is,  $\theta^*$  does not necessarily equal to  $\theta^{*,\eta}$  when training with the square error loss. In addition, this misleading effect of LC-MOS would become more significant when a higher bias rate  $\eta$  or larger subjective bias  $\mu_z$  are applied to the Eq. (9). Therefore, it is urgent to develop a robust learning framework to train the IQA model from the low-cost noisy labels.

### C. Expectation-maximization-based Dual-bias Calibration

We propose an expectation-maximization-based dual-bias calibration framework to alleviate the above-mentioned challenges of biased LC-MOS. The framework is illustrated in Fig. 2. The algorithm alternates between model update and bias update. During the model update process, the neural network parameters are updated through backpropagation, and image features are learned for quality evaluation. During the bias update process, we repeatedly estimate the subjective bias. The latter subjective bias in turn helps the model learn. The whole process is a process of mutual promotion so that the robustness of the model is improved.

According to Eq. (9), if we know  $z$  and replace  $y^\eta$  by  $y^\eta - z$  in Eq. (7), both the subjective and model biases would be pushed to zero, which improves the noisy label tolerance of IQA model learning. To this end, we consider  $Z = \{(z_i)_{i=1}^n\}$  as the latent variable of  $n$  training sample and represent the likelihood of the unknown parameter set  $\Omega = \{\theta, (\omega_i)_{i=1}^n\}$  by

$$L(\Omega; Y^\eta) = \prod_{i=1}^n p(y_i^\eta | \theta, \omega_i) \quad , \quad (10)$$

where  $Y^\eta = \{(y_i^\eta)_{i=1}^n\}$  denotes  $n$  independently observed noisy labels and  $\omega_i = \{\mu_{z_i}, \sigma_{z_i}\}$ . Following [49], we employ the expectation-maximization-based iterative method to derive  $\Omega$  for maximizing  $L(\Omega; Y^\eta)$ , which helps us achieve the dual-bias calibration. Let  $\Omega^t = \{\theta^t, (\omega_i^t)_{i=1}^n\}$  denote the estimated parameters in the  $t$ th iteration. We first conduct the Expectation step (E-step) by computing the conditional expectation of the log-likelihood,

$$\begin{aligned} Q(\Omega | \Omega^t) &= \mathbb{E}_{Z|Y^\eta, \Omega^t} [\log L(\Omega; Y^\eta, Z)] \\ &= \mathbb{E}_{Z|Y^\eta, \Omega^t} \left[ \log \prod_{i=1}^n p(y_i^\eta, z_i | \Omega) \right] \\ &= \sum_{i=1}^n \int p(z_i | y_i^\eta, \Omega^t) [\log p(y_i^\eta | z_i, \Omega) + \\ &\quad \log p(z_i | \Omega)] dz_i \quad . \quad (11) \end{aligned}$$

For dual-bias calibration, we want to use the calibrated label  $y_i^\eta - z_i$  to supervise the IQA model  $f_\theta(x_i)$ , which could be transformed to maximizing the posterior of the observed noisy label  $y_i^\eta$  with  $f_\theta(x_i) + z_i$ . Based on this requirement, we assume that  $y_i^\eta | z_i, \Omega \sim \mathcal{N}(f_\theta(x_i) + z_i, \sigma_{y_i^\eta | z_i, \Omega}^2)$ , which enforces  $f_\theta(x_i)$  to approach  $y_i^*$  according to Eq. (2). In addition, the conditional distribution of  $z_i | y_i^\eta, \Omega^t$  could be derived from the Bayes theorem

$$p(z_i | y_i^\eta, \Omega^t) = \frac{p(y_i^\eta | z_i, \Omega^t)p(z_i | \Omega^t)}{\int p(y_i^\eta | z_i, \Omega^t)p(z_i | \Omega^t)dz_i}, \quad (12)$$

where we obtain<sup>1</sup> that  $z_i | y_i^\eta, \Omega^t \sim \mathcal{N}(\mu_{z_i | y_i^\eta, \Omega^t}, \sigma_{z_i | y_i^\eta, \Omega^t})$  and

$$\mu_{z_i | y_i^\eta, \Omega^t} = \frac{\sigma_{y_i^\eta | z_i, \Omega^t}^2 \mu_{z_i}^t + \sigma_{z_i | \Omega^t}^2 [y_i^\eta - f_{\theta^t}(x_i)]}{\sigma_{y_i^\eta | z_i, \Omega^t}^2 + \sigma_{z_i | \Omega^t}^2}. \quad (13)$$

By plugging the probability density of  $z_i | y_i^\eta, \Omega^t$ ,  $y_i^\eta | z_i, \Omega$  and  $z_i | \Omega$  into Eq. (11), we rewrite the conditional expectation of the log-likelihood by

$$\begin{aligned} Q(\Omega | \Omega^t) &= -\frac{1}{2} \sum_{i=1}^n \left[ \frac{\sigma_{z_i | y_i^\eta, \Omega^t}^2 + (\mu_{z_i} - \mu_{z_i | y_i^\eta, \Omega^t})^2}{\sigma_{z_i | \Omega}^2} \right. \\ &\quad \left. + \frac{\sigma_{z_i | y_i^\eta, \Omega^t}^2 + [y_i^\eta - f_{\theta^t}(x_i) - \mu_{z_i | y_i^\eta, \Omega^t}]^2}{\sigma_{y_i^\eta | z_i, \Omega}^2} \right. \\ &\quad \left. + \log(2\pi\sigma_{z_i | \Omega}^2) + \log(2\pi\sigma_{y_i^\eta | z_i, \Omega}^2) \right] \end{aligned} \quad . \quad (14)$$

Then, by setting the derivative of  $Q(\Omega | \Omega^t)$  w.r.t.  $\mu_{z_i}$  to zero, we could obtain the updated parameter of subjective bias in the Maximization step (M step), i.e.,

$$\mu_{z_i}^{t+1} = \alpha \mu_{z_i}^t + (1 - \alpha) c_i^t, \quad (15)$$

where  $\alpha = \frac{\sigma_{y_i^\eta | z_i, \Omega}^2}{\sigma_{z_i | \Omega}^2 + \sigma_{y_i^\eta | z_i, \Omega}^2}$ ,  $c_i^t = y_i^\eta - f_{\theta^t}(x_i)$ , and  $\mu_{z_i}^0 = 0$ . It is noted that the fitting error  $c_i^t$  usually converges to a small value when training with clean labels in several iterations [50].

Unbounded updating of Eq. (15) may result in overadjustment. To address this issue, we develop a Gated Dual-Bias Calibration (GDBC) module by measuring the stability of IQA model learning in neighboring iterations, i.e.,

$$\mu_{z_i}^{t+1} = \begin{cases} \alpha \mu_{z_i}^t + (1 - \alpha) c_i^t, & \|C\|_1 > t_h \epsilon \\ \mu_{z_i}^t, & \text{otherwise} \end{cases}, \quad (16)$$

where  $C = [c_i^{t-t_h}, \dots, c_i^t]^T$  represents the fitting errors of the IQA model in the neighboring  $t_h$  iterations, and the subjective bias calibration only activates when  $C$ 's  $l_1$  norm exceeds a threshold  $\epsilon$ .

Since  $\mu_{z_i}^{t+1}$  maximizes the probability of  $z_i$ , we could suppress the model bias by removing  $\mu_{z_i}^{t+1}$  from the noisy LC-MOS  $y_i^\eta$ , and optimize the IQA model parameters via

$$\theta^{t+1} = \theta^t - \lambda \nabla_\theta \left[ \frac{1}{n} \sum_{i=1}^n \mathcal{L}(f_{\theta^t}(x_i), y_i^\eta - \mu_{z_i}^{t+1}) \right], \quad (17)$$

where  $\lambda$  is the learning rate, and  $\nabla_\theta$  denotes the gradient operator. We repeat the alternative subjective bias and model

<sup>1</sup>Detailed proof is given in the supplementary material.

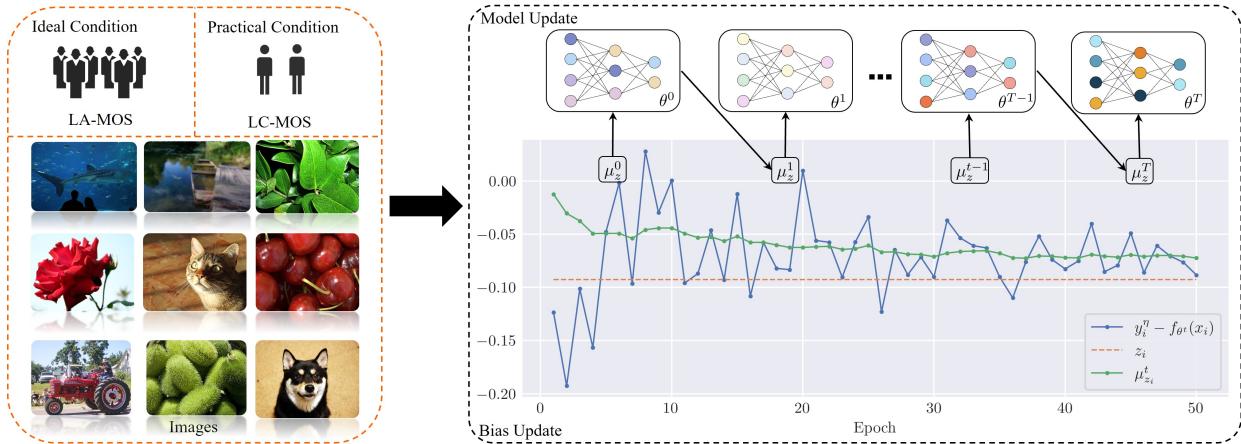


Fig. 2. The proposed framework. The left side demonstrates that subjects of practical conditions differ from ideal conditions. The right above shows an alternate optimization process the expectation maximization (EM) algorithm. In each iteration, the network parameter  $\theta^t$  and bias  $\mu_z^t$  are updated to improve the subsequent iterations  $\theta^{t+1}$  and bias  $\mu_z^{t+1}$ , the colour changes of the network nodes in the figure represent the update of the network parameter  $\theta$ . the right bottom shows an actual case for the estimated bias update process, the real bias here is close to -0.09. The blue, yellow, and green lines in the figure represent the fitting error  $y_i^n - f_{\theta^t}(x_i)$ , the real bias  $z_i$ , and our estimated bias  $\mu_z^t$ , respectively. The horizontal axis represents the number of iterations, and the vertical axis represents the corresponding value.

### Algorithm 1 Pseudo-code of our method.

**Require:** input image  $x$ , MOS target with bias  $y^n$ , Subject bias  $\mu$ , fitting error  $c$ , parameters  $\alpha$ , threshold  $\epsilon$ .

```

1: for  $t = 1, \dots, T$  do
2:   for  $i = 1, \dots, n$  do                                 $\triangleright$  Process batch
3:      $f_{\theta^t}(x_i)$                                  $\triangleright$  Forward pass
4:      $c_i^t = y_i^n - f_{\theta^t}(x_i)$                  $\triangleright$  Update fitting error
5:      $C = [c_1^{t-t_h}, \dots, c_n^{t-t_h}]^T$        $\triangleright$  Update  $\mu_z^{t+1}$ 
6:     if  $\|C\|_1 > t_h \epsilon$  then                   $\triangleright$  Update  $C$ 
7:        $\mu_z^{t+1} \leftarrow \alpha \mu_z^t + (1 - \alpha) c_i^t$ 
8:     else
9:        $\mu_z^{t+1} \leftarrow \mu_z^t$ 
10:    loss  $\leftarrow [\frac{1}{n} \sum_{i=1}^n \mathcal{L}(f_{\theta^t}(x_i), y_i^n - \mu_z^{t+1})]$ ,
11:     $\theta^{t+1} \leftarrow \theta^t - \lambda \nabla_{\theta}$  loss.  $\triangleright$  Update model parameters
Return:  $f_{\theta^G} = f_{\theta^T}$                                  $\triangleright$  Final output after  $T$  iterations

```

bias calibrations until the maximum iteration steps are reached. Finally, we could output a robust IQA model  $f_{\theta^G}$  even when learning with noisy LC-MOS. Algorithm 1 provides the pseudo-code of our method.

## IV. EXPERIMENTS

### A. Protocol

We evaluate the proposed GDBC method on four popular IQA databases, i.e., VCL [5], CSIQ [51], LIVEC [52] and KONIQ [53], which only provide the LA-MOS for all images. Let  $M$  denote the annotation number to be simulated in the experiments. We develop the following three LC-MOS settings according to the annotation resources of different databases:

- 1) When raw opinion scores of all subjects are available (such as VCL [5]), we randomly sample  $M$  scores for each image.

- 2) When pairwise MOS and standard deviation are available (such as CSIQ [51] and LIVEC [52]), we use them to simulate a normal distribution of subjective ratings [54], [55], and randomly sample  $M$  scores from this distribution.
- 3) When the empirical distribution of raw opinion scores is available (such as KONIQ [53]), we randomly sample  $M$  scores from this empirical distribution.

For each image, the mean value of the previously sampled  $M$  scores is used as the LC-MOS, where  $M$  is smaller than the minimum requirement of ITU recommendations [10], [11]. More specifically, we investigate four candidate numbers, i.e.,  $M = \{1, 2, 4, 8\}$ . To validate the universality of the proposed method, we select eight representative deep neural networks for IQA model learning, i.e., RET [56], NIA [15], DNN [14], HET [16], TIQ [57], MQA [58], DQT [59] and MIQ [60].

Following the criterion of [14], [16], [15], we randomly split each database into non-overlapped training and testing sets, which cover 80% and 20% samples respectively. To eliminate the performance bias for specific LC-MOS or train-test split, we repeat the random LC-MOS sampling and train-test splitting 10 times for each database, and report the median results across all trials for evaluation.

Let LA-MOS denote the ground-truth subjective quality of each image. Four widely used metrics are used for evaluating the performance of the IQA models learned from the LA-MOS, LC-MOS, and our GDBC method, i.e., the Pearson's Linear Correlation Coefficient (PLCC) [61], the Spearman's Rank Order Correlation Coefficient (SRCC) [62], Kendall's Rank Correlation Coefficient (KRCC) and Root Mean Square Error (RMSE). In addition, to highlight the performance improvement of the proposed method when training with LC-

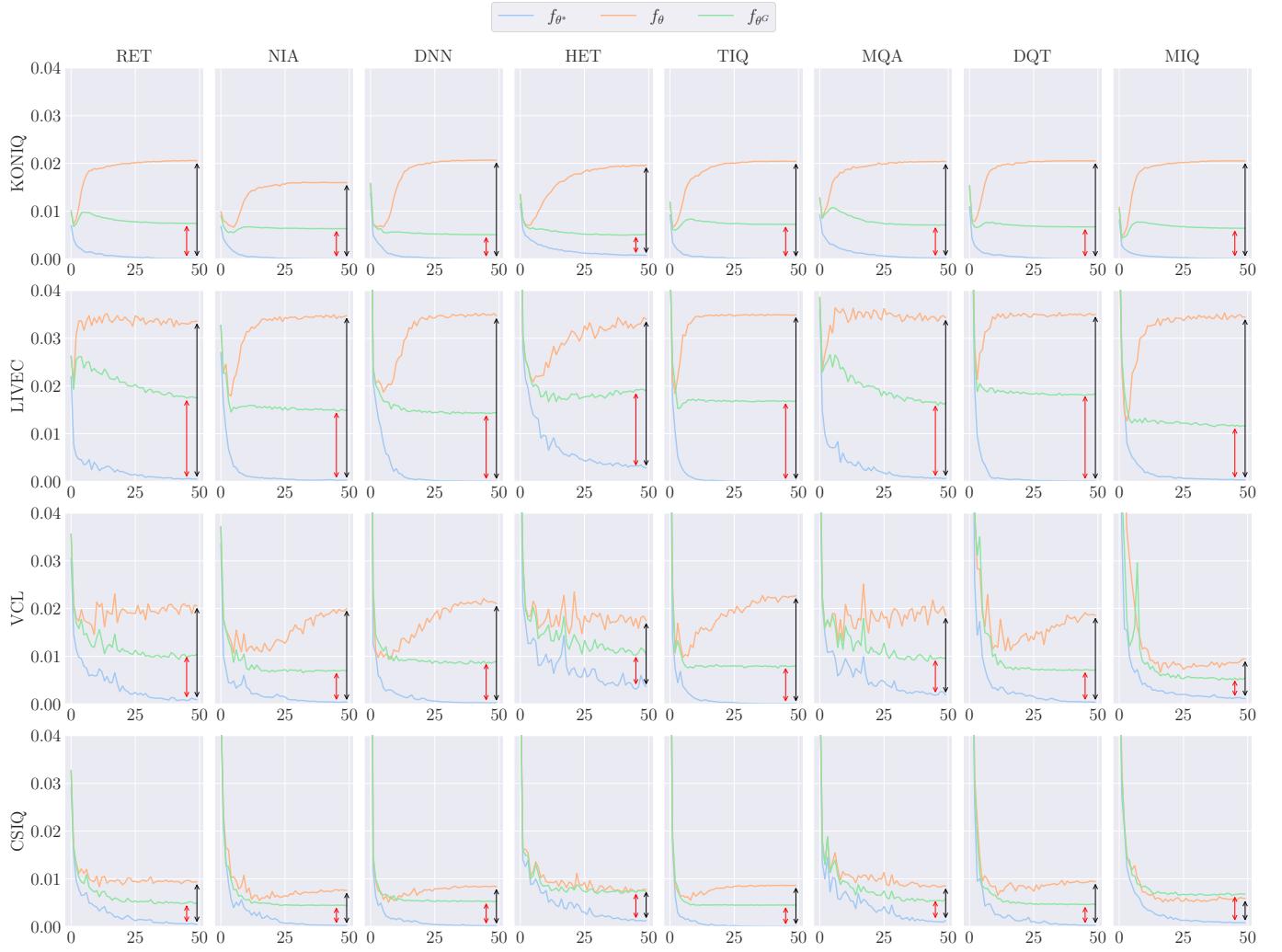


Fig. 3. The MSE curves w.r.t.  $y^*$  for various MOS patterns are depicted.  $f_{\theta^G}$  shows the model with GDBC for  $y^\eta$ , while  $f_{\theta^*}$  and  $f_\theta$  represent models trained under  $y^*$  and  $y^\eta$ , respectively. Red double arrow lines shows discrepancies of  $f_{\theta^G}$  compared to  $f_{\theta^*}$ . In contrast, the black arrow line shows discrepancies for  $f_\theta$  compared to  $f_{\theta^*}$ .

MOS, we also introduce a relative index  $\Delta(\%)$ , i.e.,

$$\Delta(\%) = \frac{m_w / \text{GDBC} - m_{w/o} \text{GDBC}}{m_{w/o} \text{GDBC}} \times 100, \quad (18)$$

where  $m_w / \text{GDBC}$  and  $m_{w/o} \text{GDBC}$  represent the evaluation metrics of the IQA model trained with and without the GDBC module, respectively. The variable  $m$  denotes the corresponding metrics, including SRCC, PLCC, KRCC, and RMSE, as previously discussed.

In our experiment, we train the IQA models with Adam optimizer [63] and set  $\alpha$  to 0.9, we experimentally set  $t_h$  using grid search [64],  $\epsilon$  to 0.01 or 0.1, epoch to 50, batch size to 16. The optimal learning rates are founded by grid search and scheduled by the cosine annealing rule [65]. During training and inference, we scale and center crop  $320 \times 320 \times 3$  sub-images from the original image without changing their aspect ratio. All experiments are performed on a workstation with a single NVIDIA GeForce RTX 3090 GPU.

TABLE I  
MSE BETWEEN CALIBRATED LC-MOS AND LA-MOS.

Datasets	LC-MOS	DNN	HET	NIA	RET	TIQ	MQA	DQT	MIQ
KONIQ	0.021	0.006	<b>0.005</b>	0.006	0.008	0.007	0.007	0.007	0.008
LIVEC	0.035	0.014	0.016	0.014	0.019	0.017	0.016	0.018	<b>0.012</b>
VCL	0.024	0.009	0.010	0.007	0.012	0.008	0.009	0.007	<b>0.005</b>
CSIQ	0.009	0.005	0.004	<b>0.003</b>	0.004	0.005	0.005	0.005	0.010

### B. Performance evaluation for the GDBC

To demonstrate the effectiveness of the proposed GDBC method, we first investigate the most challenging case with the setting of  $\eta = 100\%$  and  $M = 1$ , which means that all training samples are labeled with the noisy LC-MOS and only one annotation is available for each image.

1) *Effectiveness of subjective bias calibration:* Let  $y_i^\eta - \mu_{z_i}^{t+1}$  in Eq. 17 denote the calibrated LC-MOS. In Table I,

we report the subjective bias calibration results when GDBC is applied to different deep IQA models, which measure the MSE between the calibrated LC-MOS and its corresponding LA-MOS. Due to the difference in LC-MOS settings, the MSE between the raw LC-MOS and LA-MOS are different across different IQA databases, where the LIVEC and CSIQ present the largest and smallest bias respectively. Meanwhile, it is seen that the MSE values of the calibrated LC-MOS are much smaller than the raw LC-MOS when GDBC is applied to all deep IQA models, whose MSE reductions range from 64% to 77%. These results verify the effectiveness and universality of the proposed method in reducing subjective bias. In the following, we will further investigate the benefits of subjective bias calibration for IQA model learning.

*2) Guidance for the IQA model learning:* In Fig. 3, we present the MSE curves with  $y^*$  for different IQA models under  $y^*$  or  $y^\eta$  on the KONIQ dataset, illustrating the learning behavior under different MOS patterns. Notably,  $f_\theta$  shows significant overfitting, evidenced by rising MSE over epochs, due to noisy subjective bias that complicates network modeling [50]. Conversely,  $f_{\theta^G}$  exhibits reduced MSE, because the GDBC module effectively adjusts these biases, simplifying model learning and mitigating overfitting. Visually, the red double arrow lines indicate smaller discrepancies between  $f_{\theta^G}$  and  $f_{\theta^*}$ , suggesting a closer alignment due to the GDBC. In contrast, the black arrow lines show larger discrepancies for  $f_\theta$ , indicating greater deviation from  $f_{\theta^*}$ . This indicates  $f_{\theta^G}$  is closer to  $f_{\theta^*}$  than  $f_\theta$ .

*3) Effectiveness of model bias calibration:* In this section, we evaluate the LA-MOS prediction performance of different deep IQA models when they are trained with different labels and strategies. As shown in Table II, all IQA models trained with the LA-MOS achieve better performance than their counterparts trained with the LC-MOS. These results further confirm the issue of LC-MOS overfitting observed in Fig. 3. Meanwhile, we can also find that the proposed GDBC efficiently improves the performance of IQA models when they are trained with LC-MOS, where all  $\Delta(\%)$  report positive values across different databases.

In addition, as shown in each row of Table II, we can find that the performance improvement of GDBC is relatively smaller in CSIQ than in the other databases. There are two possible reasons to account for this result. Firstly, the CSIQ database contains the smallest LC-MOS bias as shown in Table I, which limits the improvement space of GDBC. Secondly, in comparison with the authentic distortion databases (i.e., KONIQ and LIVEC), the diversity of the CSIQ database is limited, which reduces the discrepancy between the training and testing sets and the potential overfitting risk. When we look at each column of Table II, it is seen that the performance improvement of GDBC changes significantly across different IQA models, where the DNN and HET obtain higher  $\Delta(\%)$ . A possible reason lies in the bias propagation. Actually, unlike the plain single-branch architecture in RET and NIA, DNN and HET are composed of multiple branches, which are beneficial for capturing more comprehensive quality-aware features. However, when training with LC-MOS, the subjective bias would easily propagate between different branches and

enlarge the misleading effect of LC-MOS. In these cases, the superiority of GDBC would become more significant.

*4) Comparison of continual noisy IQA tasks performance:* To explore the compatibility of our model, we conduct experiments in more realistic scenarios, namely the performance of IQA under continuous and subjective bias scenarios. We organize these datasets in order, i.e., KONIQ → LIVEC → VCL → CSIQ. The setting of noisy labels is added to be consistent with Section IV-B3. We select two models for the continual noisy IQA task, i.e., SLF [66], CQA [22] with or without our GDBC module. The CQA is trained with preprocessed labels and uses fidelity loss, while SLF model is trained by normalizing labels and using MSE loss. We present the quantitative and qualitative results when no previous data is available during new task learning. We summarize the SRCC results continually in Table V.

Note that all methods begin training on KONIQ, there are several useful findings. CQA uses a distillation loss to memorize previous results. However, the previous results themselves are in a noisy overfitting situation. SLF uses a pruned strategy to avoid excessive adjustments to original parameters. However, the representation ability of the pruned network is even worse in a noisy environment. They are even worse than CQA, which uses a direct distillation loss. On the contrary, the GDBC module is conducive to alleviating forgetting

We use the proposed mean SRCC (mSRCC), mean plasticity index (mPI), mean stability index (mSI), and mean plasticity-stability index (mPSI) to benchmark the performance [22]. From Table VI, we observe several interesting findings. Since CQA uses distillation loss to retain previous results, its performance remains more stable in terms of mSI. However, because it also retains incorrect labels, its performance in mSRCC is not satisfactory. Since SLF uses a pruned strategy to avoid excessive adjustments to original parameters, the mSRCC of the pruned network is worse in a noisy environment. In contrast, the GDBC module achieves superior results across multiple metrics. This improvement is due to the calibration of labels, which mitigates the impact of subjective bias. Additionally, the corrected labels facilitate the retention of previous model results when encountering new tasks, thereby enhancing model stability and supporting continuous learning.

*5) Ablation study on the timing performance:* We conducted an investigation experiment of the running time (Seconds) per iteration in training phase across various IQA models with/without the GDBC module, as shown in Table VII. We define the running time percentage change index as  $T\% = \frac{(T_w - T_{w/o})}{T_{w/o}} \times 100$ , where  $T_w$  and  $T_{w/o}$  represents running time with/without the GDBC module, respectively. The reported results further verify the efficiency of the proposed GDBC, whose maximum training time increase is only 1.2%.

### C. Robustness toward different subjective bias settings

Besides different IQA models and databases, we also investigate the robustness of the proposed method toward different subjective bias settings by changing the bias rate and annotation number. More specifically, a higher bias rate and a smaller

TABLE II

PERFORMANCE OF DIFFERENT METHODS ON DIFFERENT DATASETS WITH BIASED LABELS (BIAS RATE IS 100% AND THE ANNOTATION NUMBER IS 1). “-A”, “-C”, “-G” INDICATES LA-MOS, LC-MOS AND LC-MOS WITH GDBC RESPECTIVELY.

Methods	KONIQ				LIVEC				VCL				CSIQ			
	↑SRCC	↑PLCC	↑KRCC	↓RMSE	↑SRCC	↑PLCC	↑KRCC	↓RMSE	↑SRCC	↑PLCC	↑KRCC	↓RMSE	↑SRCC	↑PLCC	↑KRCC	↓RMSE
RET-A	0.8985	0.9214	0.7290	0.0557	0.8215	0.8624	0.6327	0.1085	0.9359	0.9234	0.7741	0.1278	0.9242	0.9238	0.7575	0.1423
RET-C	0.7905	0.8361	0.5990	0.0780	0.7648	0.8021	0.5699	0.1241	0.8503	0.8376	0.6469	0.2145	0.8992	0.8961	0.7276	0.1650
RET-G	0.8294	0.8659	0.6425	0.0716	0.7819	0.8179	0.5885	0.1253	0.8723	0.8531	0.6724	0.1833	0.9126	0.9224	0.7533	0.1365
Δ(%)	<b>4.9209</b>	<b>3.5642</b>	<b>7.2621</b>	<b>-8.2644</b>	<b>2.2359</b>	<b>1.9698</b>	<b>3.2637</b>	<b>0.9879</b>	<b>2.5873</b>	<b>1.8505</b>	<b>3.9419</b>	<b>-14.5601</b>	<b>1.4902</b>	<b>2.9349</b>	<b>3.5322</b>	<b>-17.2214</b>
NIA-A	0.8981	0.9179	0.7287	0.0563	0.8011	0.8404	0.6112	0.1123	0.9389	0.9067	0.7759	0.1702	0.9263	0.9369	0.7661	0.1057
NIA-C	0.7551	0.7363	0.5626	0.1080	0.6797	0.7017	0.4916	0.1559	0.8846	0.8405	0.6987	0.1869	0.9101	0.9222	0.7362	0.1198
NIA-G	0.8216	0.8519	0.6357	0.0743	0.7108	0.7376	0.5226	0.1403	0.9325	0.9216	0.7754	0.1704	0.9229	0.9349	0.7664	0.0898
Δ(%)	<b>8.8068</b>	<b>15.7001</b>	<b>12.9932</b>	<b>-31.2628</b>	<b>4.5755</b>	<b>5.1161</b>	<b>6.3059</b>	<b>-10.0343</b>	<b>5.4149</b>	<b>9.6490</b>	<b>10.9770</b>	<b>-8.8235</b>	<b>1.4080</b>	<b>1.3769</b>	<b>4.0958</b>	<b>-25.0117</b>
DNN-A	0.8899	0.9083	0.7161	0.0583	0.7390	0.7651	0.5615	0.1314	0.9209	0.9199	0.7443	0.1288	0.9498	0.9604	0.8096	0.0931
DNN-C	0.7087	0.7474	0.5189	0.0987	0.6579	0.6812	0.4756	0.1790	0.8443	0.8633	0.6478	0.2140	0.9233	0.9301	0.7600	0.0895
DNN-G	0.8053	0.8424	0.6137	0.0748	0.7007	0.7095	0.5181	0.1496	0.9074	0.9159	0.7377	0.1464	0.9463	0.9552	0.8067	0.0842
Δ(%)	<b>13.6306</b>	<b>12.7107</b>	<b>18.2694</b>	<b>-24.1734</b>	<b>6.5055</b>	<b>4.1544</b>	<b>8.9361</b>	<b>-16.4262</b>	<b>7.4736</b>	<b>6.0929</b>	<b>13.8777</b>	<b>-31.5893</b>	<b>2.4903</b>	<b>2.7033</b>	<b>6.1507</b>	<b>-5.9328</b>
HET-A	0.8792	0.9104	0.7029	0.0630	0.8183	0.8637	0.6387	0.1065	0.9437	0.9279	0.7908	0.1175	0.9191	0.9344	0.7523	0.1037
HET-C	0.6709	0.7304	0.4868	0.1013	0.6482	0.6943	0.4694	0.1528	0.8734	0.8374	0.6842	0.1667	0.8705	0.8760	0.6962	0.1551
HET-G	0.8012	0.8452	0.6109	0.0768	0.7200	0.7564	0.5298	0.1560	0.9390	0.9239	0.7860	0.1468	0.9119	0.9348	0.7471	0.1217
Δ(%)	<b>19.4217</b>	<b>15.7174</b>	<b>25.4930</b>	<b>-24.2490</b>	<b>11.0760</b>	<b>8.9443</b>	<b>12.8670</b>	<b>2.0301</b>	<b>7.5109</b>	<b>10.3290</b>	<b>14.878</b>	<b>-11.9075</b>	<b>4.7504</b>	<b>6.7099</b>	<b>7.3149</b>	<b>-21.5172</b>
TIQ-A	0.9052	0.9286	0.7375	0.0513	0.8445	0.8722	0.6588	0.1016	0.9379	0.9244	0.7829	0.1513	0.9220	0.9348	0.7565	0.1155
TIQ-C	0.6617	0.7313	0.4761	0.1155	0.7381	0.7637	0.5447	0.1360	0.8596	0.8070	0.6557	0.1764	0.9098	0.9156	0.7331	0.1298
TIQ-G	0.8100	0.8319	0.6213	0.0788	0.7563	0.7974	0.5662	0.1264	0.9287	0.9103	0.7632	0.1397	0.9543	0.9541	0.8107	0.0991
Δ(%)	<b>22.4045</b>	<b>13.7693</b>	<b>30.5076</b>	<b>-31.7999</b>	<b>2.4598</b>	<b>4.4115</b>	<b>3.9397</b>	<b>-7.0238</b>	<b>8.0336</b>	<b>12.7951</b>	<b>16.3880</b>	<b>-20.8165</b>	<b>4.8940</b>	<b>4.2056</b>	<b>10.5827</b>	<b>-23.6467</b>
MQA-A	0.9032	0.9220	0.7328	0.0555	0.8384	0.8893	0.6610	0.0998	0.9665	0.9652	0.8430	0.0928	0.9265	0.9377	0.7670	0.1110
MQA-C	0.7432	0.7760	0.5513	0.0926	0.7340	0.8074	0.5451	0.1258	0.9212	0.9150	0.7557	0.1247	0.8921	0.9041	0.7262	0.1327
MQA-G	0.8023	0.8458	0.6130	0.0745	0.7755	0.8314	0.5861	0.1274	0.9560	0.9520	0.8162	0.1358	0.9072	0.9238	0.7408	0.1212
Δ(%)	<b>7.9408</b>	<b>8.9999</b>	<b>11.1907</b>	<b>-19.5506</b>	<b>5.6631</b>	<b>2.9667</b>	<b>7.5210</b>	<b>1.2566</b>	<b>3.7710</b>	<b>4.0368</b>	<b>8.0093</b>	<b>8.8725</b>	<b>1.6913</b>	<b>2.1709</b>	<b>2.0047</b>	<b>-8.6814</b>
DQT-A	0.8958	0.9196	0.7245	0.0549	0.8153	0.8734	0.6347	0.1041	0.9292	0.9130	0.7610	0.1257	0.9544	0.9605	0.8122	0.0978
DQT-C	0.7129	0.7374	0.5234	0.1027	0.6636	0.7296	0.4831	0.1443	0.8382	0.8350	0.6469	0.1643	0.9129	0.9208	0.7408	0.1514
DQT-G	0.7993	0.8389	0.6062	0.0753	0.6996	0.7558	0.5124	0.1364	0.8786	0.8725	0.6868	0.1471	0.9386	0.9519	0.7853	0.1117
Δ(%)	<b>12.1185</b>	<b>13.7675</b>	<b>15.8345</b>	<b>-26.6781</b>	<b>5.4193</b>	<b>3.5813</b>	<b>6.0499</b>	<b>-5.4269</b>	<b>4.8159</b>	<b>4.4847</b>	<b>6.1695</b>	<b>-10.4479</b>	<b>2.8187</b>	<b>3.3771</b>	<b>6.0062</b>	<b>-26.2182</b>
MIQ-A	0.9274	0.9414	0.7689	0.0486	0.8688	0.9070	0.6925	0.0881	0.9712	0.9683	0.8588	0.0701	0.9423	0.9492	0.7912	0.1133
MIQ-C	0.8024	0.8194	0.6094	0.0845	0.7625	0.8146	0.5756	0.1231	0.8592	0.8401	0.6807	0.1348	0.8956	0.9076	0.7200	0.1428
MIQ-G	0.8578	0.8921	0.6728	0.0643	0.8035	0.8469	0.6145	0.1186	0.9649	0.9591	0.8333	0.0904	0.9182	0.9313	0.7533	0.1280
Δ(%)	<b>6.9046</b>	<b>8.8680</b>	<b>10.4092</b>	<b>-23.9063</b>	<b>5.3774</b>	<b>3.9578</b>	<b>6.7489</b>	<b>-3.6057</b>	<b>12.3007</b>	<b>14.1558</b>	<b>22.4227</b>	<b>-32.9903</b>	<b>2.5244</b>	<b>2.6132</b>	<b>4.6303</b>	<b>-10.3862</b>

annotation number would causes greater subjective bias. In this section, all experiments are conducted on the KONIQ database.

1) *Different bias rates*: Focusing on the challenging cases, we first set the annotation number to the smallest value, i.e.,  $M = 1$ , and investigate the performance of different IQA models learned with LC-MOS under the following bias rates  $\eta = \{100\%, 80\%, 60\%, 40\%, 0\%\}$ . As shown in Table III, the performance of all IQA models gradually increase with a decreasing  $\eta$ . By contrast, the performance improvement  $\Delta(\%)$  continues to decrease with the decreasing  $\eta$ . Since the impact of subjective bias reduces with a decreasing bias rate, the overfitting risk toward LC-MOS would also reduce, which may limit the room for IQA models' improvement from GDBC. Nevertheless, the proposed GDBC improves the LC-MOS prediction accuracy for all IQA models across different

bias rates, whose SRCC, PLCC, KRCC,  $\Delta(\%)$  values are all positive. When  $\eta$  is equal to 0, the SRCC performance only decreases by less than 1.2%. This verifies that the proposed GDBC is robust to the variations of bias rates.

2) *Different annotation numbers*: On the other hand, we set the bias rate to the highest value, i.e.,  $\eta = 100\%$ , and investigate the performance of different IQA models with the following annotation numbers  $M = \{1, 2, 4, 8\}$ . When  $M$  increases, the LC-MOS keeps approaching LA-MOS, which also reduces the impact of subjective bias. Similarly, in Table IV, we can find that all IQA models' performance increase as the annotation number goes up. Meanwhile, the superiority of the proposed GDBC keeps shrinking. However, our method still achieves positive  $\Delta(\%)$  across all annotation numbers in this investigation, which verifies the robustness of GDBC to the variations in annotation numbers. In addition, we

TABLE III  
PERFORMANCE OF DIFFERENT METHODS WITH DIFFERENT BIAS RATES  $\eta$ . “-G” INDICATES LC-MOS WITH GDBC.

Methods	$\eta = 100\%$				$\eta = 80\%$				$\eta = 60\%$				$\eta = 40\%$				$\eta = 0\%$			
	$\uparrow$ SRCC	$\uparrow$ PLCC	$\uparrow$ KRCC	$\downarrow$ RMSE	$\uparrow$ SRCC	$\uparrow$ PLCC	$\uparrow$ KRCC	$\downarrow$ RMSE	$\uparrow$ SRCC	$\uparrow$ PLCC	$\uparrow$ KRCC	$\downarrow$ RMSE	$\uparrow$ SRCC	$\uparrow$ PLCC	$\uparrow$ KRCC	$\downarrow$ RMSE	$\uparrow$ SRCC	$\uparrow$ PLCC	$\uparrow$ KRCC	$\downarrow$ RMSE
RET	0.7905	0.8361	0.5990	0.0775	0.8152	0.8572	0.6264	0.0726	0.8385	0.8762	0.6536	0.0671	0.8594	0.8946	0.6755	0.0620	0.8920	0.9165	0.7213	0.0565
RET-G	0.8294	0.8659	0.6425	0.0703	0.841	0.8778	0.6557	0.0676	0.8593	0.8899	0.6785	0.0642	0.8719	0.8994	0.6924	0.0617	0.8916	0.9139	0.7208	0.0574
$\Delta(\%)$	<b>4.9209</b>	<b>3.5642</b>	<b>7.2621</b>	<b>-9.3674</b>	<b>3.1649</b>	<b>2.4032</b>	<b>4.6775</b>	<b>-6.9124</b>	<b>2.4806</b>	<b>1.5636</b>	<b>3.8097</b>	<b>-4.3670</b>	<b>1.4545</b>	<b>0.5366</b>	<b>2.5019</b>	<b>-0.5501</b>	<b>-0.0413</b>	<b>-0.2826</b>	<b>-0.0668</b>	<b>1.6519</b>
NIA	0.7551	0.7363	0.5626	0.1032	0.7914	0.8301	0.5993	0.0854	0.8099	0.8428	0.6185	0.0726	0.8470	0.8767	0.6617	0.0668	0.8879	0.9108	0.7154	0.0577
NIA-G	0.8216	0.8519	0.6357	0.0736	0.8268	0.8645	0.6396	0.0684	0.8376	0.8688	0.6517	0.0757	0.8605	0.889	0.6792	0.0638	0.8829	0.9050	0.7091	0.0625
$\Delta(\%)$	<b>8.8068</b>	<b>15.7001</b>	<b>12.9932</b>	<b>-28.6645</b>	<b>4.4682</b>	<b>4.1444</b>	<b>6.7335</b>	<b>-19.8378</b>	<b>3.4202</b>	<b>3.085</b>	<b>5.3678</b>	<b>4.3253</b>	<b>1.5939</b>	<b>1.403</b>	<b>2.6447</b>	<b>-4.5333</b>	<b>-0.5684</b>	<b>-0.6345</b>	<b>-0.8815</b>	<b>8.1802</b>
DNN	0.7087	0.7474	0.5189	0.1015	0.7710	0.8191	0.5788	0.0880	0.7954	0.8334	0.6033	0.0786	0.8229	0.8553	0.6320	0.0730	0.8833	0.9050	0.7074	0.0589
DNN-G	0.8053	0.8424	0.6137	0.0761	0.8318	0.8655	0.6431	0.0696	0.8363	0.8676	0.6491	0.0690	0.8514	0.8781	0.6663	0.0667	0.8832	0.9021	0.7065	0.0606
$\Delta(\%)$	<b>13.6306</b>	<b>12.7107</b>	<b>18.2694</b>	<b>-24.9923</b>	<b>7.8859</b>	<b>5.6648</b>	<b>11.1092</b>	<b>-20.9905</b>	<b>5.1421</b>	<b>4.1037</b>	<b>7.5916</b>	<b>-12.1487</b>	<b>3.4634</b>	<b>2.6657</b>	<b>5.4272</b>	<b>-8.6287</b>	<b>-0.0111</b>	<b>-0.3241</b>	<b>-0.1238</b>	<b>2.7439</b>
HET	0.6709	0.7304	0.4868	0.1013	0.7740	0.8107	0.5823	0.0836	0.8014	0.8448	0.6118	0.0714	0.8409	0.8752	0.6532	0.0669	0.8639	0.8771	0.6888	0.0669
HET-G	0.8012	0.8452	0.6109	0.0768	0.8345	0.8740	0.6474	0.0698	0.8425	0.8788	0.6563	0.0685	0.8536	0.8877	0.6717	0.0675	0.8566	0.8329	0.6769	0.0781
$\Delta(\%)$	<b>19.4217</b>	<b>15.7174</b>	<b>25.4930</b>	<b>-24.2490</b>	<b>7.8165</b>	<b>7.8081</b>	<b>11.1798</b>	<b>-16.4712</b>	<b>5.1285</b>	<b>4.0246</b>	<b>7.2736</b>	<b>-3.9628</b>	<b>1.5103</b>	<b>1.4282</b>	<b>2.8322</b>	<b>0.9524</b>	<b>-0.8412</b>	<b>-5.0372</b>	<b>-1.7323</b>	<b>16.6841</b>
TIQ	0.6617	0.7313	0.4761	0.1155	0.7375	0.7835	0.5469	0.0942	0.8046	0.8477	0.6146	0.0765	0.8159	0.8597	0.6261	0.0715	0.9052	0.9286	0.7375	0.0513
TIQ-G	0.8100	0.8319	0.6213	0.0788	0.8121	0.8528	0.6229	0.0724	0.8385	0.8807	0.6532	0.0655	0.8537	0.8918	0.6719	0.0628	0.9030	0.9262	0.7355	0.0528
$\Delta(\%)$	<b>22.4045</b>	<b>13.7693</b>	<b>30.5076</b>	<b>-31.7999</b>	<b>10.1147</b>	<b>8.8514</b>	<b>13.8950</b>	<b>-23.1412</b>	<b>4.2122</b>	<b>3.8873</b>	<b>6.2717</b>	<b>-14.3703</b>	<b>4.6315</b>	<b>3.7282</b>	<b>7.3153</b>	<b>-12.1345</b>	<b>-0.2381</b>	<b>-0.2562</b>	<b>-0.2722</b>	<b>2.9547</b>
MQA	0.7432	0.7760	0.5513	0.0926	0.7833	0.8186	0.5933	0.0800	0.8279	0.8628	0.6390	0.0704	0.8502	0.8871	0.6675	0.0645	0.9032	0.9220	0.7328	0.0555
MQA-G	0.8023	0.8458	0.6130	0.0745	0.8309	0.8657	0.6416	0.0721	0.8460	0.8806	0.6605	0.0699	0.8633	0.8935	0.6835	0.0677	0.8932	0.9124	0.7197	0.0631
$\Delta(\%)$	<b>7.9408</b>	<b>8.9999</b>	<b>11.1907</b>	<b>-19.5506</b>	<b>6.0831</b>	<b>5.7516</b>	<b>8.1382</b>	<b>-9.8041</b>	<b>2.1875</b>	<b>2.0672</b>	<b>3.3733</b>	<b>-0.7794</b>	<b>1.5451</b>	<b>0.7205</b>	<b>2.4044</b>	<b>5.0016</b>	<b>-1.1051</b>	<b>-1.0410</b>	<b>-1.7986</b>	<b>13.6888</b>
DQT	0.7129	0.7374	0.5234	0.1027	0.7596	0.8053	0.5684	0.0846	0.8073	0.8497	0.6188	0.0732	0.8337	0.8745	0.6484	0.0672	0.8958	0.9196	0.7245	0.0549
DQT-G	0.7993	0.8389	0.6062	0.0753	0.8109	0.8570	0.6222	0.0718	0.8276	0.8689	0.6401	0.0695	0.8436	0.8829	0.6591	0.0662	0.8897	0.9142	0.7166	0.0578
$\Delta(\%)$	<b>12.1185</b>	<b>13.7675</b>	<b>15.8345</b>	<b>-26.6781</b>	<b>6.7556</b>	<b>6.4206</b>	<b>9.4587</b>	<b>-15.1113</b>	<b>2.5252</b>	<b>2.2595</b>	<b>3.4537</b>	<b>-5.1256</b>	<b>1.1923</b>	<b>0.9634</b>	<b>1.6454</b>	<b>-1.3938</b>	<b>-0.6853</b>	<b>-0.5827</b>	<b>-1.0914</b>	<b>5.2344</b>
MIQ	0.8024	0.8194	0.6094	0.0845	0.7951	0.8340	0.5997	0.0841	0.8206	0.8552	0.6294	0.0778	0.8400	0.8803	0.6516	0.0682	0.9274	0.9414	0.7689	0.0486
MIQ-G	0.8578	0.8921	0.6728	0.0643	0.8475	0.8784	0.6575	0.0672	0.8671	0.8947	0.6833	0.0628	0.8823	0.9110	0.7035	0.0575	0.9210	0.9344	0.7578	0.0533
$\Delta(\%)$	<b>6.9046</b>	<b>8.8680</b>	<b>10.4092</b>	<b>-23.9063</b>	<b>6.5855</b>	<b>5.3336</b>	<b>9.6310</b>	<b>-20.1514</b>	<b>5.6613</b>	<b>4.6203</b>	<b>8.5626</b>	<b>-19.2190</b>	<b>5.0342</b>	<b>3.4842</b>	<b>7.9634</b>	<b>-15.6801</b>	<b>-0.6927</b>	<b>-0.7510</b>	<b>-1.4433</b>	<b>9.7266</b>

also calculate the model's performance  $f_{\theta^*}$  compared to the original LA-MOS training when the subjects are limited  $f_{\theta^M}$ , where  $M$  is the annotation number. We verify the effectiveness of our method, we calculate the Percentage change in SRCC (PS) as  $\frac{(f_{\theta^M} - f_{\theta^*})}{f_{\theta^*}} \times 100$ . As shown in Fig. 4, most models can reach 90% even with minimal annotation. In the best case, most models can achieve reliability results above 96% of the original LA-MOS.

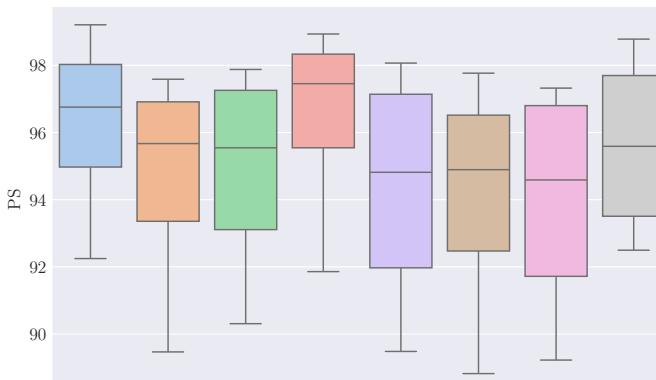


Fig. 4. The relative SRCC performance between  $f_{\theta^M}$  and  $f_{\theta^*}$ .

#### D. Parameter analysis

In this section, we further investigate the impact of the parameters  $\alpha$  and  $\epsilon$  in Eq. 16, which control the updating intensity and frequency of  $\mu_{zi}$  for our GDBC. More specifically, a larger  $\alpha$  would reduce the updating intensity and tend to keep the  $\mu_{zi}$  unchanged in each iteration. Similarly, a larger  $\epsilon$  would reduce the updating frequency of  $\mu_{zi}$  in the whole training process and vice versa. In addition, we further investigate the impact of the neighboring number  $t_h$ . All experiments are conducted on the KONIQ database under the most challenging setting, i.e.,  $\eta = 100\%$  and  $M = 1$ .

Referring to Eq. 15, we can infer that  $\alpha$  ranges from 0 to 1. Without loss of generality, we investigate the performance variation of GDBC under seven different  $\alpha$  values, i.e.,  $\{0.0, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ . As shown in Fig. 5, the performance of all IQA models undergo a process of up and down when  $\alpha$  gradually grows from 0 to 1. Actually, too large  $\alpha$  would reduce the updating intensity and result in underadjustment toward  $\mu_{zi}$ . By contrast, too small  $\alpha$  would increase the updating intensity and may result in overadjustment for  $\mu_{zi}$ . Although different IQA models prefer different  $\alpha$  settings, their performance changes become very slight when  $\alpha$  ranges from 0.5 to 0.9. In view of the average performance of all IQA models, we experimentally set  $\alpha$  to 0.9 in the proposed method.

We also investigate the performance variations of GDBC under different  $\epsilon$  values, i.e.,  $\{0.001, 0.01, 0.1, 1.0\}$ . As shown

TABLE IV  
PERFORMANCES OF DIFFERENT METHODS WITH DIFFERENT ANNOTATION NUMBERS  $M$ . “-G” INDICATES LC-MOS WITH GDBC.

Methods	$M = 1$				$M = 2$				$M = 4$				$M = 8$			
	$\uparrow$ SRCC	$\uparrow$ PLCC	$\uparrow$ KRCC	$\downarrow$ RMSE	$\uparrow$ SRCC	$\uparrow$ PLCC	$\uparrow$ KRCC	$\downarrow$ RMSE	$\uparrow$ SRCC	$\uparrow$ PLCC	$\uparrow$ KRCC	$\downarrow$ RMSE	$\uparrow$ SRCC	$\uparrow$ PLCC	$\uparrow$ KRCC	$\downarrow$ RMSE
RET	0.7905	0.8361	0.5990	0.0775	0.8445	0.8784	0.6580	0.0784	0.8654	0.8980	0.6854	0.0625	0.8779	0.9063	0.7014	0.0586
RET-G	0.8294	0.8659	0.6425	0.0703	0.8543	0.8869	0.6706	0.0650	0.8738	0.9022	0.6948	0.0611	0.8809	0.9074	0.7051	0.0588
$\Delta(\%)$	<b>4.9209</b>	<b>3.5642</b>	<b>7.2621</b>	<b>-9.3674</b>	<b>1.1604</b>	<b>0.9677</b>	<b>1.9149</b>	<b>-17.1119</b>	<b>0.9706</b>	<b>0.4677</b>	<b>1.3715</b>	<b>-2.1760</b>	<b>0.3417</b>	<b>0.1214</b>	<b>0.5275</b>	<b>0.3370</b>
NIA	0.7551	0.7363	0.5626	0.1032	0.8445	0.8795	0.6593	0.0720	0.8744	0.9012	0.6951	0.0656	0.8843	0.9062	0.7091	0.0632
NIA-G	0.8216	0.8519	0.6357	0.0736	0.8606	0.8915	0.6789	0.0690	0.8797	0.9038	0.7021	0.0665	0.8869	0.9064	0.7130	0.0653
$\Delta(\%)$	<b>8.8068</b>	<b>15.7001</b>	<b>12.9932</b>	<b>-28.6645</b>	<b>1.9065</b>	<b>1.3644</b>	<b>2.9728</b>	<b>-4.1580</b>	<b>0.6061</b>	<b>0.2885</b>	<b>1.0070</b>	<b>1.3638</b>	<b>0.2940</b>	<b>0.0221</b>	<b>0.5500</b>	<b>3.4114</b>
DNN	0.7087	0.7474	0.5189	0.1015	0.8038	0.8436	0.6135	0.0799	0.8453	0.8751	0.6574	0.0673	0.8577	0.8857	0.6740	0.0631
DNN-G	0.8053	0.8424	0.6137	0.0761	0.8418	0.8728	0.6549	0.0696	0.8631	0.8877	0.6800	0.0644	0.8707	0.8927	0.6902	0.0627
$\Delta(\%)$	<b>13.6306</b>	<b>12.7107</b>	<b>18.2694</b>	<b>-24.9923</b>	<b>4.7275</b>	<b>3.4614</b>	<b>6.7482</b>	<b>-12.8648</b>	<b>2.1058</b>	<b>1.4398</b>	<b>3.4378</b>	<b>-4.2233</b>	<b>1.5157</b>	<b>0.7903</b>	<b>2.4036</b>	<b>-0.6724</b>
HET	0.6709	0.7304	0.4868	0.0984	0.7322	0.789	0.5452	0.0696	0.8114	0.8569	0.6209	0.0644	0.8421	0.8825	0.6554	0.0618
HET-G	0.8012	0.8452	0.6109	0.0749	0.8275	0.867	0.6399	0.0701	0.8342	0.8535	0.6485	0.0679	0.8567	0.8845	0.6733	0.0670
$\Delta(\%)$	<b>19.4217</b>	<b>15.7174</b>	<b>25.4930</b>	<b>-23.8446</b>	<b>13.0156</b>	<b>9.8859</b>	<b>17.3698</b>	<b>0.6127</b>	<b>2.8100</b>	<b>-0.3968</b>	<b>4.4452</b>	<b>5.4723</b>	<b>1.7334</b>	<b>0.2238</b>	<b>2.7269</b>	<b>8.4043</b>
TIQ	0.6617	0.7313	0.4761	0.1155	0.7584	0.8223	0.5667	0.0829	0.8608	0.8953	0.6770	0.0620	0.8666	0.9018	0.6864	0.0598
TIQ-G	0.8100	0.8319	0.6213	0.0788	0.8400	0.8832	0.6568	0.0649	0.8765	0.9059	0.6979	0.0589	0.8877	0.9154	0.7142	0.0562
$\Delta(\%)$	<b>22.4045</b>	<b>13.7693</b>	<b>30.5076</b>	<b>-31.7999</b>	<b>10.7594</b>	<b>7.4055</b>	<b>15.8911</b>	<b>-21.6994</b>	<b>1.8288</b>	<b>1.1840</b>	<b>3.0773</b>	<b>-5.0036</b>	<b>2.4356</b>	<b>1.4986</b>	<b>4.0498</b>	<b>-5.9991</b>
MQA	0.7432	0.7760	0.5513	0.0926	0.8125	0.8604	0.6250	0.0707	0.8602	0.8923	0.6753	0.0629	0.8766	0.9034	0.6986	0.0606
MQA-G	0.8023	0.8458	0.6130	0.0745	0.8462	0.8841	0.6641	0.0699	0.8680	0.8974	0.6860	0.0659	0.8831	0.9039	0.7062	0.0648
$\Delta(\%)$	<b>7.9408</b>	<b>8.9999</b>	<b>11.1907</b>	<b>-19.5506</b>	<b>4.1482</b>	<b>2.7517</b>	<b>6.2562</b>	<b>-1.1643</b>	<b>0.9106</b>	<b>0.5724</b>	<b>1.5769</b>	<b>4.6561</b>	<b>0.7416</b>	<b>0.0567</b>	<b>1.0946</b>	<b>6.8662</b>
DQT	0.7129	0.7374	0.5234	0.1027	0.8047	0.8505	0.6149	0.0728	0.8551	0.8877	0.6691	0.0638	0.8679	0.8988	0.6898	0.0608
DQT-G	0.7993	0.8389	0.6062	0.0753	0.8291	0.8719	0.6426	0.0691	0.8656	0.8948	0.6830	0.0631	0.8718	0.8994	0.6928	0.0620
$\Delta(\%)$	<b>12.1185</b>	<b>13.7675</b>	<b>15.8345</b>	<b>-26.6781</b>	<b>3.0237</b>	<b>2.5136</b>	<b>4.5122</b>	<b>-5.1369</b>	<b>1.2286</b>	<b>0.7966</b>	<b>2.0741</b>	<b>-1.0286</b>	<b>0.4473</b>	<b>0.0673</b>	<b>0.4374</b>	<b>1.8877</b>
MIQ	0.8024	0.8194	0.6094	0.0845	0.8332	0.8711	0.6438	0.0715	0.8775	0.9038	0.6942	0.0608	0.9047	0.9254	0.7324	0.0530
MIQ-G	0.8578	0.8921	0.6728	0.0643	0.8703	0.9018	0.6885	0.0603	0.9027	0.9235	0.7292	0.0535	0.9161	0.9331	0.7502	0.0505
$\Delta(\%)$	<b>6.9046</b>	<b>8.8680</b>	<b>10.4092</b>	<b>-23.9063</b>	<b>4.4594</b>	<b>3.5248</b>	<b>6.9317</b>	<b>-15.6247</b>	<b>2.8752</b>	<b>2.1756</b>	<b>5.0524</b>	<b>-12.0602</b>	<b>1.2649</b>	<b>0.8252</b>	<b>2.4315</b>	<b>-4.6477</b>

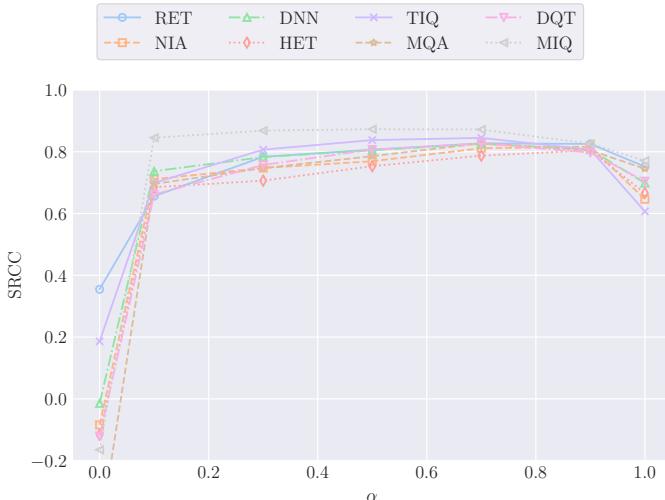


Fig. 5. The performance variation of GDBC with different  $\alpha$  on various IQA models.

in Fig. 6, when  $\epsilon$  grows from 0.001 to 0.1, the performance of GDBC are insensitive to different threshold settings, whose curves are very close to each other across all IQA models and the SRCC fluctuations are within 0.02. However, when

$\epsilon$  reaches 1, the performance of GDBC significantly drops across all IQA models. As mentioned above, too large  $\epsilon$  tends to disable the updating of  $\mu_{z_i}$  and degrade GDBC to a plain model, which directly learns from the noisy LC-MOS and suffers their misleading. Based on these investigations, we experimentally set  $\epsilon$  to 0.01 in our GDBC.

In addition, we also added parameter analysis for  $t_h$  across all IQA models on VCL dataset. As shown in Eq. 16, the  $t_h$  controls the updating latency, where the GDBC only activates after the training iterations reach  $t_h$ . That is, a higher  $t_h$  would postpone the dual-bias calibration and may result in small  $c_i^t$  due to the overfitting toward  $y_i^n$ , which reversely suppress the parameter updating. As shown in Fig. 7, when  $t_h$  is smaller than 5, the performance variations of all IQA models are very limited, whose RMSE fluctuations are within 0.04. However, when  $t_h$  reaches 11, the RMSE of all IQA models would significantly increase. Therefore,  $t_h$  can't be set to a very high number.

#### E. Comparison with separate subjective bias calibration and model bias calibration methods

Besides the effectiveness and robustness validation, we further compare our alternating optimization-based dual-bias calibration model with some separate subjective bias calibration and model bias calibration methods. On the one hand, many

TABLE V

PERFORMANCE COMPARISON IN TERMS OF SRCC BETWEEN THE PROPOSED METHOD AND ITS VARIANTS. BEST RESULTS IN EACH SESSION ARE HIGHLIGHTED IN BOLD, WHILE RESULTS OF FUTURE TASKS ARE MARKED IN GREY. “-A”, “-C”, “-G” INDICATES LA-MOS, LC-MOS AND LC-MOS WITH GDBC RESPECTIVELY.

Sequence	Dataset	Method	KONIQ	LIVEC	VCL	CSIQ
1 <sup>st</sup>	KONIQ	SLF-A	0.8863	0.7950	0.5992	0.5954
		SLF-C	0.6994	0.5944	0.4263	0.6947
		SLF-G	<b>0.7956</b>	0.6552	0.4898	0.5347
		CQA-A	0.8883	0.8021	0.6072	0.5637
		CQA-C	0.7729	0.6497	0.4482	0.4089
		CQA-G	<b>0.8284</b>	0.6780	0.5119	0.5751
2 <sup>nd</sup>	LIVEC	SLF-A	0.8487	0.8341	0.5715	0.4696
		SLF-C	0.5474	0.5965	0.3544	0.5404
		SLF-G	<b>0.7600</b>	<b>0.7008</b>	0.4193	0.4444
		CQA-A	0.8659	0.8145	0.6893	0.5744
		CQA-C	0.7863	0.7613	0.4229	0.3429
		CQA-G	<b>0.8216</b>	<b>0.7393</b>	0.5160	0.5387
3 <sup>rd</sup>	VCL	SLF-A	0.7297	0.7112	0.7401	0.5736
		SLF-C	0.3520	0.3866	0.3574	0.4701
		SLF-G	<b>0.7386</b>	<b>0.7123</b>	<b>0.6209</b>	0.6347
		CQA-A	0.8224	0.7616	0.5612	0.6140
		CQA-C	0.7801	0.7087	0.4712	0.6292
		CQA-G	<b>0.8010</b>	<b>0.7221</b>	<b>0.4799</b>	0.6316
4 <sup>th</sup>	CSIQ	SLF-A	0.7972	0.7687	0.7507	0.5567
		SLF-C	0.2676	0.3170	0.3605	0.4621
		SLF-G	<b>0.7496</b>	<b>0.7025</b>	<b>0.5959</b>	<b>0.6452</b>
		CQA-A	0.7655	0.7095	0.5336	0.6071
		CQA-C	0.7432	0.6703	0.4970	0.6050
		CQA-G	<b>0.7539</b>	<b>0.6846</b>	<b>0.5003</b>	<b>0.6436</b>

TABLE VI

PERFORMANCE COMPARISON OF DIFFERENT BIQA MODELS. ALL METHODS ARE TRAINED IN CHRONOLOGICAL ORDER. “-A”, “-C”, “-G” INDICATES LA-MOS, LC-MOS AND LC-MOS WITH GDBC RESPECTIVELY.

Methods	mSRCC	mPI	mSI	mPSI
SLF-A	0.7183	0.7543	0.9466	0.8504
SLF-C	0.3518	0.5288	0.9437	0.7363
SLF-G	<b>0.6733</b>	<b>0.6906</b>	<b>0.9544</b>	<b>0.8225</b>
CQA-A	0.6539	0.7178	0.9848	0.8513
CQA-C	0.6289	0.6526	0.9761	0.8144
CQA-G	<b>0.6456</b>	<b>0.6728</b>	<b>0.9855</b>	<b>0.8291</b>

standardized label screening models have been developed to recover high-quality MOS from multiple noisy human annotations, which focus on the subjective bias calibration, such as the subject rejection (SR) model in ITU-R BT.500 [10], the maximum likelihood estimation (MLE) model in ITU-T P.910 [11], and the subject bias removal (SBR) model in ITU-T P.913 [12]. In view of the label screening models' request for multiple annotations, we simply the most challenging case of LC-MOS learning to the following setting, i.e.,  $\eta = 100\%$  and  $M = 2$ , and the repetition number of subjective test is

TABLE VII

RUNNING TIME INVESTIGATION OF THE GDBC MODULE. “W” AND “W/O” REPRESENT WITH AND WITHOUT THE GDBC MODULE IN THE TRAINING PHASE RESPECTIVELY. S: SECONDS.

	RET	NIA	DNN	HET	TIQ	MQA	DQT	MIQ
w (s)	0.0602	0.0831	0.1069	0.0522	0.1235	0.1166	0.2096	0.4054
w/o (s)	0.0601	0.0820	0.1062	0.0517	0.1220	0.1159	0.2071	0.4042
T%	0.2366	1.2897	0.6859	0.9996	1.2295	0.6269	1.2046	0.3008

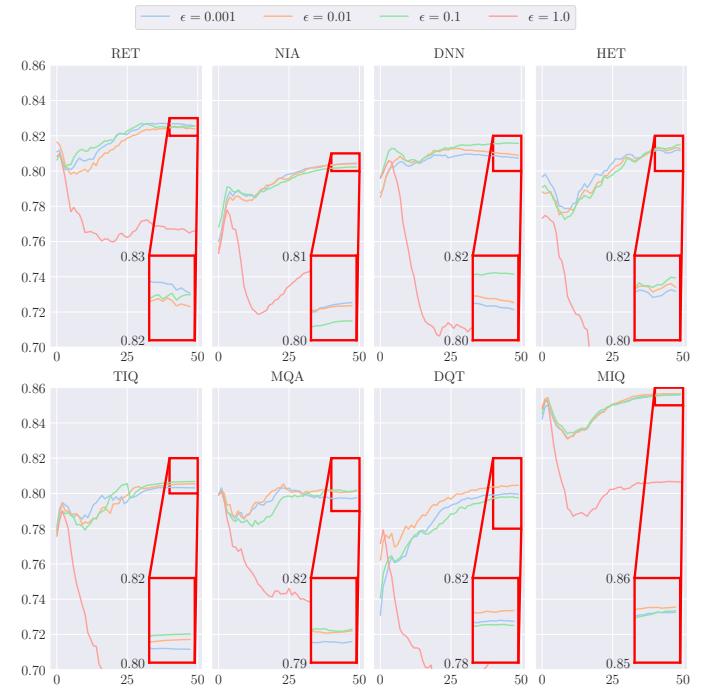


Fig. 6. The performance variation of GDBC with different  $\epsilon$  on various IQA models.

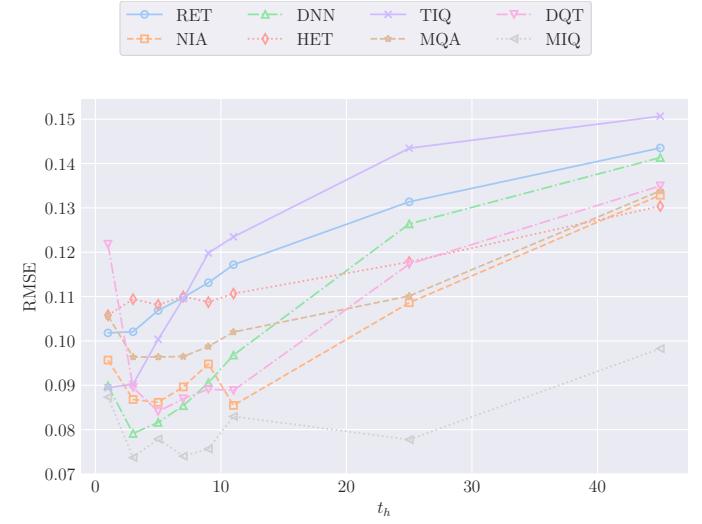


Fig. 7. The performance variation of GDBC with different  $t_h$  on VCL dataset.

set to 1. On the other hand, regarding the subjective bias as the noise, recent noisy label learning models could also be applicable to the model bias calibration in our LC-MOS learning task, such as the generalized cross entropy (GCE) loss [39] and symmetric cross entropy (SCE) loss [43]. It should be noted that these representative noisy label learning methods, such as GCE and SCE, were originally developed for multiple-output-based classification networks, which is hard to apply to most single-output-based regression networks in IQA. For compatibility, we evaluate all bias calibration models on a multiple-output-based IQA model NIA [15]. Meanwhile, in view of the availability of raw human annotations, we conduct all experiments on the VCL database in this section.

As shown in Table VIII, both the label screening and noisy label learning methods deteriorate the performance of NIA when LC-MOS is used for training. Only the proposed GDBC improves the learning of IQA model. Firstly, unlike our iterative optimization strategy, existing label screening methods like SR, MLE, and BR adopt a one-stop post-processing paradigm, which can not identify or even suppress the error in calibrating the human annotations. When the available annotations are very few, the calibration error easily becomes considerable and further interferes with the IQA model learning. Secondly, to avoid overfitting issues, existing noisy label learning methods like GCE and SCE focus on suppressing the gradient backpropagation, which slows down the rate of convergence and easily causes underfitting in turn. These reported results further verify the superiority and necessity of jointly calibrating the subjective bias and model bias.

TABLE VIII

COMPARISON RESULTS WITH THE REPRESENTATIVE LABEL SCREENING AND NOISY LABEL LEARNING METHODS.

Methods	LC-MOS	SR	MLE	SBR	GCE	SCE	GDBC
SRCC	0.9142	0.8976	0.8808	0.9090	0.7788	0.8772	<b>0.9233</b>
PLCC	0.8770	0.8419	0.8805	0.8742	0.7608	0.8587	<b>0.9053</b>
KRCC	0.7311	0.7061	0.6921	0.7303	0.6120	0.7237	<b>0.7500</b>

## V. CONCLUSION

In this paper, we explore a new challenge of learning robust IQA models from noisy Low-Cost MOS (LC-MOS), which requires very few opinion scores for each image. By jointly inferring the subjective bias and model bias, we develop a plug-and-play Gated Dual-Bias Calibration (GDBC) module, which enforces the IQA model learned from LC-MOS to approach the unbiased estimation of Labor-Abundant MOS (LA-MOS). Extensive experiments on popular IQA databases and representative deep IQA models demonstrate the effectiveness of the proposed method. It significantly outperforms the IQA models directly learned from LC-MOS and even achieves comparable performance with respect to the IQA models learned from the expensive and time-consuming LA-MOS. Additionally, we also verify the superiority of the proposed method over the existing label screening and noisy label learning methods.

## REFERENCES

- [1] G. Zhai and X. Min, "Perceptual image quality assessment: a survey," *Science China Information Sciences*, vol. 63, no. 11, pp. 1–52, 2020.
- [2] Y. Deng, C. C. Loy, and X. Tang, "Image aesthetic assessment: An experimental survey," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 80–106, 2017.
- [3] H. Lin, V. Hosu, and D. Saupe, "Kadid-10k: A large-scale artificially distorted iqas database," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019, pp. 1–3.
- [4] D. Kundu, D. Ghadiyaram, A. C. Bovik, and B. L. Evans, "Large-scale crowdsourced study for tone-mapped hdr pictures," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4725–4740, 2017.
- [5] A. Zaric, N. Tatalovic, N. Brajkovic, H. Hlevnjak, M. Loncaric, E. Dumic, and S. Grgic, "Vcl@ fer image quality assessment database," in *Proceedings ELMAR-2011*. IEEE, 2011, pp. 105–110.
- [6] T. Song, L. Li, J. Wu, Y. Yang, Y. Li, Y. Guo, and G. Shi, "Knowledge-guided blind image quality assessment with few training samples," *IEEE Transactions on Multimedia*, 2022.
- [7] L. Wang, Q. Wu, K. N. Ngan, H. Li, F. Meng, and L. Xu, "Blind tone-mapped image quality assessment and enhancement via disentangled representation learning," in *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2020, pp. 1096–1102.
- [8] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," *Majlesi Journal of Electrical Engineering*, vol. 9, no. 1, 2015.
- [9] Z. Wang and A. C. Bovik, "Modern image quality assessment," Ph.D. dissertation, Springer, 2006.
- [10] ITU-R BT.500, "Methodologies for the subjective assessment of the quality of television images," <https://www.itu.int/rec/R-REC-BT.500>.
- [11] ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," <https://www.itu.int/rec/T-REC-P.910>.
- [12] ITU-T P.913, "Methods for the subjective assessment of video quality audio quality and audiovisual quality of internet video and distribution quality television in any environment," <https://www.itu.int/rec/T-REC-P.913>.
- [13] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202–1213, 2017.
- [14] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 36–47, 2018.
- [15] H. Talebi and P. Milanfar, "Nima: Neural image assessment," *IEEE transactions on image processing*, vol. 27, no. 8, pp. 3998–4011, 2018.
- [16] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, and Y. Zhang, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3667–3676.
- [17] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Transactions on Image Processing*, vol. 30, pp. 3474–3486, 2021.
- [18] S. A. Golestan, S. Dadsetan, and K. M. Kitani, "No-reference image quality assessment via transformers, relative ranking, and self-consistency," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1220–1230.
- [19] S. Sun, T. Yu, J. Xu, W. Zhou, and Z. Chen, "Graphiq: Learning distortion graph representations for blind image quality assessment," *IEEE Transactions on Multimedia*, 2022.
- [20] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "Metaiq: Deep meta-learning for no-reference image quality assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 143–14 152.
- [21] J. Ma, J. Wu, L. Li, W. Dong, X. Xie, G. Shi, and W. Lin, "Blind image quality assessment with active inference," *IEEE Transactions on Image Processing*, vol. 30, pp. 3650–3663, 2021.
- [22] W. Zhang, D. Li, C. Ma, G. Zhai, X. Yang, and K. Ma, "Continual learning for blind image quality assessment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [23] W. Zhang, G. Zhai, Y. Wei, X. Yang, and K. Ma, "Blind image quality assessment via vision-language correspondence: A multitask learning perspective," *arXiv preprint arXiv:2303.14968*, 2023.
- [24] W. Zhang, D. Li, X. Min, G. Zhai, G. Guo, X. Yang, and K. Ma, "Perceptual attacks of no-reference image quality models with human-in-the-loop," *arXiv preprint arXiv:2210.00933*, 2022.

- [25] H. Duan, G. Zhai, X. Min, Y. Zhu, Y. Fang, and X. Yang, "Perceptual quality assessment of omnidirectional images," in *2018 IEEE international symposium on circuits and systems (ISCAS)*. IEEE, 2018, pp. 1–5.
- [26] W. Sun, X. Min, G. Zhai, K. Gu, H. Duan, and S. Ma, "Mc360iq: A multi-channel cnn for blind 360-degree image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 64–77, 2019.
- [27] V. K. Adhikarla, M. Vinkler, D. Sumin, R. Mantiuk, K. Myszkowski, H.-P. Seidel, and P. Didyk, "Towards a quality metric for dense light fields," in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [28] X. Min, G. Zhai, K. Gu, X. Yang, and X. Guan, "Objective quality evaluation of dehazed images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 2879–2892, 2018.
- [29] X. Min, G. Zhai, K. Gu, Y. Zhu, J. Zhou, G. Guo, X. Yang, X. Guan, and W. Zhang, "Quality evaluation of image dehazing methods using synthetic hazy images," *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2319–2333, 2019.
- [30] Y. Fang, H. Zhu, Y. Zeng, K. Ma, and Z. Wang, "Perceptual quality assessment of smartphone photography," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3677–3686.
- [31] N. Kawabata and M. Miyao, "Multi-view 3d cg image quality assessment for contrast enhancement based on s-cielab color space," *IEICE TRANSACTIONS on Information and Systems*, vol. 100, no. 7, pp. 1448–1462, 2017.
- [32] N. Kawabata and M. Miyao, "3d cg image quality metrics by regions with 8 viewpoints parallax barrier method," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 98, no. 8, pp. 1696–1708, 2015.
- [33] D. Arpit, S. Jastrzebski, N. Ballas, D. Krueger, E. Bengio, M. S. Kanwal, T. Maharaj, A. Fischer, A. Courville, Y. Bengio et al., "A closer look at memorization in deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 233–242.
- [34] A. Ghosh, H. Kumar, and P. Sastry, "Robust loss functions under label noise for deep neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.
- [35] D. Tanaka, D. Ikami, T. Yamasaki, and K. Aizawa, "Joint optimization framework for learning with noisy labels," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5552–5560.
- [36] J. Goldberger and E. Ben-Reuven, "Training deep neural-networks using a noise adaptation layer," in *International conference on learning representations*, 2022.
- [37] X. Xia, T. Liu, B. Han, M. Gong, J. Yu, G. Niu, and M. Sugiyama, "Sample selection with uncertainty of losses for learning with noisy labels," *arXiv preprint arXiv:2106.00445*, 2021.
- [38] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Communications of the ACM*, vol. 64, no. 3, pp. 107–115, 2021.
- [39] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in neural information processing systems*, vol. 31, 2018.
- [40] X. Zhou, X. Liu, C. Wang, D. Zhai, J. Jiang, and X. Ji, "Learning with noisy labels via sparse regularization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 72–81.
- [41] S. Hahn and H. Choi, "Self-knowledge distillation in natural language processing," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 2019, pp. 423–430.
- [42] A. Gotmare, N. S. Keskar, C. Xiong, and R. Socher, "A closer look at deep learning heuristics: Learning rate restarts, warmup and distillation," in *International Conference on Learning Representations*, 2019.
- [43] Y. Wang, X. Ma, Z. Chen, Y. Luo, J. Yi, and J. Bailey, "Symmetric cross entropy for robust learning with noisy labels," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 322–330.
- [44] H. Zhuang and J. Young, "Leveraging in-batch annotation bias for crowdsourced active learning," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 2015, pp. 243–252.
- [45] C. Hube, B. Fetahu, and U. Gadiraju, "Understanding and mitigating worker biases in the crowdsourced collection of subjective judgments," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.
- [46] G. Li, J. Wang, Y. Zheng, and M. J. Franklin, "Crowdsourced data management: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 9, pp. 2296–2319, 2016.
- [47] T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen, and J. Häkkinen, "Cid2013: A database for evaluating no-reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 390–402, 2014.
- [48] Z. Li, C. G. Bampis, L. Krasula, L. Janowski, and I. Katsavounidis, "A simple model for subject behavior in subjective experiments," *arXiv preprint arXiv:2004.02067*, 2020.
- [49] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.
- [50] L. Yang, H. Li, F. Meng, Q. Wu, and K. N. Ngan, "Task-specific loss for robust instance segmentation with noisy class labels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 1, pp. 213–227, 2023.
- [51] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti et al., "Image database tid2013: Peculiarities, results and perspectives," *Signal processing: Image communication*, vol. 30, pp. 57–77, 2015.
- [52] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372–387, 2015.
- [53] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment," *IEEE Transactions on Image Processing*, vol. 29, pp. 4041–4056, 2020.
- [54] Q. Wu, H. Li, F. Meng, and K. N. Ngan, "A perceptually weighted rank correlation indicator for objective image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2499–2513, 2018.
- [55] Q. Wu, H. Li, K. N. Ngan, and K. Ma, "Blind image quality assessment using local consistency aware retriever and uncertainty aware evaluator," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2078–2089, 2018.
- [56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [57] C. Chen, J. Mo, J. Hou, H. Wu, L. Liao, W. Sun, Q. Yan, and W. Lin, "Topiq: A top-down approach from semantics to distortions for image quality assessment," *IEEE Transactions on Image Processing*, 2024.
- [58] L. Yu, J. Li, F. Pakdaman, M. Ling, and M. Gabbouj, "Mamiqa: No-reference image quality assessment based on multiscale attention mechanism with natural scene statistics," *IEEE Signal Processing Letters*, 2023.
- [59] G. Qin, R. Hu, Y. Liu, X. Zheng, H. Liu, X. Li, and Y. Zhang, "Data-efficient image quality assessment with attention-panel decoder," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 2, 2023, pp. 2091–2100.
- [60] S. Yang, T. Wu, S. Shi, S. Lao, Y. Gong, M. Cao, J. Wang, and Y. Yang, "Maniq: Multi-dimension attention network for no-reference image quality assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1191–1200.
- [61] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*. Springer, 2009, pp. 1–4.
- [62] J. H. Zar, "Spearman rank correlation," *Encyclopedia of biostatistics*, vol. 7, 2005.
- [63] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [64] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of machine learning research*, vol. 13, no. 2, 2012.
- [65] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.
- [66] R. Ma, Q. Wu, K. N. Ngan, H. Li, F. Meng, and L. Xu, "Forgetting to remember: A scalable incremental learning framework for cross-task blind image quality assessment," *IEEE Transactions on Multimedia*, 2023.



**Lei Wang** (Student Member, IEEE) received the B.S. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2017, where he is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering. His research interests include image quality assessment, visual question answer, and adversarial robustness.



**Hongliang Li** (Senior Member, IEEE) received his Ph.D. degree in Electronics and Information Engineering from Xi'an Jiaotong University, China, in 2005. From 2005 to 2006, he joined the visual signal processing and communication laboratory (VSPC) of the Chinese University of Hong Kong (CUHK) as a Research Associate. From 2006 to 2008, he was a Postdoctoral Fellow at the same laboratory in CUHK. He is currently a Professor in the School of Information and Communication Engineering, University of Electronic Science and Technology of

China. His research interests include image and video processing, visual attention, object detection and segmentation, object recognition and parsing, multimedia content analysis, deep learning.

Dr. Li has authored or co-authored numerous technical articles in well-known international journals and conferences. He is a co-editor of a Springer book titled "Video segmentation and its applications". Dr. Li is involved in many professional activities. He received the 2019 and 2020 Best Associate Editor Awards for IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), and the 2021 Best Editor Award for Journal on Visual Communication and Image Representation. He served as a Technical Program Chair for VCIP 2016 and PCM 2017, General Chairs for ISPACS 2017 and ISPACS 2010, a Publicity Chair for IEEE VCIP 2013, a Local Chair for the IEEE ICME 2014, Area Chairs for VCIP 2022 and 2021, and a Reviewer committee member for IEEE ISCAS from 2018 to 2022. He served as an Associate Editor of IEEE Transactions on Circuits and Systems for Video Technology (2018-2021). He is now an Associate Editor of Journal on Visual Communication and Image Representation, IEEE Open Journal of Circuits and Systems, and an Area Editor of Signal Processing: Image Communication (Elsevier Science). He is selected as the IEEE Circuits and Systems Society Distinguished Lecturer for 2022-2023.



**Qingbo Wu** (Member, IEEE) received the Ph.D. degree in signal and information processing from the University of Electronic Science and Technology of China in 2015. From February 2014 to May 2014, he was Research Assistant with the Image and Video Processing (IVP) Laboratory, Chinese University of Hong Kong. From October 2014 to October 2015, he served as a Visiting Scholar with the Image and Vision Computing (IVC) Laboratory, University of Waterloo. He is currently a Professor with the School of Information and Communication

Engineering, University of Electronic Science and Technology of China. His research interests include image/video coding, quality evaluation, perceptual modeling and processing. He has served as Area Chair for ACM MM 2024, VCIP 2016, Session Chair for ACM MM 2021, ICMCT 2022, TPC/PC member of AAAI 2021-2023, APSIPA ASC 2020-2021, CICAI 2021-2023. He was also a Guest Editor of Remote Sensing and Frontiers in Neuroscience.



**Desen Yuan** (Member, IEEE) received the M.S. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2023. His research interests include image quality assessment, visual question answering, and adversarial robustness. He has served as Session Chair of IEEE SMC and as a reviewer for several prestigious conferences and journals, including ICRA, ACM MM, ICASSP, and IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), among others.



**Fanman Meng** (Member, IEEE) received the Ph.D. degree in signal and information processing from the University of Electronic Science and Technology of China, Chengdu, China, in 2014. From 2013 to 2014, he was a Research Assistant with the Division of Visual and Interactive Computing, Nanyang Technological University, Singapore. He is currently a Professor with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. He has authored or coauthored numerous technical articles in well-known international journals and conferences. His research interests include image segmentation and object detection.



**King Ngi Ngan** (Life Fellow, IEEE) received the Ph.D. degree in electrical engineering from Loughborough University, Loughborough, U.K. He is currently a Distinguished Professor at the University of Electronic Science and Technology of China in Chengdu, China, and an Emeritus Professor at the Chinese University of Hong Kong, Hong Kong. He was previously a Full Professor at the Nanyang Technological University, Singapore, and the University of Western Australia, Perth, Australia. He holds honorary and visiting professorships with numerous universities in China, Australia, and South East Asia. He has published extensively, including 3 authored books, 7 edited volumes, over 450 refereed technical papers, and has edited 9 special issues in journals. He holds 15 patents in image or video coding and communications.

Dr. Ngan served as an Associate Editor of the IEEE Transactions on Circuits and Systems for Video Technology, the Journal on Visual Communications and Image Representation, the EURASIP Journal of Signal Processing: Image Communication, and the Journal of Applied Signal Processing. He chaired a number of prestigious international conferences on video signal processing and communications, and served on the advisory and technical committees of numerous professional organizations. He co-chaired the 2010 IEEE International Conference on Image Processing in Hong Kong. He was a Fellow of IET, U.K., and IEAust, Australia, and an IEEE Distinguished Lecturer from 2006 to 2007.



**Linfeng Xu** (Member, IEEE) received the Ph.D. degree in signal and information processing from the School of Electronic Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2014. From 2014 to 2015, he was with the Ubiquitous Multimedia Laboratory, The State University of New York at Buffalo, Buffalo, NY, USA, as a Visiting Scholar. He is currently an Associate Professor with the School of Information and Communication Engineering, UESTC. His research interests include machine learning, visual attention, image and video coding, visual signal processing, and multimedia communication system. He was a Local Arrangement Chair for ISPACS 2010 and VCIP 2016.