This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3073410, IEEE Transactions on Circuits and Systems for Video Technology

Generalizable No-Reference Image Quality Assessment via Deep Meta-learning

Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi, *Fellow, IEEE*

***Abstract*—Recently, researchers have shown great interest in using convolutional neural networks (CNNs) for no-reference image quality assessment (NR-IQA). Due to the lack of big training data, the efforts of existing metrics in optimizing CNN-based NR-IQA models remain limited. Furthermore, the diversity of distortions in images result in the generalization problem of NR-IQA models when trained with known distortions and tested on unseen distortions, which is an easy task for human. Hence, we propose a NR-IQA metric via deep meta-learning, which is highly generalizable in the face of unseen distortions. The fundamental idea is to learn the meta-knowledge shared by human when evaluating the quality of images with diversified distortions. Specifically, we define NR-IQA of different distortions as a series of tasks and propose a task selection strategy to build two task sets, which are characterized by synthetic to synthetic and synthetic to authentic distortions, respectively. Based on these two task sets, an optimization-based meta-learning is proposed to learn the generalized NR-IQA model, which can be directly used to evaluate the quality of images with unseen distortions. Extensive experiments demonstrate that our NR-IQA metric outperforms the state-of-the-arts in terms of both evaluation performance and generalization ability.**

***Index Terms*—No-reference image quality assessment, generalization ability, optimization-based meta-learning, convolutional neural networks**

I. Introduction

In the past few years, the popularity of smart phones and mobile internet has produced a huge demand for images. The quality of images is closely related to people's visual experience, and it is one of the crucial factors that determine the usefulness of image information received by humans. However, images may be polluted in any stage of its life cycle, including acquisition, compression, storage, and transmission. Hence, it is necessary to quantitatively measure the degree of image distortion, which has critical applications in image processing systems and computer vision tasks. Although subjective image quality assessment (IQA) can obtain accurate and reliable results, it is expensive for people to directly rate image quality and hard to be embedded in real-time systems. Therefore, objective image quality assessment metrics [1] that can automatically evaluate image quality has been a long-lasting research topic in the image processing community, which has widespread applications in image compression [2], image restoration [3], [4], virtual reality [5], fingerprint recognition [6], and image quality monitoring systems [7], etc.

According to the availability of a perfect-quality reference image, the current IQA metrics can be divided into three categories: full-reference IQA (FR-IQA) [8], reduced-reference IQA (RR-IQA) [9], and no-reference IQA (NR-IQA) [10]. While the performances of FR-IQA and RR-IQA metrics are remarkable, reference images are usually hard to obtain in many application scenarios. Hence, NR-IQA has become the research focus in the IQA community [11]. Nonetheless, unavailable reference information leads to enormous challenges for NR-IQA metrics. Therefore, previous NR-IQA metrics are primarily for specific distortion types, such as ringing effects [12], blur [13], and blocking artifacts [14]. The precondition of these metrics is that images are distorted by only one specific type. However, the distortion types of images are often diverse and unknown in actual situations. Hence, the general-purpose NR-IQA metric has attracted growing attention from recent researchers [15], [16], [17], [18], [19], [20], [21]. These metrics establish a NR-IQA model by using hand-crafted [16] or learned [19] features to quantify the general rules of image distortions.

The powerful learning ability of deep learning [22] drives a large number of NR-IQA metrics based on deep convolutional neural networks (DCNNs) [23], [24], [25], [26], [27], [28], [29], [30], [31]. As expected, these metrics are superior to the traditional hand-crafted feature-based NR-IQA metrics [16], [17], [18]. This is because that DCNNs have a huge number of learnable parameters, which can learn the mapping relationship between images and quality scores more effectively. Therefore, the effective learning of DCNNs requires big labeled data. Although developing a large-scale IQA database [32], [33] is an effective way to improve the performance of learned deep models, it is often impossible to obtain a large amount of annotated data when facing a target NR-IQA task in practical applications [11]. As a result, how to make full use of the existing small scale and medium scale annotated IQA databases [34], [35] to learn a more generalizable IQA model has raised increasing concern [10], [27], [28], [29], [30], [31]. To this end, a straightforward approach [36] is to take advantage

This work was supported by the National Natural Science Foundation of China under Grants 61771473, 61991451 and 61379143, the Fundamental Research Funds for the Central Universities under Grant JBF2119 and 2021QN1071, the Key Project of Shaanxi Provincial Department of Education under Grant 20JY024, the Science and Technology Plan of Xi'an under Grant 20191122015KYPT011JC013, the Natural Science Foundation of Jiangsu Province under Grant BK20181354 and BK20200649, and the Six Talent Peaks High-level Talents in Jiangsu Province under Grant XYDXX-063 (*Corresponding author: Leida Li*).

H. Zhu is with the School of Artificial Intelligence, Xidian University, Xi'an 710071, China, and also with the School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China (e-mail: zhuhancheng@cumt.edu.cn).

L. Li is with the School of Artificial Intelligence and Guangzhou Institute of Technology, Xidian University, Xi'an 710071, China, and also with the Pazhou Lab, Guangzhou 510335, China (email: ldli@xidian.edu.cn).

J. Wu, W. Dong and G. Shi are with the School of Artificial Intelligence, Xidian University, Xi'an 710071, China (emails: jinjian.wu@mail.xidian.edu.cn; wsdong@mail.xidian.edu.cn; gmshi@xidian.edu.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3073410, IEEE Transactions on Circuits and Systems for Video Technology

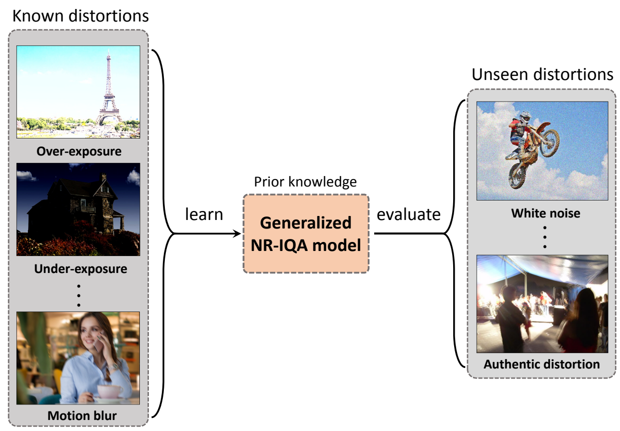


Fig. 1. An illustration of our motivation. Human have highly generalizable prior knowledge and can easily evaluate image quality without knowing the distortion type. Therefore, a generalized NR-IQA model is expected to learn prior knowledge from images with known distortions and can directly evaluate the quality of images with unseen distortions. The generalized NR-IQA model can directly evaluate the quality of images with distortions that are not encountered during model training, without doing any fine-tuning.

of the model pre-trained on ImageNet [37]. Although this approach can reduce the over-fitting of the trained NR-IQA model to some extent, the generalization ability is still quite limited. This is not hard to understand, because ImageNet is designed for image classification tasks, which are significantly different from IQA. Another common practice is to perform data augmentation by dividing an input image into patches for training deep NR-IQA model [23]. This is problematic in that any change in image size may affect its original quality. Besides, some other works propose to generate large-scale ranking data [26], [25], or hallucinated reference images [27] for learning DCNN-based NR-IQA models. These approaches are effective in learning deep NR-IQA models through data enhancement with known distortion types, but fail to deal with the quality evaluation of images with diversified (usually unseen) distortions. Therefore, the current deep IQA models are typically subject to the generalization problem, which could impede their applicability in real-world scenarios.

People have highly generalizable prior knowledge and can easily evaluate image quality without knowing the exact distortion type. Therefore, a generalized NR-IQA model should learn the prior knowledge from images with known distortions and can directly evaluate the quality of images with unseen distortions, as illustrated in Fig. 1. To this end, we propose a new NR-IQA metric via deep meta-learning [38], [39], which can learn the generalization ability of human in extracting the underlying rules across diversified distortions. In contrast to the existing metrics, the proposed NR-IQA model can learn the shared meta-knowledge of people when evaluating images with diversified distortions, directly handling the NR-IQA task of unseen distortions without any model updating or additional fine-tuning.

This paper extends our previous CVPR paper [10], which is named Meta-learning based Image Quality Assessment (MetaIQA). The extensions are multiple folds. First and foremost, MetaIQA requires an additional fine-tuning stage using the target domain data when dealing with new distortions. By contrast, the proposed metric can evaluate new unseen distortions without doing any model updating or fine-tuning, which is achieved by significantly improving the meta-optimization process. Second, we propose a task selection strategy across distortions to reduce the redundant information of training samples and improve the generalization performance of our NR-IQA model. Third, MetaIQA needs to change images to a fixed size for adapting to the input of deep models, which leads to the quality change of input images. To deal with this problem, we propose a deep regression network by introducing a spatial pyramid pooling (SPP) [40] module that does not require a fixed input image size, which further improves the generalization ability of the trained NR-IQA model. Finally, we perform new Leave-One-Distortion-Out cross-validation on synthetically distorted IQA databases that can eliminate the influence of image content for training and testing. Besides, we provide a more in-depth analysis of the proposed metric on authentically distorted IQA databases. Visual analysis is further conducted to illustrate the generalization ability of the proposed model, which is not provided in the original work.

Therefore, we term the proposed model "MetaIQA+". Compared to the original MetaIQA, the proposed metric has a significant improvement in terms of both prediction accuracy and generalization ability. In summary, the contributions of the proposed work are as follows.

- We propose a NR-IQA metric via deep meta-learning, aiming at the generalization problem of applying CNN to the NR-IQA task. It can effectively build a highly generalizable NR-IQA model by a cross-distortion meta-optimization from synthetic to synthetic and synthetic to authentic distortions, which can directly evaluate the quality of images with unseen distortions.
- We propose a task selection strategy to select the NR-IQA tasks with representative synthetic distortions for the subsequent meta-optimization, which can effectively reduce the requirement of training data and significantly improve the generalization ability of the proposed NR-IQA model.
- We offer two experimental settings for quality evaluation of synthetic and authentic distortions, respectively. The experimental results demonstrate that the proposed model significantly outperforms our previous MetaIQA and the state-of-the-arts for NR-IQA with unseen distortions. Besides, the proposed NR-IQA model can be used as efficient prior knowledge for adapting to new NR-IQA tasks through fine-tuning.

The remainder of this paper is organized as follows. The related works on no-reference image quality assessment (NR-IQA) and deep meta-learning are briefly introduced in Section II. In Section III, we present details of the proposed optimization-based deep meta-learning for NR-IQA. Experimental results and visual analysis are presented in Section IV, and finally, conclusions are drawn in Section V.

II. Related Work

A. No-reference Image Quality Assessment

NR-IQA consists of two categories: distortion-specific metrics [12], [14], [13] and general-purpose metrics [15], [16], [17], [18], [19], [20], [21]. For distortion-specific metrics, image quality is quantified by characterizing the degree of known distortions. Since the characteristics of distortion types are known, these metrics have achieved significant evaluation performance. Nonetheless, the distortion types of images are often unknown in actual situations, which makes the application scope of these metrics is restricted [41]. Consequently, general-purpose NR-IQA metrics have become the focus of recent researchers [42].

In general, conventional general-purpose NR-IQA models are based on hand-crafted features, which can be classified into two categories: natural scene statistics (NSS)-based metrics [17], [15], [16] and learning-based metrics [19], [21]. The NSS-based metrics find that the statistical characteristics of natural images vary with the degree of distortions. Moorthy *et al.* [15] proposed a NR-IQA metric by extracting the NSS features from the discrete wavelet transform (DWT) domain of images. Saad *et al.* [16] evaluated the quality of images by using the statistical features from the discrete cosine transform (DCT) domain. Mittal *et al.* [17] used NSS features from the spatial domain to evaluate the quality of images and achieved very encouraging performance. In addition to the above metrics, learning-based metrics have also drawn attention. Ye *et al.* [19] proposed a codebook representation approach based on Support Machine Regression (SVR) model to predict the quality scores of images. Zhang *et al.* [21] proposed a NR-IQA metric by combining the semantic-level features that influence the human visual system with local features.

Recently, DCNN-based general-purpose NR-IQA metrics [23], [24], [25], [26], [27], [28], [29], [30], [31], [36] have shown better evaluation performance than hand-crafted feature-based metrics. The challenge of these metrics is that the effective learning of the deep model requires big labeled data, but IQA is typically a small sample learning task in practical applications [11]. In [23], Kang *et al.* first used CNN to learn deep features for the NR-IQA task and partitioned an input image into multiple 32×32 patches to alleviate the lack of training samples. But it is counterintuitive that the quality of a patch used for training is inherited from that of the corresponding image. Bianco *et al.* [36] proposed a NR-IQA metric to fine-tune the pre-trained deep model on the ImageNet database [37]. Talebi *et al.* [24] predicted the perceptual distribution of quality opinion scores through a DCNNs-based model, whose parameters were also initialized by pre-training on the large-scale database for image classification tasks [37]. However, the discrepancy between image classification and NR-IQA restricts the generalization performance of the model. In [25], [26], the authors leveraged a variety of distortion types to generate a large number of image pairs to train a prior model of quality ranking and then fine-tuned the prior model on several small-scale IQA databases to obtain quality evaluation models. Zhang *et al.* [31] proposed two-streams of pre-trained CNN model for synthetic and authentic distortions, respectively. Then, a bilinear pooling module was introduced for fine-tuning on a target NR-IQA task.

Although the above-mentioned metrics can efficaciously handle the NR-IQA task with known (trained) distortions, the generalizable ability of these NR-IQA models to unseen (untrained) distortions is often unsatisfactory. To tackle this problem, in this paper, we propose a radically different approach by using deep meta-learning [38] to achieve a highly generalizable NR-IQA model, which can directly evaluate the quality of an image with unseen distortion.

B. Deep Meta-learning

Deep meta-learning is a knowledge-driven learning framework, trying to deal with the issue of learning to learn [38]. Humans have a fast learning ability for new tasks, which largely depends on their prior knowledge learned from related tasks. Based on this idea, meta-learning is used to imitate this fast learning ability of humans. Generally, meta-learning can be classified into Recurrent Neural Networks (RNNs) memory-based models [43], [44], metric-based models [45], [46] and optimization-based models [47], [48]. The RNN memory-based models learn a new task by mainly using RNNs with memories to store experience knowledge from previous related tasks [43], [44]. The metric-based models first map the input space of some related tasks to a new embedding space by learning an embedding function and then using the nearest neighbor or linear classifiers to deal with a new task [45], [46]. The optimization-based models attempt to learn the initialization parameters of a model that can be quickly adapted to a new task by fine-tuning with a small number of training samples [47].

The above meta-learning models are elaborated for few-shot learning in image classification tasks [39], and each class usually contains only a few training samples. In contrast, the NR-IQA model needs to quantitatively predict image quality, which makes it more difficult and challenging to obtain training samples. Therefore, we aim to handle an unseen NR-IQA task directly without any model updating or fine-tuning. In view of this, we propose a no-reference image quality assessment metric via optimization-based meta-learning, which can learn a NR-IQA model with generalizable prior knowledge. To this end, we conduct meta-optimization across NR-IQA tasks with diversified distortions and learn a generalized NR-IQA model that can directly deal with a target NR-IQA task with unseen distortions.

III. Our Approach

In this section, we model the no-reference image quality assessment (NR-IQA) task in a deep meta-learning framework, which is a deep regression network with a spatial pyramid pooling module. Based on the deep meta-learning framework, we first generate a series of NR-IQA tasks from synthetically and authentically distorted IQA databases. Then, we propose a task selection strategy to build two task sets for the cross-distortion meta-optimization, enabling the NR-IQA model to capture the generalization ability from synthetic to synthetic and synthetic to authentic distortions. Finally, we leverage

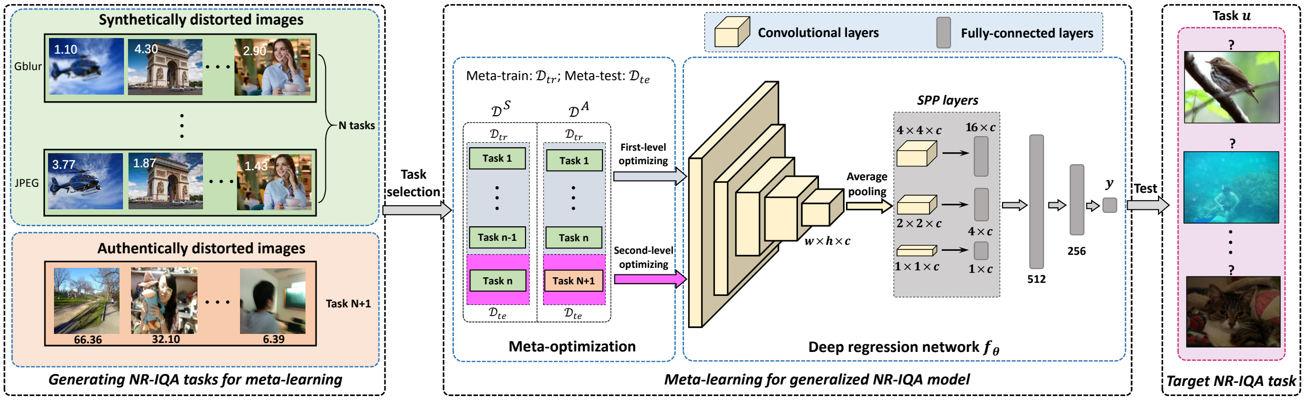


Fig. 2. The overview of our proposed MetaIQA+ framework, which is based on a deep regression network with a spatial pyramid pooling module. MetaIQA+ contains three parts: generating NR-IQA tasks for meta-learning, task selection for building two task sets, and bi-level meta-optimization to learn a generalized NR-IQA model. Once trained on NR-IQA tasks with known distortions, the proposed MetaIQA+ model can be directly tested on a target NR-IQA task with unseen distortions.

a bi-level meta-optimization to learn a generalized NR-IQA model across NR-IQA tasks in each task set. The learned NR-IQA model can directly test a target IQA task without model updating. The framework of the proposed model is illustrated in Fig. 2, which is termed Meta-learning-based IQA plus (MetaIQA+).

A. Deep Meta-learning Framework

To avoid the change of original image quality caused by the arbitrary change of input size, a deep regression network f_θ implemented with a spatial pyramid pooling (SPP) module [40] is proposed to build our NR-IQA model. The deep regression network includes convolutional layers, SPP layers, and fully-connected layers, as shown on the right side of Fig. 2. Particularly, the convolutional layers come from a prevalent DCNN, and the SPP operation on top of the last convolutional layer is employed for yielding a fully-connected layer. With the fixed spatial bin size (1×1, 2×2, 4×4), we make sure that the fully-connected layer gets the fixed shape input regardless of input image size. Then, another two fully-connected layers are added to produce the output of the proposed deep regression network.

Specifically, we fed an image x into the proposed deep regression network f_θ for obtaining the predicted quality score $\hat{y}$, which can be formulated as

$$\hat{y} = f_\theta(x;\theta), \tag{1}$$

where θ represents the initialized parameters of our deep network. We use the squared Euclidean distance as the loss function of our model for minimizing the difference between the ground-truth and predicted quality scores of the image x. The loss function can be defined as

$$\mathcal{L} = \|f_\theta(x;\theta) - y\|_2^2, \tag{2}$$

where y indicates the ground-truth quality score.

B. Generating NR-IQA Tasks

As mentioned in [42], the traditional NR-IQA models are distortion-aware and usually trained on images with several common distortions, which can only effectively evaluate the quality of images with known (trained) distortion types. This restricts the generalization ability of the traditional NR-IQA models for images with unseen distortion types. Consequently, a generalized NR-IQA model should have prior knowledge learned from known distortions and generalized to unseen distortions.

Inspired by the concept of learning to learn in deep meta-learning [38], we could learn the generalization ability of the deep model from one task to another task during the meta-optimization process. To learn a generalizable NR-IQA model from different distortions, we need to generate a series of NR-IQA tasks that can be used for meta-learning. A NR-IQA task is an IQA task for a specific distortion type (authentic distortion is regarded as one distortion type). For images with synthetic distortions, N NR-IQA tasks for specific distortion types (e.g., JPEG or blur) can be produced. For images with authentic distortion, we consider it as the $(N+1)$-th NR-IQA task, as shown on the left side of Fig. 2.

C. Task Selection

In real-world applications, since the subjective experiment of image quality assessment is time-consuming and laborious, the available types of synthetic distortions that can be used in our experiment are limited. Therefore, we expect that our NR-IQA model can be learned from as few known distortions as possible while ensuring the evaluation performance for images with unseen distortions. In this way, we can reduce the requirement of training data and learn a more generalizable NR-IQA model. For this purpose, it is essential to select the NR-IQA tasks with representative synthetic distortions for the subsequent meta-optimization. Inspired by [48], we exploit a gradient direction similarity approach that measures the gradient agreement among N NR-IQA tasks with synthetic distortions.

Suppose $\mathcal{D}^{p(\tau)} = \{\mathcal{D}^{\tau_1}, \mathcal{D}^{\tau_2}, ..., \mathcal{D}^{\tau_N}\}$ is the set of N NR-IQA tasks with synthetic distortions. For the i-th task in $\mathcal{D}^{p(\tau)}$ ($i \in \{1, 2, ..., N\}$), we first compute the gradients of loss function $\mathcal{L}_{\mathcal{D}^{\tau_i}}$ that are related to all network parameters θ for R times

$$g_i = \sum_{r=1}^{R} \alpha_{se} \nabla_\theta \mathcal{L}_{\mathcal{D}^{\tau_i}}(f_\theta), \tag{3}$$

where α_{se} is the learning rate of selection. Similar to [48], we then calculate the gradient direction similarity factor w_i between the i-th task and all the other tasks, which can be defined as

$$w_i = \frac{\sum_{\mathcal{D}^{\tau_j} \in \mathcal{D}^{p(\tau)}} (g_i^T g_j)}{\sum_{\mathcal{D}^{\tau_h} \in \mathcal{D}^{p(\tau)}} |\sum_{\mathcal{D}^{\tau_j} \in \mathcal{D}^{p(\tau)}} (g_h^T g_j)|}. \tag{4}$$

The higher the w_i, the more agreement between the i-th distortion and all the other distortions. In other words, if the gradient direction of one distortion is more similar to that of all other distortions, it means that the distortion has a better general direction than other distortions. Finally, we rank all the N kinds of distortions according to the gradient direction similarity factor $\mathbf{W} = \{w_i\}_{i=1}^{N}$, which can be defined as

$$\mathbf{W}_{rank} = sort(\mathbf{W}), \tag{5}$$

where $sort(\cdot)$ is the descending function.

Next, we build two task sets for the subsequent meta-optimization. First, to achieve the generalization between synthetic distortions, top n $(1 \leq n \leq N)$ NR-IQA tasks are selected on the basis of $\mathbf{W}_{rank}$ to build a task set $\mathcal{D}^S$. Second, to obtain the generalization from synthetic to authentic distortions, the NR-IQA task with authentic distortion and the selected n specific distortion NR-IQA tasks are used to build another task set $\mathcal{D}^A$. In the task set $\mathcal{D}^S$, we randomly split n tasks into a meta-train set $\mathcal{D}_{tr}$ with $n-1$ source tasks and a meta-test set $\mathcal{D}_{te}$ with 1 target task, which simulates the generalization problem from known synthetic distortion to unseen synthetic distortion. In the task set $\mathcal{D}^A$, we divide the n tasks with synthetic distortions into meta-train set $\mathcal{D}_{tr}$ and take the task with authentic distortion as the meta-test set $\mathcal{D}_{te}$. The overview of building two task sets is shown in the middle of Fig. 2. Through the two task sets, the NR-IQA model is encouraged to learn prior knowledge that can generalize from known distortions to unseen distortions. In contrast to MetaIQA [10], the main difference of generated task sets is to learn the cross-task generalization ability between different distortion types, so it can obtain better generalization performance for NR-IQA tasks with unseen distortions. Different from previous optimization-based meta-learning, the proposed metric generates a series of tasks according to distortion types and aims at learning the generalization ability across different NR-IQA tasks.

D. Meta-optimization

In our metric, the learned NR-IQA model is expected to directly evaluate the quality of images with unseen distortions. Hence, we use a bi-level meta-optimization to learn the generalization ability of our NR-IQA model. First, we sample some NR-IQA tasks from the meta-train set to calculate all the gradients of our NR-IQA model and tentatively update them with first-level gradient descent. Then, we conduct second-level gradient descent on the NR-IQA task in the meta-test set to verify whether the updated model is executed effectively. By this means, our NR-IQA model can capture the generalization ability from the tasks in the meta-train set to the tasks in the meta-test set. The two-level gradient descent strategy from meta-train set to meta-test set is called bi-level meta-optimization.

Specifically, we randomly sample k tasks from the meta-train set and one task in the meta-test as a meta-batch $\mathcal{D}_b = \{\mathcal{D}_{btr}, \mathcal{D}_{bte}\}$, where $\mathcal{D}_{btr} = \{\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_k\}$ is the training set of meta-batch and $\mathcal{D}_{bte} = \{\mathcal{D}_{k+1}\}$ is the testing set of meta-batch ($1 \leq k \leq n-1$ for $\mathcal{D}^S$, and $1 \leq k \leq n$ for $\mathcal{D}^A$). For the i-th task $\mathcal{D}_i$ in the training set of meta-batch $\mathcal{D}_{btr}$, the loss function can be formulated as $\mathcal{L}_{\mathcal{D}_i}$ $(i \in \{1, 2, \ldots, k\})$ that is computed by Eq. 2. Because our NR-IQA model is more complex than the model of classification task in [47] and more training samples are available in each NR-IQA task, a more effective gradient descent strategy is used to optimize the proposed model. Hence, we first compute the first-order gradients of loss function $\mathcal{L}_{\mathcal{D}_i}$ that are related to the parameters of our model θ, which can be formulated as

$$g_\theta = \nabla_\theta \mathcal{L}_{\mathcal{D}_i}(f_\theta). \tag{6}$$

Then, our model is updated for P times by leveraging the Adam [49] optimizer on the training set of meta-batch $\mathcal{D}_i (i = 1, 2, ..., k)$, which can be defined as

$$\theta_i^{'} \leftarrow \theta - G_{ad}(\mathcal{L}_{\mathcal{D}_i}; \theta). \tag{7}$$

The $G_{ad}(\mathcal{L}_{\mathcal{D}_i}; \theta)$ can be calculated by

$$G_{ad}(\mathcal{L}_{\mathcal{D}_i}; \theta) = \alpha \sum_{j=1}^{P} \frac{m_{\theta(j)}}{\sqrt{v_{\theta(j)}} + \epsilon}, \tag{8}$$

where α denotes the inner learning rate and $\epsilon = 1e-8$. $m_{\theta(j)}$ and $v_{\theta(j)}$ denote the first moment and second raw moment of gradients $g_{\theta(j)}$, which are calculated by

$$m_{\theta(j)} = \mu_1 m_{\theta(j-1)} + (1-\mu_1) g_{\theta(j)}, \tag{9}$$

$$v_{\theta(j)} = \mu_2 v_{\theta(j-1)} + (1-\mu_2) g_{\theta(j)}^2, \tag{10}$$

where $m_{\theta(0)} = 0$ and $v_{\theta(0)} = 0$. μ_1 and μ_2 denote the exponential decay rates of $m_{\theta(j)}$ and $v_{\theta(j)}$, respectively. $g_{\theta(j)}$ represents the gradients of the updated model in j-th $(j \in \{1, 2, ..., P\})$ time.

As we mentioned previously, the proposed NR-IQA model updated with the meta-train set is expected to perform well on the meta-test set. Instead computing second-order gradients in [47], we then calculate the first-order gradients of loss function $\mathcal{L}_{\mathcal{D}_{k+1}}$ that are related to the parameters of updated model $\theta_i^{'}$ for a second time, which can be defined as

$$g_{\theta_i^{'}} = \nabla_{\theta_i^{'}} \mathcal{L}_{\mathcal{D}_{k+1}}(f_{\theta_i^{'}}). \tag{11}$$

We update the model parameters $\theta_i^{'}$ for P times by using Adam optimizer on the testing set of meta-batch $\mathcal{D}_{k+1}$, which takes the following form

$$\theta_i \leftarrow \theta_i^{'} - G_{ad}(\mathcal{L}_{\mathcal{D}_{k+1}}; \theta_i^{'}). \tag{12}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3073410, IEEE Transactions on Circuits and Systems for Video Technology

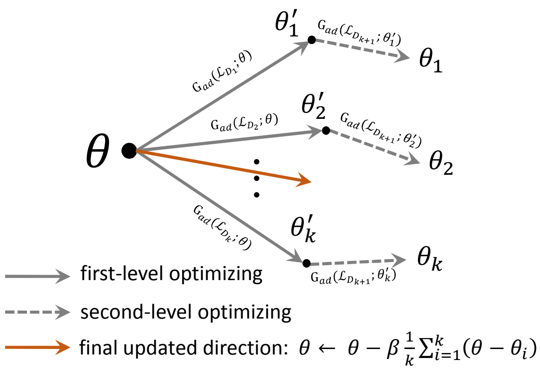


Fig. 3. The overview of meta-optimization in a meta-batch $\mathcal{D}_b$. The meta-batch consists of a training set with k tasks $\mathcal{D}_{btr} = \{\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_k\}$ and a testing set with one task $\mathcal{D}_{bte} = \{\mathcal{D}_{k+1}\}$. The gradients of the model is first calculated on the training set of meta-batch $\mathcal{D}_{btr}$. Then, we calculate the updating gradient of model on the testing set of meta-batch $\mathcal{D}_{bte}$ for the second time. Finally, the model is updated to a direction that can learn the generalization ability from the training set to testing set in the meta-batch by integrating of all gradients.

The $G_{ad}(\mathcal{L}_{\mathcal{D}_{k+1}}; \theta_i^{'})$ can be calculated by

$$G_{ad}(\mathcal{L}_{\mathcal{D}_{k+1}}; \theta_i^{'}) = \alpha \sum_{j=1}^{P} \frac{m_{\theta_i^{'(j)}}}{\sqrt{v_{\theta_i^{'(j)}}} + \epsilon}, \tag{13}$$

where $m_{\theta_i^{'(j)}}$ and $v_{\theta_i^{'(j)}}$ represent the first moment and second raw moment of gradients $g_{\theta_i^{'(j)}}$. For the meta-batch, the gradients of all tasks are integrated to update the final model parameters, which can be defined as

$$\theta \leftarrow \theta - \beta \frac{1}{k} \sum_{i=1}^{k} (\theta - \theta_i), \tag{14}$$

where β is the outer learning rate. In this way, k tasks are iteratively sampled from the meta-train set $\mathcal{D}_{tr}$ and one task in the meta-test set $\mathcal{D}_{te}$ is used to build a meta-batch $\mathcal{D}_b$ for optimizing the proposed deep regression network f_θ. Ultimately, the generalized NR-IQA model learned from known distortions to unseen distortions can be obtained by the bi-level meta-optimization.

After training our model through NR-IQA tasks with known distortions, we can directly use this model to test a target NR-IQA task u with unseen distortions. For a query image x_q in the target NR-IQA task, the quality score $\hat{y_q}$ can be predicted by using our NR-IQA model $\hat{y_q} = f_\theta(x_q; \theta)$. The procedure of our MetaIQA+ is summed up in Algorithm 1 and the overview of meta-optimization in a meta-batch is illustrated in Fig. 3.

Algorithm 1 The proposed MetaIQA+

Input: N NR-IQA tasks from synthetically distorted IQA databases and one task from authentically distorted IQA databases, a query image x_q with unseen distortions in the target NR-IQA task u, learning rate α, β

Output: Predicted quality score $\hat{y}_q$ for x_q

```
Deep regression network f with initial parameters θ;
Select n NR-IQA tasks with representative synthetic distortions through gradient direction similarity;
Build two task sets D^S and D^A;
for iteration = 1, 2, ... do
    Randomly split D^S (or D^A) into D_tr and D_te;
    /* meta-optimization */
    Sample k tasks D_btr = {D_1, D_2, ..., D_k} from D_tr and one task D_btr = {D_{k+1}} from D_tr;
    Build a meta-batch D_b = {D_btr, D_bte};
    for i = 1, 2, ..., k do
        /* first-level optimizing */
        Compute θ'_i = θ − G_ad(L_{D_i}; θ) on D_i ;
        /* second-level optimizing */
        Compute θ_i = θ'_i − G_ad(L_{D_{k+1}}; θ'_i) on D_{k+1};
    end for
    Update θ ← θ − β (1/k) Σ_{i=1}^{k} (θ − θ_i);
end for
Input x_q into the trained NR-IQA model f_θ;
return ŷ_q.
```

TABLE I

SUMMARY OF SYNTHETICALLY DISTORTED IQA DATABASES RELATING TO NUMBERS OF REFERENCE IMAGES (REF.), DISTORTION IMAGES (DIST.), DISTORTION TYPES (DIST. TYPES), AND SCORE RANGE. A HIGHER SCORE REPRESENTS BETTER QUALITY.

Databases	Ref.	Dist.	Dist. Types	Score Range
TID2013 [35]	25	3,000	24	[0, 9]
KADID-10k [50]	81	10,125	25	[1, 5]

IV. EXPERIMENTS

A. Image Quality Databases

The performance of the proposed NR-IQA model is verified on two types of databases: synthetically distorted IQA databases and authentically distorted IQA databases.

Synthetically distorted IQA databases are used to evaluate the generalization performance of our NR-IQA model for unseen synthetic distortions, which include TID2013 [35] and KADID-10k [50]. Table I summarizes the information of these two databases.

Authentically distorted IQA databases can be used to verify the generalization performance of the learned NR-IQA model for unseen authentic distortions, which include CID2013 [51], LIVE in the wild image quality challenge (LIVE challenge) [41] and KonIQ-10k [52] databases. The CID2013 database consists of six subsets with a total of 480 authentically distorted images. These images were captured by 79 digital cameras. Participants were employed in the user study to rate the quality scores of images that range from 0 to 100, and a higher score indicates better quality. The LIVE challenge database consists of 1,162 images with authentic distortions, such as JPEG compression, overexposure or underexposure, noise, and motion blur. Therefore, the distortion types of images are unknown and there are no available reference images. The quality scores of images were acquired

TABLE II
COMPARISON RESULTS (SROCC) IN LEAVE-ONE-DISTORTION-OUT CROSS-VALIDATION ON TID2013 AND KADID-10K DATABASES.

	Dist. type	BLIINDS-II [16]	BRISQUE [17]	ILNIQE [18]	CORNIA [19]	HOSA [20]	WaDIQaM-NR [28]	DB-CNN [31]	MetaIQA [10]	SPP-FT	**MetaIQA+**
TID2013	AGN	0.6938	0.7582	**0.8720**	0.3492	0.3431	0.7745	0.7585	0.7334	0.7157	0.8415
	ANC	0.6931	0.4670	0.8349	0.0966	0.3178	0.7008	**0.8423**	0.6352	0.6134	0.6662
	SCN	0.8738	0.6246	**0.9172**	0.6908	0.2731	0.7265	0.7800	0.8197	0.7254	0.7608
	MN	0.1908	0.5125	0.5128	0.5062	0.3158	0.5540	0.4501	0.5039	0.5475	**0.6131**
	HFN	0.8162	0.8285	0.8735	0.8482	0.6263	0.7900	0.8192	0.7640	0.8214	**0.8808**
	IN	0.7123	0.1889	0.7751	0.3362	0.0673	**0.7308**	0.5798	0.6584	0.6215	0.6300
	QN	0.6500	0.4145	**0.8626**	0.4885	0.1331	0.6985	0.7502	0.6808	0.6451	0.7369
	GB	0.8962	0.7823	0.7967	0.8923	0.4492	0.7899	0.9015	0.8981	0.8457	**0.9183**
	DEN	0.7692	0.5436	0.7377	0.5597	0.5531	0.6232	**0.8103**	0.7767	0.7564	0.7962
	JPEG	0.8608	0.8318	0.8326	**0.8785**	0.7686	0.6495	0.8131	0.8582	0.8341	0.8623
	JP2K	0.8254	0.5097	0.8417	0.7798	0.4155	0.7168	0.8162	0.8663	0.7891	**0.8752**
	JGTE	0.1892	0.4494	0.2956	0.6931	0.3869	0.3785	0.7023	0.7488	0.7159	**0.7898**
	J2TE	0.6738	0.1405	0.5289	0.2638	0.1700	0.4639	0.6577	0.7035	0.6816	**0.7469**
	NEPN	-0.0723	0.2163	-0.0788	0.4108	0.2247	0.2183	0.3774	0.3729	0.3559	**0.4224**
	Block	0.3672	0.3767	-0.1401	-0.0258	0.2869	0.2740	0.4169	0.4930	0.5124	**0.6244**
	MS	0.3038	0.0633	0.1998	0.2077	**0.4615**	0.3062	0.1023	0.2654	0.3017	0.3634
	CTC	-0.0569	0.0466	0.0268	0.0800	0.0969	0.3849	0.2946	0.3069	0.3451	**0.3973**
	CCS	0.1338	-0.1390	-0.1720	0.2908	-0.0038	0.4281	0.3423	0.4308	0.4128	**0.4457**
	MGN	0.7054	0.5491	0.6962	0.3516	0.3415	0.7408	0.6700	0.7633	0.7125	**0.7855**
	CN	0.1823	0.3740	0.3621	0.6444	0.1254	0.6971	0.5544	0.6864	0.6483	**0.7148**
	LCNI	0.3477	0.5053	**0.8181**	0.7623	0.2569	0.5455	0.7600	0.6920	0.6753	0.7397
	ICQD	0.5685	0.7936	0.7365	0.7792	0.6562	0.6508	0.7608	0.7836	0.7458	**0.7988**
	CHA	0.7677	0.6657	0.6649	0.8102	0.5877	0.7488	0.7097	0.8016	0.8064	**0.8130**
	SSR	0.6246	0.8273	**0.8770**	0.8569	0.7254	0.6806	0.7700	0.8015	0.7861	0.8121
	Average	0.5299	0.4720	0.5697	0.5230	0.3575	0.5947	0.6433	0.6685	0.6506	**0.7098**
KADID-10k	GB	0.7637	0.7789	**0.8931**	0.8426	0.8006	0.8296	0.8378	0.8012	0.7842	0.8412
	LB	0.5974	0.6703	0.8555	0.7998	0.6627	0.6924	0.7868	0.7588	0.7831	**0.8645**
	MB	0.3986	0.5093	0.7782	0.4486	0.6004	0.5930	0.6450	0.7815	0.7428	**0.7915**
	CD	0.6075	0.4833	0.6552	0.1001	0.6847	0.7605	0.7851	0.7564	0.7485	**0.8064**
	CS	0.1808	-0.0268	0.0751	0.0393	-0.0048	0.3871	0.6219	0.6134	0.6074	**0.6734**
	ICQD	0.7626	0.7091	0.6736	0.3100	0.6166	0.7616	**0.7962**	0.6487	0.6411	0.6587
	CCS	0.2443	0.1415	0.0454	0.0979	0.1609	0.1222	0.1695	0.2354	0.2249	**0.2554**
	CCS-Lab	0.6208	0.3249	0.6853	0.3511	0.6101	0.7191	0.8184	0.8121	0.7873	**0.8211**
	JP2K	0.4921	0.5342	0.7772	0.3268	0.5695	0.4992	0.7483	0.7804	0.7451	**0.8138**
	JPEG	0.3049	0.7692	0.7989	0.5243	0.5158	0.5078	0.7424	0.7713	0.7498	**0.8013**
	AGN	0.6560	0.5946	0.7595	0.3479	0.6334	**0.8533**	0.8171	0.7557	0.7394	0.8057
	ANC	0.8082	0.5898	**0.8602**	0.2603	0.7083	0.7771	0.8529	0.8309	0.8109	0.8209
	IN	0.3675	-0.1336	**0.8213**	0.2582	0.2212	0.7916	0.7055	0.7457	0.7149	0.7396
	MGN	0.5804	0.5963	0.6786	0.2246	0.5286	0.7397	0.7532	0.7883	0.7547	**0.8083**
	DEN	0.3150	0.1590	0.7399	0.2459	0.2071	0.7221	**0.8318**	0.7634	0.7341	0.7586
	Brighten	0.3473	0.2670	0.2912	0.1999	0.5082	0.6362	0.6291	0.6511	0.6412	**0.6617**
	Darken	0.3037	0.4207	0.4416	0.2498	0.4935	0.2502	0.4925	0.5157	0.5034	**0.5471**
	MS	0.0657	0.2142	0.3154	0.1259	0.4450	0.3090	0.4149	0.5067	0.4876	**0.5248**
	Jitter	0.0390	0.6270	0.4470	0.6521	0.3569	0.7115	0.4889	0.6454	0.6176	**0.7245**
	NEPN	0.2820	0.0372	0.2337	0.2030	**0.4051**	0.3292	0.3085	0.3852	0.3479	0.3533
	Pixelate	-0.0168	0.6718	0.5619	0.6123	0.6382	0.6659	0.6265	0.7375	0.7198	**0.7889**
	QN	0.6194	0.7248	0.5714	0.2275	0.6233	0.6881	**0.7483**	0.7186	0.7018	0.7271
	Block	0.0240	-0.0867	0.0066	0.0694	0.3417	0.2232	0.3614	0.3953	0.3715	**0.4459**
	HS	-0.0755	0.3988	**0.6766**	0.2419	0.1780	0.5430	0.4694	0.4264	0.4371	0.4772
	CTC	0.0413	0.1374	0.0548	0.0347	0.3892	0.3925	0.3383	0.4463	0.4219	**0.4933**
	Average	0.3732	0.4045	0.5479	0.3118	0.4758	0.5802	0.6316	0.6509	0.6327	**0.6802**

by crowdsourcing experiments, which range from 0 to 100. We also include a recently released IQA database: KonIQ-10k, which is a relatively large-scale consisting of 10,073 authentically distorted images [52]. Each image received five-point ratings from about 120 workers.

B. Experimental Setup

To verify the generalization performance of our NR-IQA model for synthetic and authentic distortions, we conduct two experimental settings, respectively.

1) Experiment I: In the first experiment, we verify whether the learned NR-IQA model from the task set $\mathcal{D}^S$ has a highly generalizable performance for unseen synthetic distortions. To this end, we conduct Leave-One-Distortion-Out cross-validation on two synthetically distorted IQA databases: TID2013 [35] and KADID-10k [50]. Therefore, this experiment does not require the operation of task selection. Different from the experimental settings in MetaIQA [10], on the basis of ensuring that distortion types do not overlap, we further make the image content in training and testing do not overlap. Specifically, we randomly divide the reference images and the corresponding distorted images into two subsets with 80% for training and 20% for testing. Suppose there are M distortions in each database. The images with $M-1$ distortions in the training set are used as the task set $\mathcal{D}^S$, and the images with left-out one distortion in the test set are used for a target task u. Our NR-IQA model is learned from the task set $\mathcal{D}^S$ through meta-optimization, and then directly used to evaluate the performance on the target task u without fine-tuning. We do 10 repeated experiments on the partition of training-testing to avoid random bias and report the average results of 10 times.

2) Experiment II: In the second experiment, we further evaluate the proposed NR-IQA model learned from two task sets $\mathcal{D}^S$ and $\mathcal{D}^A$ for unseen authentic distortions. As mentioned in Section III-C, we need to use the task selection operation for determining the number of selected tasks. In this experiment, we merge images with the same distortion on TID2013 and KADID-10k databases into one task and obtain a total of 34 ($N = 34$) NR-IQA tasks. We train our NR-IQA model using the selected n tasks with synthetic distortions and one task with authentic distortion and then directly test it on a target u with authentic distortion. For the NR-IQA tasks of CID2013 [51] and LIVE challenge [41], we use the KonIQ-10k [52] as a task in the meta-test of $\mathcal{D}^A$. For the NR-IQA task of KonIQ-10k [52], we use the LIVE challenge [41] as a task in the meta-test of $\mathcal{D}^A$. In this way, we ensure that all authentically distorted images in the model test are not used for model training. To address the problem of nonlinear alignment of subjective scores in these five IQA databases, we use a nonlinear module for each database to learn the nonlinear mapping between predicted scores and subjective

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3073410, IEEE Transactions on Circuits and Systems for Video Technology

Fig. 4. Example images (a)-(d) with the first four distortions according to the $\mathbf{W}_{rank}$. (a) Lens blur (LB). (b) JPEG compression (JPEG). (c) JPEG2000 compression (JP2K). (d) Gaussian blur (GB).

scores during mixed databases training. Inspired by [53], this nonlinear module is composed of two linear layers and a sigmoid activation function (Linear+Sigmoid+Linear).

3) Implementation details and Evaluation criteria: In our NR-IQA model, we use ResNet18 [22] as the convolutional layers of the proposed deep regression network. In task selection, we set the learning rate of selection α_{se} and the number of selected tasks n to $1e-4$ and 24 respectively. In bi-level meta-optimization, we set the inner learning rate α and the outer learning rate β to $1e-4$ and $1e-2$ respectively. The number of images sampled in a mini-batch is set to 25. All the learning rates drop down to a factor of 0.9 after every 50 iterations and the total number of iterations is 500. Besides, we also set the following hyper-parameters: weight decay of $1e-5$, meta-batch size k of 4, the number of learning in the inner loop P of 5, exponential decay rate μ_1 of 0.9, exponential decay rate μ_2 of 0.99. The proposed model is implemented based on Pytorch [54].

For the NR-IQA task, the ranking and linear consistency between predicted and ground-truth quality scores are two critical evaluation criteria. Consequently, we use Spearman Rank Order Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient (PLCC) to evaluate the performance of the proposed and several state-of-the-art metrics, which is similar to [28], [30]. The PLCC and SROCC are in the range [-1, 1] and higher absolute values represent better evaluation performance.

C. Performance on synthetically distorted IQA databases

In Experiment I, we compare the proposed MetaIQA+ metric with seven state-of-the-art general-purpose NR-IQA metrics and our previous MetaIQA [10] on TID2013 [35] and KADID-10k [50] databases. These metrics are BLIINDS-II [16], BRISQUE [17], ILNIQE [18], CORNIA [19], HOSA [20], WaDIQaM-NR [28] and DB-CNN [31]. Besides, to verify the effectiveness of the proposed meta-optimization, we also include a baseline model that fine-tunes our deep regression network implemented with the SPP module by using the Adam optimizer directly (called SPP-FT). For a fair comparison, we use the source codes of these metrics provided by the authors and conduct the experiments under the same training-testing setting (Leave-One-Distortion-Out cross-validation).

The comparison results in terms of SROCC are summarized in Table II, in which we show the best result for each distortion

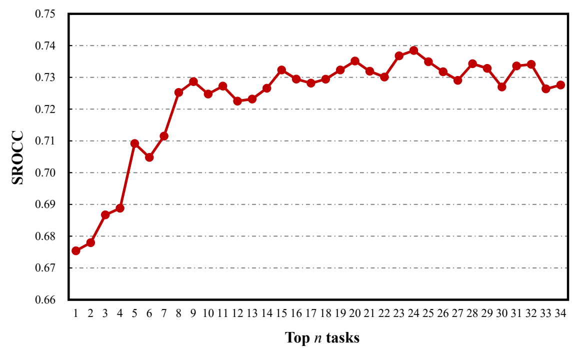


Fig. 5. Performance comparisons (SROCC) on the LIVE challenge database with different top n tasks used for meta-optimization.

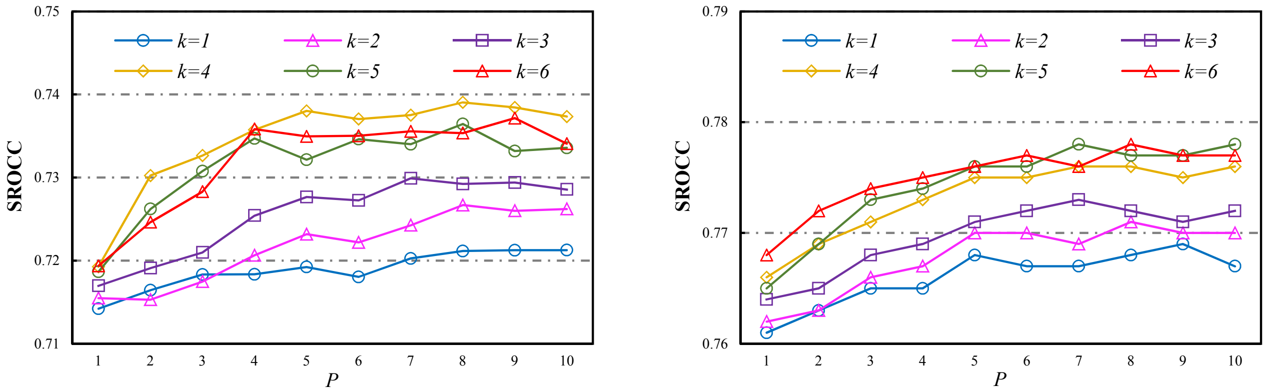
type in bold. It can be easily observed from Table II that the proposed metric is significantly better than other metrics in terms of overall performance (average results) on both databases. The proposed model can obtain the best performance for more than half of the distortion types (14 out of 24 on TID2013 and 16 out of 25 on KADID-10k). Moreover, the experimental results also demonstrate that the proposed MetaIQA+ based on cross-distortion meta-optimization is superior to our previous MetaIQA and the proposed baseline (SPP-FT). This is mainly because the cross-distortion meta-optimization in the training process makes the optimized NR-IQA model obtain the generalizability for unseen distortions without model fine-tuning.

D. Performance on authentically distorted IQA databases

In Experiment II, we first determine the number of selected tasks n for meta-optimization (Section IV-D1), then we discuss the effectiveness of two critical parameters in the meta-optimization of the proposed NR-IQA model (Section IV-D2), and finally compare the proposed metric with some state-of-the-art NR-IQA metrics on CID2013 [51], LIVE challenge [41] and KonIQ-10k [52] databases (Section IV-D3).

1) Task selection: To demonstrate that our task selection can effectively select the representative distortion types for meta-optimization, Fig. 4 shows example images with the first four distortions according to the $\mathbf{W}_{rank}$. As shown in Fig. 4, the first four distortions are all caused by the loss of high-frequency information in images. Compared with other distortions in images, the loss of high-frequency information in images (e.g., LB, JPEG, JP2K, and GB) is a more generally

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3073410, IEEE Transactions on Circuits and Systems for Video Technology



(a) (b)

Fig. 6. The effectiveness of parameters k and P in the proposed meta-optimization on two authentically distorted IQA databases. (a) LIVE challenge. (b) KonIQ-10k.

representative distortion type [55]. The NR-IQA model learned from a small number of representative distortion types can obtain the general knowledge of distortion and effectively generalize to the unseen distortions.

Furthermore, different top n tasks ($1 \leq n \leq 34$) and the task of KonIQ-10k are used to train NR-IQA models, which are directly tested on the LIVE challenge database. The performance comparisons in terms of SROCC are shown in Fig. 5. Since the first four distortions are similar (as shown in Fig. 4), the performances on the LIVE challenge database increase slowly when n increases from 1 to 4. The fifth distortion (Color saturation in LAB space) is quite different from the first four distortions, which increases the distortion diversity for model training. Therefore, the performance is greatly improved when n is from 4 to 5. When n increases from 5 to 6, the performance slightly decreased. This may be due to the reasons that the sixth distortion (Color saturation in HSV space) is similar to the fifth distortion and quite different from authentic distortion in the LIVE challenge database. When n is larger than 9, the performances on the LIVE challenge database tend to be stable. This demonstrates that the top nine tasks are easier to reach an agreement and the corresponding distortions have more universally representative. Also, the SROCC results do not increase continuously with an increase of n. This demonstrates the necessity of task selection for meta-optimization in our approach. When n is 24, our model can obtain the best performance on the LIVE challenge database. Therefore, we set $n = 24$ in our proposed MetaIQA+.

2) Parameters discussion: We first discuss the efficacy of two critical parameters in the proposed meta-optimization, i.e. k to dominate the number of NR-IQA tasks in a meta-batch and P to control the number of learning in the inner loop. Since our NR-IQA model is for authentic distortion, we set k and P in the proposed meta-optimization to different values and show the test results (SROCC) on two commonly used IQA databases (LIVE challenge and KonIQ-10k) in Fig. 6. Overall, the evaluation performances of our NR-IQA model improve with the increase of k and P. In the LIVE challenge database, the evaluation performance starts to deteriorate when

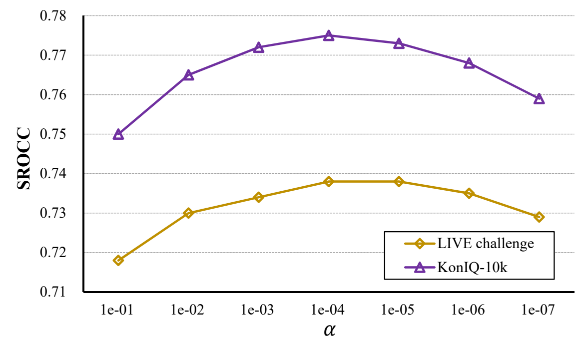


Fig. 7. The effectiveness of parameter α in the proposed model on the LIVE challenge and KonIQ-10k databases.

TABLE III
COMPARISON RESULTS (PLCC AND SROCC) OF OUR METRIC WITH SEVEN STATE-OF-THE-ART NR-IQA METRICS AND OUR PREVIOUS METAIQA BY DIRECTLY TESTING ON THREE AUTHENTICALLY DISTORTED IQA DATABASES: CID2013 [51], LIVE CHALLENGE [41] AND KONIQ-10K [52].

Metrics	CID2013		LIVE challenge		KonIQ-10k	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
BLIINDS-II [16]	0.213	0.124	0.242	0.171	0.108	0.089
BRISQUE [17]	0.536	0.521	0.428	0.400	0.355	0.310
ILNIQE [18]	0.538	0.346	0.589	0.594	0.537	0.501
CORNIA [19]	0.577	0.529	0.543	0.519	0.345	0.315
HOSA [20]	0.588	0.544	0.558	0.522	0.377	0.355
WaDIQaM-NR [28]	0.524	0.512	0.608	0.607	0.689	0.676
DB-CNN [31]	0.834	0.799	0.726	0.713	0.764	0.734
MetaIQA [10]	0.753	0.721	0.743	0.716	0.776	0.747
MetaIQA+	**0.872**	**0.868**	**0.767**	**0.738**	**0.813**	**0.775**

k is greater than 4. Moreover, the test results (SROCC) begin to increase slowly when P exceeds 5. In the KonIQ-10k database, our NR-IQA model can also achieve stable evaluation performance when k and P are larger than 4 and 5, respectively. In summary, k and P are set to 4 and 5 in all the subsequent experiments, respectively. Then, we further discuss the influence of another key parameter (the inner learning rate α) on our model. The parameter α is set to different values and the test results (SROCC) on the LIVE challenge and KonIQ-10k databases are shown in Fig. 7. As can be seen from the

TABLE IV
COMPARISON RESULTS (PLCC AND SROCC) OF OUR METRIC WITH FIFTEEN STATE-OF-THE-ART NR-IQA METRICS AND OUR PREVIOUS METAIQA BY FINE-TUNING ON THREE AUTHENTICALLY DISTORTED IQA DATABASES: CID2013 [51], LIVE CHALLENGE [41] AND KONIQ-10K [52].

Metrics	CID2013		LIVE challenge		KonIQ-10k	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
BLIINDS-II [16]	0.565	0.487	0.507	0.463	0.615	0.529
BRISQUE [17]	0.648	0.615	0.645	0.607	0.537	0.473
ILNIQE [18]	0.538	0.346	0.589	0.594	0.537	0.501
CORNIA [19]	0.680	0.624	0.662	0.618	0.795	0.780
HOSA [20]	0.685	0.663	0.678	0.659	0.813	0.805
deepIQA [56]	/	/	0.482	0.493	0.606	0.604
BIECON [57]	0.620	0.606	0.613	0.595	/	/
dipIQ [26]	0.709	0.702	0.290	0.187	0.437	0.228
MEON [29]	0.703	0.701	0.693	0.688	/	/
DistNet-Q3 [58]	/	/	0.601	0.570	0.710	0.702
DeepFL-IQA [33]	/	/	0.769	0.734	0.887	0.877
DIQA [59]	0.720	0.708	0.704	0.703	/	/
NSSADNN [30]	0.825	0.748	0.813	0.745	/	/
WaDIQaM-NR [28]	0.729	0.708	0.680	0.671	0.761	0.739
DB-CNN [31]	0.871	0.863	0.869	0.851	0.869	0.856
MetaIQA [10]	0.784	0.766	0.835	0.802	0.887	0.850
MetaIQA+	**0.891**	**0.882**	**0.872**	**0.852**	**0.921**	**0.909**

TABLE V
ABLATION STUDY RESULTS (PLCC AND SROCC) ON THREE AUTHENTICALLY DISTORTED IQA DATABASES: CID2013 [51], LIVE CHALLENGE [41] AND KONIQ-10K [52].

Metrics	CID2013		LIVE challenge		KonIQ-10k	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
ResNet-FT	0.716	0.701	0.721	0.703	0.768	0.737
SPP-FT	0.810	0.799	0.744	0.713	0.782	0.753
MetaIQA+	**0.872**	**0.868**	**0.767**	**0.738**	**0.813**	**0.775**

figure, our NR-IQA model can achieve the best evaluation performance in both databases when $\alpha = 1e-4$. Therefore, α is set to $1e-4$ in our experiments.

3) Comparisons with the state-of-the-art NR-IQA metrics: To examine the generalizable performance of our NR-IQA model, we compare the proposed metric with seven state-of-the-art general-purpose NR-IQA metrics and our previous MetaIQA on CID2013 [51], LIVE challenge [41] and KonIQ-10k [52] databases. All these metrics are tested under the settings of Experiment II, and the directly tested results on all images of the three databases are listed in Table III, where the best results among the NR-IQA metrics for each database are shown boldfaced. From the results in Table III, we can conclude that our MetaIQA+ model can obtain the best performance through direct testing on the authentically distorted IQA databases. In particular, the proposed MetaIQA+ is significantly superior to MetaIQA on the CID2013 database. The reasons may be that the size of images in the database is larger than the other two databases, and our NR-IQA model with SPP module can greatly improve the evaluation performance by inputting images with original size. This demonstrates that the proposed MetaIQA+ model has better generalization ability for unseen authentic distortions.

To further verify the adaptability of our NR-IQA model, we also fine-tune it on the three authentically distorted IQA databases and compare it with more state-of-the-art NR-IQA metrics and our previous MetaIQA. For a fair comparison with the reported results of existing NR-IQA metrics, we do the same training-testing setting as [28], [30]. In the CID2013 database, we use four out of six subsets as training samples and the remaining two subsets as testing samples. In LIVE challenge or KonIQ-10k database, we randomly split all images into 80% training samples and 20% testing samples. Specifically, we use the training samples of each database to fine-tune our meta-optimized MetaIQA+ model, which will further promote the evaluation performance of the test samples. In the fine-tuning procedure, the squared Euclidean distance between the ground-truth and predicted quality scores of training samples is used as the loss function to optimize our NR-IQA model. After using the training samples of each database for fine-tuning, a new NR-IQA model can be obtained for evaluating the quality of the test samples.

We do 10 repeated experiments on the partition of training-testing to avoid random bias and summarize the average results of the three databases in Table IV. Since KonIQ-10k is a recently released database and CID2013 is relatively small, several DCNN-based NR-IQA metrics have not released their testing results on the two databases. We can see that our metric is superior to the state-of-the-arts on CID2013 and KonIQ-10k databases. MetaIQA+ has obtained slightly better performance than DB-CNN on the LIVE challenge database, which outperforms other NR-IQA metrics by a large margin. This demonstrates that our NR-IQA model based on deep meta-learning can also be used as an effective prior model for adapting to NR-IQA tasks with authentic distortions.

4) Ablation study: To further examine whether the advantage of MetaIQA+ is achieved from the proposed deep regression network or the optimization-based meta-learning, we also conduct ablation studies. We compare MetaIQA+ with two baseline metrics: fine-tuning the ResNet18 [22] (the last layer replaced with regression prediction) pre-trained on ImageNet by using the Adam optimizer directly (called ResNet-FT), fine-tuning the proposed deep regression network implemented with the SPP module by using the Adam optimizer directly (called SPP-FT).

The testing results on CID2013 [51], LIVE challenge [41] and KonIQ-10k [52] databases are listed in Table V. As can be seen in this table, the proposed MetaIQA+ metric is significantly better than the two baseline metrics on all these databases. From the comparison results of ResNet-FT and SPP-FT, it is known that our network implemented with the SPP module can effectively improve the evaluation performance. Besides, SPP-FT and MetaIQA+ have the same network structure but are optimized by two different approaches. Compared with SPP-FT, MetaIQA+ has superior evaluation performance and can promote the performance of the NR-IQA model without changing the network structure. This indicates that both meta-optimization and SPP module have significant contributions to the proposed NR-IQA model.

E. Visual Analysis

To validate the effectiveness of our IQA model for learning prior knowledge from different distortions, we conduct a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3073410, IEEE Transactions on Circuits and Systems for Video Technology

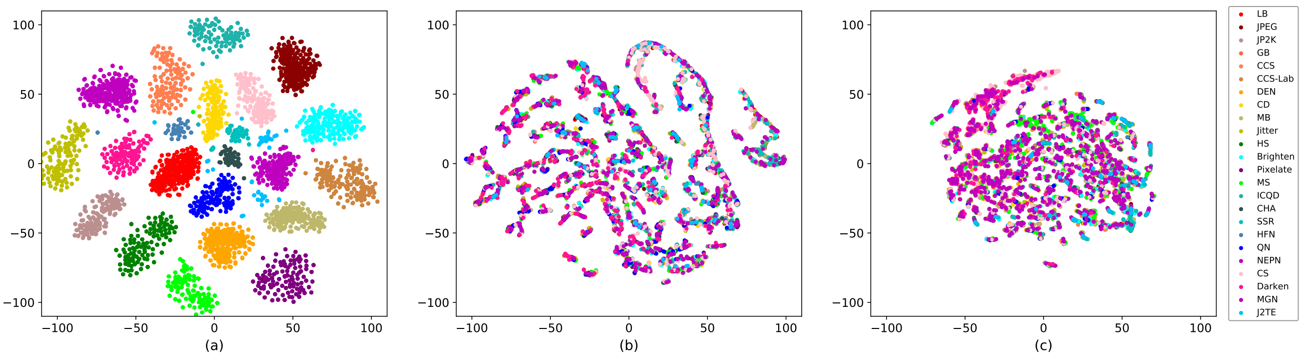


Fig. 8. The t-SNE 2D scatter plots of deep features extracted by three models on the merged database based on TID2013 and KADID-10k with 24 selected distortions: (a) The model trained by using a single distortion type, respectively; (b) The model trained by directly using all distortion types; (c) The proposed model based on meta-learning.

visual experiment by using the t-distributed stochastic neighbor embedding (t-SNE) algorithm [60]. Particularly, t-SNE is used to visualize the deep features of images with different distortion types, which come from the penultimate layer of our deep regression network f_θ optimized by three learning metrics. The first one is to train the network by using a single distortion type, respectively. The second one is to train a model by directly using multiple distortion types (called Multi-distortion model). The last one is our model based on deep meta-learning (MetaIQA+). The visualization is plotted on the merged database based on TID2013 and KADID-10k with 24 shared distortions. It's worth noting that the three models are all trained on the same deep regression network.

The t-SNE 2D scatter plots of deep features extracted by three models are shown in Fig. 8, where different colors represent different distortion types and each point represents an image. As can be seen, the clusters of different distortion types are naturally separated in Fig. 8(a). This indicates that there are huge differences among the NR-IQA models trained with diverse distortions. When the Multi-distortion model is trained by directly using all 24 distortion types, the clusters of different distortion types begin to mix in Fig. 8(b). However, there still exist some distortions that are clustered separately. In Fig. 8(c), all the distortion types are better mixed together compared with Fig. 8(b). This shows that the shared prior knowledge of varied distortions in images can be efficaciously captured through our MetaIQA+ model.

To quantitatively verify the generalization performance of the Multi-distortion model and MetaIQA+, we use different amounts of distortion to train these two models and plot the t-SNE scatter of models for 24 distortions. Then, the average distance of 24 distortions can be measured by the average result of the minimum distance from any dot of each distortion to points from the other distortions in t-SNE scatter plots. The average distance indicates the degree of dispersion in t-SNE scatter plots. The smaller the dispersion of the scattered points, the better the generalization performance of models for distortions. Fig. 9 shows the average distance of these two models for 24 distortions versus the number of distortions used in the training procedure. From the figure, the average distance of MetaIQA+ decreases faster than that of the Multi-distortion model when increasing the number

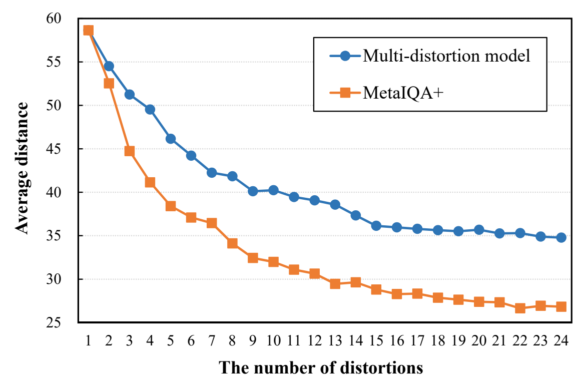


Fig. 9. The average distance of the Multi-distortion model and MetaIQA+ for 24 distortions versus the number of distortions used in the training procedure.

of distortions. This indicates that the t-SNE scatter plots of MetaIQA+ are more concentrated than the Multi-distortion model, which further proves that our MetaIQA+ model has a highly generalizable ability for unseen distortions.

V. Conclusion

In this paper, we propose a novel optimization-based deep meta-learning approach to address the generalization problem of NR-IQA. The proposed deep regression network with spatial pyramid pooling can effectively extract the features of the original image for quality evaluation. Since the proposed metric can capture the shared prior knowledge from a series of known distortions like human, the learned NR-IQA model can directly evaluate the quality of an image with unseen distortions. Also, we exploit a task selection approach based on gradient direction similarity to further reduce the requirement of training samples and improve the generalization ability of our NR-IQA model. Two experiments conducted on five public IQA databases have shown that the proposed metric outperforms the state-of-the-art NR-IQA metrics in terms of both evaluation accuracy and generalization ability. Besides, MetaIQA+ can also be used as an efficient prior model for adapting to NR-IQA tasks with authentic distortions, which will enlighten the design of NR-IQA models in future real-world applications.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3073410, IEEE Transactions on Circuits and Systems for Video Technology

References

[1] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, and A. C. Bovik, "Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 130–141, Nov. 2017.

[2] X. Zhang, S. Kwong, and C. . C. J. Kuo, "Data-driven transform based compressed image quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, 2020, DOI:10.1109/TCSVT.2020.3041639.

[3] H. Liang and D. S. Weller, "Comparison-based image quality assessment for selecting image restoration parameters," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5118–5130, Nov. 2016.

[4] Q. Wu, L. Wang, K. N. Ngan, H. Li, F. Meng, and L. Xu, "Subjective and objective de-raining quality assessment towards authentic rain image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3883–3897, Nov. 2020.

[5] H. G. Kim, H. Lim, and Y. M. Ro, "Deep virtual reality image quality assessment with human perception guider for omnidirectional image," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 4, pp. 917–928, Apr. 2020.

[6] W. Xia, Y. Yang, J. Xue, and J. Xiao, "Domain fingerprints for no-reference image quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, 2020, DOI:10.1109/TCSVT.2020.3002662.

[7] C. Lai and C. Chiu, "Using image processing technology for water quality monitoring system," in *Proc. IEEE Int. Conf. Mach. Learn. Cybern.*, vol. 4, Jul. 2011, pp. 1856–1861.

[8] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[9] S. Golestaneh and L. J. Karam, "Reduced-reference quality assessment based on the entropy of dwt coefficients of locally weighted gradient magnitudes," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5293–5303, Nov. 2016.

[10] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "MetaIQA: deep meta-learning for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 14 143–14 152.

[11] K. Ma, Z. Duanmu, Z. Wang, Q. Wu, W. Liu, H. Yong, H. Li, and L. Zhang, "Group maximum differentiation competition: Model comparison with few samples," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 851–864, Apr. 2020.

[12] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 529–539, Apr. 2010.

[13] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 39–50, Jan. 2016.

[14] L. Li, H. Zhu, G. Yang, and J. Qian, "Referenceless measure of blocking artifacts by tchebichef kernel analysis," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 122–125, Jan. 2014.

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

[16] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[17] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[18] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.

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

[20] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4444–4457, Sep. 2016.

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

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

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

[24] H. Talebi and P. Milanfar, "NIMA: neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.

[25] X. Liu, J. V. D. Weijer, and A. D. Bagdanov, "RankIQA: learning from rankings for no-reference image quality assessment," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1040–1049.

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

[27] K. Lin and G. Wang, "Hallucinated-IQA: No-reference image quality assessment via adversarial learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 732–741.

[28] S. Bosse, D. Maniry, K. Mller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan 2018.

[29] 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 Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018.

[30] B. Yan, B. Bare, and W. Tan, "Naturalness-aware deep no-reference image quality assessment," *IEEE Trans. Multimedia*, vol. 21, no. 10, pp. 2603–2615, Oct. 2019.

[31] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 36–47, Jan. 2020.

[32] J. Wu, J. Ma, F. Liang, W. Dong, G. Shi, and W. Lin, "End-to-end blind image quality prediction with cascaded deep neural network," *IEEE Trans. Image Process.*, vol. 29, pp. 7414–7426, Dec. 2020.

[33] H. Lin, V. Hosu, and D. Saupe, "DeepFL-IQA: weak supervision for deep iqa feature learning," *CoRR*, 2020. [Online]. Available: https://arxiv.org/abs/2001.08113

[34] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," *http://live. ece. utexas. edu/research/quality*, 2005.

[35] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, and F. Battisti, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process.-Image Commun.*, vol. 30, pp. 57–77, 2015.

[36] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal Image Video Process.*, vol. 12, no. 2, pp. 355–362, 2018.

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst.*, Dec. 2012, pp. 1097–1105.

[38] J. Vanschoren, "Meta-learning," in *Proc. Autom. Mach. Learn.*, 2019, pp. 35–61.

[39] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–34, Jun. 2020.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

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

[42] K. Ma, X. Liu, Y. Fang, and E. P. Simoncelli, "Blind image quality assessment by learning from multiple annotators," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2019, pp. 2344–2348.

[43] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. P. Lillicrap, "Meta-learning with memory-augmented neural networks," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2016, pp. 1842–1850.

[44] T. Munkhdalai and H. Yu, "Meta networks," in *Proc. Int. Conf. Mach. Learn.*, Aug. 2017, pp. 2554–2563.

[45] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," in *Proc. Neural Inf. Process. Syst.*, Dec. 2017, pp. 4080–4090.

[46] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1199–1208.

[47] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn.*, Aug. 2017, pp. 1126–1135.

[48] A. E. Eshratifar, D. Eigen, and M. Pedram, "Gradient agreement as an optimization objective for meta-learning," *CoRR*, 2018. [Online]. Available: https://arxiv.org/abs/1810.08178

[49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, May 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3073410, IEEE Transactions on Circuits and Systems for Video Technology

[50] H. Lin, V. Hosu, and D. Saupe, "KADID-10k: a large-scale artificially distorted IQA database," in *Proc. IEEE Int. Conf. Qual. Multimedia Experience*, Jun. 2019, pp. 1–3.

[51] T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen, and J. Häkkinen, "CID2013: a database for evaluating no-reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 390–402, Jan. 2015.

[52] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "KonIQ-10k: an ecologically valid database for deep learning of blind image quality assessment," *IEEE Trans. Image Process.*, vol. 29, pp. 4041–4056, 2020.

[53] D. Li, T. Jiang, and M. Jiang, "Unified quality assessment of in-the-wild videos with mixed datasets training," *Int. J. Comput. Vis.*, 2021, DOI:10.1007/s11263-020-01408-w.

[54] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *Proc. Neural Inf. Process. Syst. Workshop*, 2017.

[55] Y. Fang, R. Du, Y. Zuo, W. Wen, and L. Li, "Perceptual quality assessment for screen content images by spatial continuity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 4050–4063, Nov. 2020.

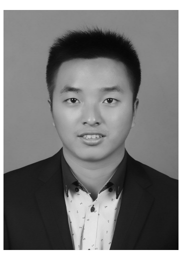
[56] S. Bosse, D. Maniry, T. Wiegand, and W. Samek, "A deep neural network for image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Aug. 2016, pp. 3773–3777.

[57] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 1, pp. 206–220, Feb. 2017.

[58] S. V. R. Dendi, C. Dev, N. Kothari, and S. S. Channappayya, "Generating image distortion maps using convolutional autoencoders with application to no reference image quality assessment," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 89–93, Jan. 2019.

[59] J. Kim, A. Nguyen, and S. Lee, "Deep cnn-based blind image quality predictor," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 11–24, Jan. 2019.

[60] L. V. D. Maaten and G. E. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 2605, pp. 2579–2605, Nov. 2008.



Hancheng Zhu received the B.S. degree from Changzhou Institute of Technology, Changzhou, China, in 2012, and the M.S. and Ph.D. degrees from China University of Mining and Technology, Xuzhou, China, in 2015 and 2020, respectively. He is currently a Postdoctoral Fellow in the School of Computer Science and Technology, China University of Mining and Technology, China. His research interests include image quality assessment and affective computing.

Leida Li received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2009, respectively. In 2008, he was a Research Assistant with the Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan. From 2014 to 2015, he was a Visiting Research Fellow with the Rapid-Rich Object Search (ROSE) Laboratory, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he was a Senior Research Fellow from 2016 to 2017. From 2009 to 2019, he worked in the School of Information and Control Engineering, China University of Mining and Technology, as Assistant Professor, Associate Professor and Professor, respectively. Currently, he is a Professor with the School of Artificial Intelligence, Xidian University.

His research interests include multimedia quality assessment, affective computing, information hiding, and image forensics. He has served as SPC for IJCAI 2019-2020, Session Chair for ICMR 2019 and PCM 2015, and TPC for AAAI 2019, ACM MM 2019-2020, ACM MM-Asia 2019, ACII 2019, PCM 2016. He is now an Associate Editor of the Journal of Visual Communication and Image Representation and the EURASIP Journal on Image and Video Processing.

Jinjian Wu received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2008 and 2013, respectively. From 2011 to 2013, he was a Research Assistant with Nanyang Technological University, Singapore, where he was a Post-Doctoral Research Fellow from 2013 to 2014. From 2015 to 2019, he was an Associate Professor with Xidian University, where he had been a Professor since 2019. His research interests include visual perceptual modeling, biomimetic imaging, quality evaluation, and object detection. He received the Best Student Paper Award at the ISCAS 2013. He has served as associate editor for the journal of Circuits, Systems and Signal Processing (CSSP), the Special Section Chair for the IEEE Visual Communications and Image Processing (VCIP) 2017, and the Section Chair/Organizer/TPC member for the ICME 2014-2015, PCM 2015-2016, ICIP 2015, VCIP 2018, and AAAI 2019.

Weisheng Dong received the B.S. degree in electronic engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2004, and the Ph.D. degree in circuits and system from Xidian University, Xi'an, China, in 2010. He was a Visiting Student with Microsoft Research Asia, Beijing, China, in 2006. From 2009 to 2010, he was a Research Assistant with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. In 2010, he joined Xidian University, as a Lecturer, where has been a Professor since 2016. He is currently with the School of Artificial Intelligence, Xidian University. His research interests include inverse problems in image processing, sparse signal representation, and image compression. He was a recipient of the Best Paper Award at the SPIE Visual Communication and Image Processing (VCIP) in 2010. He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and the SIAM Journal on Imaging Sciences.

Guangming Shi received the B.S. degree in automatic control, the M.S. degree in computer control, and the Ph.D. degree in electronic information technology from Xidian University, Xi'an, China, in 1985, 1988, and 2002, respectively. He had studied at the University of Illinois and University of Hong Kong. Since 2003, he has been a Professor with the School of Electronic Engineering, Xidian University. He was awarded the Cheung Kong Scholar Chair Professor by the Ministry of Education in 2012. He is currently the Academic Leader on circuits and systems, Xidian University. He has authored and coauthored more than 200 papers in journals and conferences. His research interests include compressed sensing, brain cognition theory, multirate filter banks, image denoising, low-bitrate image and video coding, and implementation of algorithms for intelligent signal processing. He served as the Chair for the 90th MPEG and 50th JPEG of the international standards organization, and Technical Program Chair for FSKD06, VSPC 2009, IEEE Pulse Code Modulation 2009, SPIE Visual Communications and Image Processing 2010, and IEEE International Symposium on Circuits and Systems 2013.