MetaIQA: Deep Meta-learning for No-Reference Image Quality Assessment

Hancheng Zhu[a,b], Leida Li[a*], Jinjian Wu[a], Weisheng Dong[a], and Guangming Shi[a]

[a]School of Artificial Intelligence, Xidian University
[b]School of Information and Control Engineering, China University of Mining and Technology
[a]ldli@xidian.edu.cn, [a]{jinjian.wu, wsdong}@mail.xidian.edu.cn, [a]gmshi@xidian.edu.cn
[b]zhuhancheng@cumt.edu.cn

Abstract

Recently, increasing interest has been drawn in exploiting deep convolutional neural networks (DCNNs) for no-reference image quality assessment (NR-IQA). Despite of the notable success achieved, there is a broad consensus that training DCNNs heavily relies on massive annotated data. Unfortunately, IQA is a typical small sample problem. Therefore, most of the existing DCNN-based IQA metrics operate based on pre-trained networks. However, these pre-trained networks are not designed for IQA task, leading to generalization problem when evaluating different types of distortions. With this motivation, this paper presents a no-reference IQA metric based on deep meta-learning. The underlying idea is to learn the meta-knowledge shared by human when evaluating the quality of images with various distortions, which can then be adapted to unknown distortions easily. Specifically, we first collect a number of NR-IQA tasks for different distortions. Then meta-learning is adopted to learn the prior knowledge shared by diversified distortions. Finally, the quality prior model is fine-tuned on a target NR-IQA task for quickly obtaining the quality model. Extensive experiments demonstrate that the proposed metric outperforms the state-of-the-arts by a large margin. Furthermore, the meta-model learned from synthetic distortions can also be easily generalized to authentic distortions, which is highly desired in real-world applications of IQA metrics.

1. Introduction

In recent years, the explosive growth of social networks has produced massive amounts of images. Digital images could be distorted in any stage of their life cycle, from acquisition, compression, storage to transmission, which leads to the loss of received visual information. Consequently, a reliable quality assessment metric of digital images is in great need to pick out high quality images for the end users. Although human's subjective evaluation of images is accurate and reliable, it is time-consuming and laborious in practical applications. Hence, objective image quality assessment (IQA) [20] is needed to imitate human beings to automatically assess image quality, which has extensive applications in image restoration [6], image retrieval [13] and image quality monitoring systems [23], etc.

Typically, IQA methods can be divided into three categories: full-reference IQA (FR-IQA) [17], reduced-reference IQA (RR-IQA) [11], and no-reference IQA (NR-IQA) [54], depending on the amount of reference information needed during quality evaluation. Although FR-IQA and RR-IQA methods can achieve promising performance, reference images are often not available in real-world situations. Hence, NR-IQA methods have attracted extensive attention recently, as they operate on the distorted images directly. Meantime, the lacking of reference information poses huge challenge for NR-IQA methods. Early NR-IQA methods mainly focus on specific distortion types, such as blocking artifacts [25], blur [24] and ringing effects [29]. The prerequisite of these approaches is that there is only one known type of distortion in the images. Since the distortion types are usually not known in advance in real-world applications, more and more attention has been drawn in general-purpose NR-IQA methods over the past few years [39, 32, 56, 51, 50, 12, 54, 49, 10]. These metrics attempt to characterize the general rules of image distortions through hand-crafted [39] or learned [54] features, based on which an image quality prediction model can be established.

Recent years have witnessed the great success of Deep Convolutional Neural Networks (DCNNs) [14] in many computer vision tasks [4, 5], which has also spawned a number of DCNNs-based NR-IQA approaches [16, 2, 28, 30, 52, 58, 57]. These approaches have achieved signifi-

*Corresponding author: Leida Li

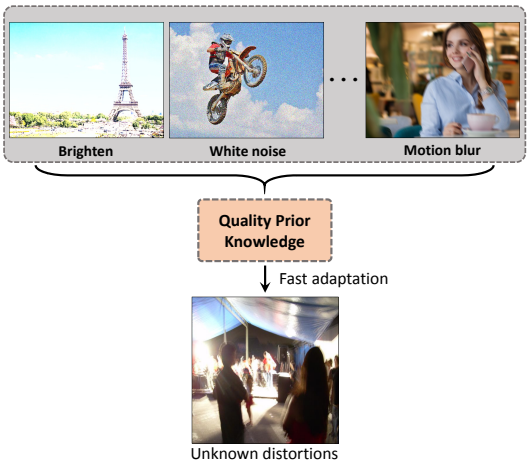

Figure 1. An illustration of our motivation. Humans can use the quality prior knowledge learned from various distortions (e.g. brighten, white noise, and motion blur) for fast adapting to unknown distortions (e.g. images captured by mobile cameras). Hence, it is necessary to make the NR-IQA model learn such quality prior knowledge to achieve high generalization performance.

cantly better performance than the traditional hand-crafted feature-based NR-IQA methods [39, 32, 56, 51, 50, 12]. The main reason is that DCNNs consist of massive parameters, which are helpful in learning the intricate relationship between image data and human perceived quality. At the same time, it is a broad sense that training DCNNs requires huge amount of annotated data. Unfortunately, collecting huge image quality data for training DCNNs-based IQA models is difficult, since annotating image quality by human is extremely expensive and time-consuming. As a result, the scale of existing annotated IQA databases [41, 38] is usually limited, thus training deep IQA models directly using these databases easily leads to over-fitting. To tackle the problem, existing works usually resort to pre-trained network models where big training data is available, e.g. ImageNet image classification task [1, 44]. Although these metrics can alleviate the over-fitting problem to some extent, the generalization performance is unsatisfactory when facing images with unknown distortions. In our opinion, this mainly attributes to the fact that the pre-trained models are not designed for IQA task, so they cannot easily adapt to new types of distortions.

In real-world situations, human beings can easily obtain quality prior knowledge from images with various distortions and quickly adapt to the quality evaluation of unknown distorted images, as shown in Figure 1. Therefore, it is critical for NR-IQA method to learn the shared prior knowledge of humans in evaluating the quality of images with various distortions. With this motivation, this paper presents a novel NR-IQA metric based on deep meta-learning that can make machines learn to learn, that is, to have the ability to learn quickly through a relatively small amount of training samples for a related new task [45, 48]. In particular, the proposed approach leverages a bi-level gradient descent strategy based on a number of distortion-specific NR-IQA tasks to learn a meta-model. The distortion-specific NR-IQA task is actually an IQA task for a specific distortion type (e.g., JPEG or blur). Different from the existing approaches, the learned meta-model can capture the shared meta-knowledge of humans when evaluating images with various distortions, enabling fast adaptation to the NR-IQA task of unknown distortions. The contributions of our work are summarized as follows.

- We propose a no-reference image quality metric based on deep meta-learning. Different from the existing IQA metrics, the proposed NR-IQA model is characterized by good generalization ability, in that it can perform well on diversified distortions.
- We adopt meta-learning to learn the shared meta-knowledge among different types of distortions when human evaluate image quality. This is achieved using bi-level gradient optimization based on a number of distortion-specific NR-IQA tasks. The meta-knowledge serves as an ideal pre-trained model for fast adapting to unknown distortions.
- We have done extensive experiments on five public IQA databases containing both synthetic and authentic distortions. The results demonstrate that the proposed model significantly outperforms the state-of-the-art NR-IQA methods in terms of generalization ability and evaluation accuracy.

2. Related Work

2.1. No-reference image quality assessment

NR-IQA can be classified into distortion-specific methods [25, 24, 29, 47] and general-purpose methods [39, 32, 56, 51, 50, 12, 54, 49, 10]. In distortion-specific methods, the image quality is evaluated by extracting features of known distortion types. This kind of metrics have achieved remarkable consistency with human perception. However, their application scope is rather limited, considering the fact that the distortion type is usually unknown in real applications [15, 9]. Thus, general-purpose NR-IQA approaches have received increasingly more attention recently [31]. Generally, conventional hand-crafted feature-based general-purpose NR-IQA methods can be divided into natural scene statistics (NSS) based metrics [8, 32, 33, 39] and learning-based metrics [53, 54, 36]. The NSS-based methods hold that natural images have certain statistical characteristics, which will be changed under different distortions. Moorthy *et al.* [33] proposed to extract NSS features from the discrete wavelet transform (DWT) domain for blind image quality assessment. Saad *et al.* [39] leveraged the statistical features of discrete cosine transform (DCT) to estimate

the image quality. Mittal *et al.* [32] proposed a general-purpose NR-IQA metric by extracting NSS features in the spatial domain and achieved promising performance. In additional to the NSS-based approaches, learning-based approaches have also been developed. The codebook representation approaches [53, 54] were proposed to predict subjective image quality scores by Support Machine Regression (SVR) model. Zhang *et al.* [36] combined the semantic-level features that influence human vision system with local features for image quality estimation.

In recent years, the deep learning-based general-purpose NR-IQA methods have demonstrated superior prediction performance over traditional methods [1, 44, 55, 28, 2, 30, 52, 58, 57]. One key issue of deep learning is that it requires abundant labeled data, but IQA is a typical small sample problem. In [1], Bianco *et al.* pre-trained a deep model on the large-scale database for image classification task and then fine-tuned it for NR-IQA task. Talebi *et al.* [44] proposed a DCNNs-based model by predicting the perceptual distribution of subjective quality opinion scores, and the model parameters were initialized by pre-training on ImageNet database [22]. Zeng *et al.* [55] also fine-tuned several popular pre-trained deep CNN models on IQA databases to learn a probabilistic quality representation (PQR). These approaches use the deep semantic features learned from image classification task as prior knowledge to assist in the learning of the NR-IQA task. However, image classification and quality assessment are quite different in nature, which leads to the generalization problem of deep NR-IQA models. In contrast to these approaches, in this paper, we take the advantage of meta-learning [45] to explore a more effective prior knowledge for the NR-IQA task.

2.2. Deep meta-learning

Deep meta-learning is a knowledge-driven machine learning framework, attempting to solve the problem of how to learn [45]. Human beings can effectively learn a new task from limited training data, largely relying on prior knowledge of related tasks. Meta-learning is to acquire a prior knowledge model by imitating this ability of human beings. Typically, meta-learning can be divided into three main approaches: Recurrent Neural Networks (RNNs) memory-based methods [40, 34], metric-based methods [42, 43] and optimization-based methods [7, 35]. The RNN memory-based methods use RNNs with memories to store experience knowledge from previous tasks for learning new task [40, 34]. The metric-based methods mainly learn an embedding function that maps the input space to a new embedding space, and leverage nearest neighbour or linear classifiers for image classification [42, 43]. The optimization-based methods aim to learn the initialization parameters of a model that can quickly learn new tasks by fine-tuning the model using few training samples [7, 35]. Although these methods are designed for few-shot learning in image classification task [48], the optimization-based method is easier to be extended because it is based on gradient optimization without limiting network structures [7]. Inspired by this, we propose an optimization-based meta-learning approach for NR-IQA task, which uses a number of distortion-specific NR-IQA tasks to learn the shared prior knowledge of various distortions in images. The NR-IQA task requires a quantitative measure of the perceptual quality of image, making it more complex and difficult than image classification task. Hence, we tailor a deep meta-learning with more efficient gradient optimization.

3. Our Approach

In this section, we detail our deep meta-learning approach for no-reference image quality assessment. The diversity of distortions in images leads to the generalization problem of deep NR-IQA models. In view of this, our approach leverages meta-learning to seek the general rules of image distortion through multiple distortion-specific NR-IQA tasks. That is, we learn a shared quality prior model through a number of NR-IQA tasks with known distortion types, and then fine-tune it for the NR-IQA task with unknown distortions. The overall framework of our approach is shown in Figure 2, which is composed of two steps, i.e., meta-training for quality prior model and fine-tuning for NR-IQA of unknown distortions. In the first step, we leverage a number of distortion-specific NR-IQA tasks to establish a meta-training set, which is further divided into two sets: support set and query set. Then, a bi-level gradient descent method from support set to query set is used to learn the quality prior model. In the second step, we fine-tune the quality prior model on a target NR-IQA task to obtain the quality model. Our method is termed Meta-learning based Image Quality Assessment (**MetaIQA**).

3.1. Meta-training for quality prior model

Shared quality prior knowledge among distortions. As mentioned in [31], most of the existing NR-IQA methods are distortion-aware, which are sensitive to image distortion types. Moreover, the available training data on current IQA databases cannot directly train an effective deep NR-IQA model. This limits the generalization ability of the trained NR-IQA model among images with different distortion types. Therefore, we need to learn a shared quality prior knowledge model from various distortions of images and make it quickly generalize to unknown distortions. Motivated by learning to learn in deep meta-learning [45], an optimization-based approach is introduced to learn the model parameters of shared quality prior knowledge from a number of NR-IQA tasks. For the NR-IQA task, we expect that the learned model can be quickly generalized to images with unknowable distortions. Hence, we use a two-level

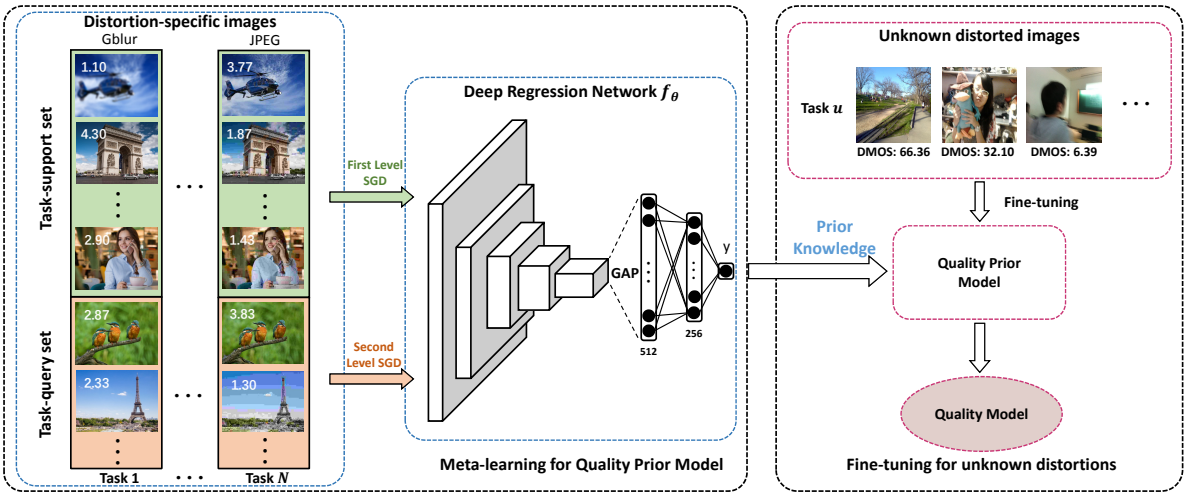

Figure 2. The overview framework of our deep meta-learning approach for no-reference image quality assessment.

gradient descent method to learn this generalization ability. First, the training data of each NR-IQA task is divided into support set and query set. Then, we use the support set to calculate the gradients of the model parameters and tentatively update them with stochastic gradient descent. Finally, the query set is used to verify whether the updated model is effectively performed or not. In this way, the model can learn the fast generalization ability among NR-IQA tasks with diversified distortions. The two-level gradient descent approach from support set to query set is called bi-level gradient optimization.

Meta-learning with bi-level gradient optimization. Since the optimization-based meta-learning method can be easily applied to any deep network using stochastic gradient descent, we introduce a deep regression network f_θ for the NR-IQA task. As shown in Figure 2, the deep regression network consists of convolutional layers and fully-connected layers. The convolutional layers derive from a popular deep network and we employ a Global Average Pooling (GAP) operation for yielding a fully-connected layer. Then, we add another fully-connected layer to generate the output of our deep regression network. In particular, for an input image x, we fed it into the deep network to generate the predicted quality score of the image $\hat{y}$, which is defined as

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

where θ denotes the initialized network parameters. Since we expect to minimize the difference between the predicted and ground-truth quality scores of the image x, the squared Euclidean distance is used as loss function, which takes the following form

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

where y denotes the ground-truth quality score of the input image x.

The purpose of our approach is to learn a shared prior model among various distortions when human evaluate image quality. Therefore, we obtain the meta-training set $\mathcal{D}_{meta}^{p(\tau)} = \{\mathcal{D}_s^{\tau_n}, \mathcal{D}_q^{\tau_n}\}_{n=1}^N$ from a number of distortion-specific NR-IQA tasks, where $\mathcal{D}_q^{\tau_n}$ and $\mathcal{D}_s^{\tau_n}$ are the support set and query set of each task, and N is the total number of tasks. In order to capture a generalized model among different NR-IQA tasks, we randomly sample k tasks as a mini-batch from the meta-training set ($1 < k \leq N$). For the i-th support set $\mathcal{D}_s^{\tau_i}$ in the mini-batch, the loss can be calculated by Eq. 2 and denoted as $\mathcal{L}_{\tau_i}$ ($i \in \{1, 2, \ldots, k\}$). Since our deep regression network is more complex than the classification network in [7] and there are more samples available for training, we leverage a more efficient stochastic gradient descent (SGD) approach to optimize our model. Therefore, we first calculate the first-order gradients of loss function $\mathcal{L}_{\tau_i}$ relating to all model parameters and it is defined as

$$g_\theta = \nabla_\theta \mathcal{L}_{\tau_i}(f_\theta). \tag{3}$$

Then, we update the model parameters for S steps using the Adam [21] optimizer on the support set $\mathcal{D}_s^{\tau_i} (i = 1, 2, ..., k)$, which is defined as

$$Adam(\mathcal{L}_{\tau_i}, \theta) : \theta_i^{'} \leftarrow \theta - \alpha \sum_{s=1}^{S} \frac{m_{\theta^{(s)}}}{\sqrt{v_{\theta^{(s)}}} + \epsilon}, \tag{4}$$

where $\epsilon = 1e-8$ and α is the inner learning rate. $m_{\theta^{(s)}}$ and $v_{\theta^{(s)}}$ denote the first moment and second raw moment of gradients, which are formulated as

$$m_{\theta^{(s)}} = \mu_1 m_{\theta^{(s-1)}} + (1 - \mu_1) g_{\theta^{(s)}}, \tag{5}$$

$$v_{\theta^{(s)}} = \mu_2 v_{\theta^{(s-1)}} + (1 - \mu_2) g_{\theta^{(s)}}^2, \tag{6}$$

where $m_{\theta^{(0)}} = 0$ and $v_{\theta^{(0)}} = 0$. μ_1 and μ_2 are the exponential decay rates of $m_{\theta^{(s)}}$ and $v_{\theta^{(s)}}$, respectively. $g_{\theta^{(s)}}$ denote

the updated gradients in step s ($s \in \{1, 2, ..., S\}$). As we mentioned previously, we expect that the quality model updated with the support set can perform well on the query set. In contrast to calculating second-order gradients in [7], we then compute the first-order gradients of updated model parameters for a second time to reduce the computational complexity of our model. The model parameters $\theta_i^{'}$ are updated with Adam optimizer for S steps on the query set $\mathcal{D}_q^{\tau_i} (i = 1, 2, ..., k)$, which takes the following form

$$Adam(\mathcal{L}_{\tau_i}, \theta_i^{'}) : \theta_i \leftarrow \theta_i^{'} - \alpha \sum_{s=1}^{S} \frac{m_{\theta'(s)}}{\sqrt{v_{\theta'(s)}} + \epsilon}, \quad (7)$$

where $m_{\theta'(s)}$ and $v_{\theta'(s)}$ are the first moment and second raw moment of gradients. For the mini-batch of k tasks, the gradients of all tasks are integrated to update the final model parameters, which is defined as

$$\theta \leftarrow \theta - \beta \frac{1}{k} \sum_{i=1}^{k} (\theta - \theta_i), \quad (8)$$

where β is the outer learning rate. With this approach, we iteratively sample k NR-IQA tasks on the meta-training set $\mathcal{D}_{meta}^{p(\tau)}$ to train our deep regression network f_θ. Finally, the quality prior model shared with various image distortions can be obtained by the meta-learning with bi-level gradient optimization.

3.2. Fine-tuning for unknown distortions

After training the quality prior model from a number of distortion-specific NR-IQA tasks, we then use this model as prior knowledge for fine-tuning on NR-IQA task with unknown distortions. Given M training images with annotated quality scores from a target NR-IQA task, we denote the predicted and ground-truth quality scores of i-th image as $\hat{y}_i$ and y_i ($i = 1, 2, ..., M$), respectively. We first use the squared Euclidean distance as loss function, which is formulated as

$$\mathcal{L} = \frac{1}{M} \sum_{i=1}^{M} \|\hat{y}_i - y_i\|_2^2. \quad (9)$$

Then, we leverage Adam optimizer to update the quality prior model for P steps on the NR-IQA task and it is defined as

$$Adam(\mathcal{L}, \theta) : \theta_{te} \leftarrow \theta - \alpha_f \sum_{p=1}^{P} \frac{m_{\theta(p)}}{\sqrt{v_{\theta(p)}} + \epsilon}, \quad (10)$$

where α_f is the learning rate of fine-tuning. $m_{\theta(p)}$ and $v_{\theta(p)}$ are first moment and second raw moment of gradients. Finally, the quality model can be obtained for assessing the quality of images with unknown distortions. It is worth noting that the fine-tuning process of our approach does not need to learn additional parameters, which greatly improves the learning efficiency and enhances the generalization ability of our model.

For a query image x, the predicted quality score $\hat{y}$ can be obtained by capturing the output of the quality model $\hat{y} = f_{\theta_{te}}(x; \theta_{te})$. The whole procedure of the proposed MetaIQA is summarized in Algorithm 1.

Algorithm 1 Meta-learning based IQA (MetaIQA)

Input: Meta-training set $\mathcal{D}_{meta}^{p(\tau)} = \{\mathcal{D}_s^{\tau_i}, \mathcal{D}_q^{\tau_i}\}_{i=1}^{N}$, where $\mathcal{D}_{tr_q}^{\tau_i}$ and $\mathcal{D}_{tr_s}^{\tau_i}$ are task-support set and task-query set, and N is the total number of tasks, a target NR-IQA task with M training images, query image x, learning rate β

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

```
Initialize model parameters θ;
/⋆ meta-training for prior model ⋆/
for iteration = 1, 2, ... do
    Sample a mini-batch of k tasks in D_meta^{p(τ)};
    for i = 1, 2, ..., k do
        /⋆ first level computing ⋆/
        Compute θ_i' = Adam(L_{τ_i}, θ) on D_s^{τ_i};
        /⋆ second level computing ⋆/
        Compute θ_i = Adam(L_{τ_i}, θ_i') on D_q^{τ_i};
    end for
    update θ ← θ − β (1/k) Σ_{i=1}^{k} (θ − θ_i);
end for
/⋆ fine-tuning for NR-IQA task ⋆/
Update θ_te = Adam(L, θ) on the NR-IQA task;
Input x into the quality model f_{θ_te};
return ŷ.
```

4. Experiments

4.1. Databases

We evaluate the performance of our approach on two kind of databases: synthetically distorted IQA databases and authentically distorted IQA databases.

Synthetically distorted IQA databases can be used for generating the meta-training set and evaluating the generalization performance of our quality prior model for unseen distortions, including TID2013 [38] and KADID-10K [27]. The information for each database is listed in Table 1.

Authentically distorted IQA databases are used to verify the generalization performance of our quality prior model for real distorted images, including CID2013 [46], LIVE challenge [9] and KonIQ-10K [26]. The CID2013 database contains six subsets with a total of 480 authentically distorted images captured by 79 different digital cameras. Subjects participated in the user study to evaluate the quality scores of images, which are in the range $[0, 100]$, and the higher the score, the higher the quality. The LIVE challenge

Table 1. Summary of synthetically distorted IQA databases with respect to numbers of reference images (Ref.), distortion images (Dist.), distortion types (Dist. Types) and score range. higher score indicates higher quality.

Databases	Ref.	Dist.	Dist. Types	Score Range
TID2013 [38]	25	3,000	24	[0, 9]
KADID-10K [27]	81	10,125	25	[1, 5]

database contains 1,162 images with authentic distortions taken from mobile cameras, such as motion blur, overexposure or underexposure, noise and JPEG compression. The quality scores of images were obtained by crowdsourcing experiments, which are in the range $[0,\ 100]$, and higher score indicates higher quality. Recently, a relatively large-scale IQA database, KonIQ-10K, consisting of 10,073 images was introduced in [26]. The quality score of each image is averaged by the five-point ratings of about 120 workers, which are in the range $[1,\ 5]$, and higher score also indicates higher quality.

4.2. Implementation details

In the proposed model, a popular network architecture, ResNet18 [14], is adopted as our backbone network. All training images are randomly cropped to 224×224 pixel patches for feeding into the proposed model. We train our model using bi-level gradient optimization with the inner learning rate α of $1e-4$ and the outer learning rate β of $1e-2$, which is implemented based on Pytorch [37]. We set the fine-tuning learning rate α_f to $1e-5$. These learning rates drop to a factor of 0.8 after every five epochs and the total epoch is 100. For both model training and fine-tuning, the weight decay is $1e-5$. The other hyper-parameters are set as follows: mini-batch size k of 5, exponential decay rate μ_1 of 0.9, exponential decay rate μ_2 of 0.99, learning steps S of 6, learning steps P of 15.

4.3. Evaluation criteria

In our experiments, Spearman's rank order correlation coefficient (SROCC) and Pearson's linear correlation coefficient (PLCC) are employed to evaluate the performance of NR-IQA methods [2, 52]. For N testing images, the PLCC is defined as

$$\text{PLCC} = \frac{\sum_{i=1}^{N}(s_i - \mu_{s_i})(\hat{s}_i - \mu_{\hat{s}_i})}{\sqrt{\sum_{i=1}^{N}(s_i - \mu_{s_i})^2}\sqrt{\sum_{i=1}^{N}(\hat{s}_i - \mu_{\hat{s}_i})^2}}, \quad (11)$$

where s_i and $\hat{s}_i$ denote the ground-truth and predicted quality scores of i-th image, and μ_{s_i} and $\mu_{\hat{s}_i}$ denote the average of each. Let d_i denote the difference between the ranks of i-th test image in ground-truth and predicted quality scores, the SROCC is defined as

$$\text{SROCC} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)}. \quad (12)$$

The PLCC and SROCC range from -1 to 1, and higher absolute value indicates better prediction performance.

4.4. Comparisons with the state-of-the-arts

Evaluation on synthetically distorted images. To validate the generalization performance of our meta-model for unknown distortions, we compare our method with six state-of-the-art general-purpose NR-IQA methods by using the Leave-One-Distortion-Out cross validation on TID2013 [38] and KADID-10K [27] databases. In implementation, suppose there are N kinds of distortions in a database, we use $(N-1)$ kinds of distortions for training and the remaining one kind of distortion is used for performance test. These methods are BLIINDS-II [39], BRISQUE [32], ILNIQE [56], CORNIA [54], HOSA [49] and WaDIQaM-NR [2]. For a fair comparison, all the source codes of NR-IQA methods released by original authors are conducted under the same training-testing strategy.

The tested SROCC values of our approach and state-of-the-art NR-IQA methods are listed in Table 2 and the best result for each distortion type is marked in bold. As can be seen, our approach is superior to other methods in overall performance (average results) on both databases by a large margin. For most of the distortion types (19 out of 24 on TID2013 and 19 out of 25 on KADID-10K), our method can achieve the best evaluation performance. In TID2013 database, the SROCC values of our method are higher than 0.9 for more than half of the distortion types, which indicates that our meta-learning based NR-IQA method can effectively learn a shared quality prior model and fast adapt to a NR-IQA task with unknown distortion types.

Generalization performance on authentically distorted images. To further evaluate the generalization performance of the quality prior model learned from synthetic distortions for the IQA of authentic distortions, we compare the proposed method with five state-of-the-art hand-crafted feature-based and six state-of-the-art deep learning-based general-purpose NR-IQA methods. The five hand-crafted feature-based NR-IQA methods are BLIINDS-II [39], BRISQUE [32], ILNIQE [56], CORNIA [54] and HOSA [49], while the six deep learning-based NR-IQA methods are BIECON [18], MEON [30], WaDIQaM-NR [2], DistNet-Q3 [3], DIQA [19] and NSSADNN [52]. For a fair comparison with the reported results of these methods on CID2013 [46], LIVE challenge [9] and KonIQ-10K [26] databases, we follow the same experimental setup in [2, 49, 52]. In CID2013 database, four out of six subsets are used for model training, and the remaining two subsets are used for testing. In LIVE challenge and KonIQ-10K databases, all images are randomly divided into 80% training samples and 20% testing samples. All experiments are conducted 10 times to avoid the bias of randomness and the average results of PLCC and SROCC are reported.

Table 2. SROCC values comparison in leave-one-distortion-out cross validation on TID2013 and KADID-10K databases.

	Dist. type	BLIINDS-II [39]	BRISQUE [32]	ILNIQE [56]	CORNIA [54]	HOSA [49]	WaDIQaM-NR [2]	**MetaIQA**
TID2013	AGN	0.7984	0.9356	0.8760	0.4465	0.7582	0.9080	**0.9473**
	ANC	0.8454	0.8114	0.8159	0.1020	0.4670	0.8700	**0.9240**
	SCN	0.6477	0.5457	0.9233	0.6697	0.6246	0.8802	**0.9534**
	MN	0.2045	0.5852	0.5120	0.6096	0.5125	**0.8065**	0.7277
	HFN	0.7590	0.8965	0.8685	0.8402	0.8285	0.9314	**0.9518**
	IN	0.5061	0.6559	0.7551	0.3526	0.1889	**0.8779**	0.8653
	QN	0.3086	0.6555	**0.8730**	0.3723	0.4145	0.8541	0.7454
	GB	0.9069	0.8656	0.8142	0.8879	0.7823	0.7520	**0.9767**
	DEN	0.7642	0.6143	0.7500	0.6475	0.5436	0.7680	**0.9383**
	JPEG	0.7951	0.5186	0.8349	0.8295	0.8318	0.7841	**0.9340**
	JP2K	0.8221	0.7592	0.8578	0.8611	0.5097	0.8706	**0.9586**
	JGTE	0.4509	0.5604	0.2827	0.7282	0.4494	0.5191	**0.9297**
	J2TE	0.7281	0.7003	0.5248	0.4817	0.1405	0.4322	**0.9034**
	NEPN	0.1219	0.3111	-0.0805	0.3571	0.2163	0.1230	**0.7238**
	Block	0.2789	0.2659	-0.1357	0.2345	0.3767	**0.4059**	0.3899
	MS	0.0970	0.1852	0.1845	0.1775	0.0633	**0.4596**	0.4016
	CTC	0.3125	0.0182	0.0141	0.2122	0.0466	0.5401	**0.7637**
	CCS	0.0480	0.2142	-0.1628	0.2299	-0.1390	0.5640	**0.8294**
	MGN	0.7641	0.8777	0.6932	0.4931	0.5491	0.8810	**0.9392**
	CN	0.0870	0.4706	0.3599	0.5069	0.3740	0.6466	**0.9516**
	LCNI	0.4480	0.8238	0.8287	0.7191	0.5053	0.6882	**0.9779**
	ICQD	0.7953	0.4883	0.7487	0.7757	0.8036	0.7965	**0.8597**
	CHA	0.5417	0.7470	0.6793	0.6937	0.6657	0.7950	**0.9269**
	SSR	0.7416	0.7727	0.8650	0.8867	0.8273	0.8220	**0.9744**
	Average	0.5322	0.5950	0.5701	0.5465	0.4725	0.7073	**0.8539**
KADID-10K	GB	0.8799	0.8118	0.8831	0.8655	0.8522	0.8792	**0.9461**
	LB	0.7810	0.6738	0.8459	0.8109	0.7152	0.7299	**0.9168**
	MB	0.4816	0.4226	0.7794	0.5323	0.6515	0.7304	**0.9262**
	CD	0.5719	0.5440	0.6780	0.2432	0.7272	0.8325	**0.8917**
	CS	-0.1392	-0.1821	0.0898	-0.0023	0.0495	0.4209	**0.7850**
	CQ	0.6695	0.6670	0.6763	0.3226	0.6617	**0.8055**	0.7170
	CSA1	0.0906	0.0706	0.0266	-0.0194	0.2158	0.1479	**0.3039**
	CSA2	0.6017	0.3746	0.6771	0.1197	0.8408	0.8358	**0.9310**
	JP2K	0.6546	0.5159	0.7895	0.3417	0.6078	0.5387	**0.9452**
	JPEG	0.4140	0.7821	0.8036	0.5561	0.5823	0.5298	**0.9115**
	WN	0.6277	0.7080	0.7757	0.3574	0.6796	0.8966	**0.9047**
	WNCC	0.7567	0.7182	0.8409	0.4183	0.7445	0.9247	**0.9303**
	IN	0.5469	-0.5425	0.8082	0.2188	0.2535	0.8142	**0.8673**
	MN	0.7017	0.6741	0.6824	0.3060	0.7757	0.8841	**0.9247**
	Denoise	0.4566	0.2213	0.8562	0.2293	0.2466	0.7648	**0.8985**
	Brighten	0.4583	0.5754	0.3008	0.2272	0.7525	0.6845	**0.7827**
	Darken	0.4391	0.4050	0.4363	0.2060	**0.7436**	0.2715	0.6219
	MS	0.1119	0.1441	0.3150	0.1215	**0.5907**	0.3475	0.5555
	Jitter	0.6287	0.6719	0.4412	0.7186	0.3907	0.7781	**0.9278**
	NEP	0.0832	0.1911	0.2178	0.1206	**0.4607**	0.3478	0.4184
	Pixelate	0.1956	0.6477	0.5770	0.5868	0.7021	0.6998	**0.8090**
	Quantization	0.7812	0.7135	0.5714	0.2592	0.6811	0.7345	**0.8770**
	CB	-0.0204	0.0673	0.0029	0.0937	0.3879	0.1602	**0.5132**
	HS	-0.0151	0.3611	**0.6809**	0.1142	0.2302	0.5581	0.4374
	CC	0.0616	0.1048	0.0723	0.1253	**0.4521**	0.4214	0.4377
	Average	0.4328	0.4136	0.5528	0.3149	0.5598	0.6295	**0.7672**

In our approach, we first normalize the subjective scores of images on TID2013 and KADID-10K databases to $[0,\ 1]$ and then use the generated NR-IQA tasks to train our network for obtaining a quality prior model. Finally, we fine-tune the quality prior model on the training set of CID2013, LIVE challenge and KonIQ-10K. Table 3 summarizes the testing results on the three IQA databases and the best results among the NR-IQA methods for each database are shown boldfaced. We can see that our approach achieves the best evaluation performance on LIVE challenge and KonIQ-10K databases. Our method and NSSADNN have achieved comparable results on CID2013 database, which are significantly better than other NR-IQA methods. This indicates that our method based on meta-learning can capture the quality prior model shared by human when evaluating the perceived quality images with various synthesized distortions, and then quickly adapt to a NR-IQA task with authentic distortions.

4.5. Visual analysis for quality prior model

In this section, we performed a visual experiment to demonstrate the effectiveness of our quality prior model. Particularly, we use a CNN visualization code[1] to show the gradient maps in pixel-wisely with various distortions. We learn the quality prior model from distortion-specific images on TID2013 and KADID-10K databases, and then randomly select four severely distorted images in the LIVE challenge database for visualization experiment. The images as well as the corresponding gradient maps are shown in Figure 3. As can be seen, the gradient maps can accurately capture the location of authentic distortions in images, such as overexposure in Figure 3(a), underexposure in Figure 3(b), motion blur in Figure 3(c) and noise in Figure 3(d). This strongly demonstrates that the shared prior knowledge of various distortions in images can be effectively learned from a number of NR-IQA tasks through meta-learning.

[1] https://github.com/sar-gupta/convisualize_nb

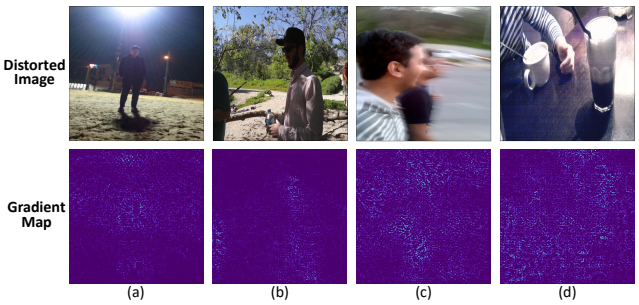

Figure 3. The gradient maps of some authentically distorted images in LIVE challenge database.

Table 3. Comparison results (PLCC and SROCC) of our approach with several state-of-the-art NR-IQA methods on authentically distorted IQA databases (i.e., CID2013 [46], LIVE challenge [9] and KonIQ-10K [26]).

Methods	CID2013		LIVE challenge		KonIQ-10K	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
BLIINDS-II [39]	0.565	0.487	0.507	0.463	0.615	0.529
BRISQUE [32]	0.648	0.615	0.645	0.607	0.537	0.473
ILNIQE [56]	0.538	0.346	0.589	0.594	0.537	0.501
CORNIA [54]	0.680	0.624	0.662	0.618	0.795	0.780
HOSA [49]	0.685	0.663	0.678	0.659	0.813	0.805
BIECON [18]	0.620	0.606	0.613	0.595	/	/
MEON [30]	0.703	0.701	0.693	0.688	/	/
WaDIQaM-NR [2]	0.729	0.708	0.680	0.671	0.761	0.739
DistNet-Q3 [3]	/	/	0.601	0.570	0.710	0.702
DIQA [19]	0.720	0.708	0.704	0.703	/	/
NSSADNN [52]	**0.825**	0.748	0.813	0.745	/	/
MetaIQA	0.784	**0.766**	**0.835**	**0.802**	**0.887**	**0.850**

Table 4. Ablation study results (PLCC and SROCC) on authentically distorted IQA databases (i.e., CID2013 [46], LIVE challenge [9] and KonIQ-10K [26]).

Methods	CID2013		LIVE challenge		KonIQ-10K	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
Baseline	0.727	0.712	0.801	0.743	0.832	0.816
MetaIQA	**0.784**	**0.766**	**0.835**	**0.802**	**0.887**	**0.850**

4.6. Ablation study

To further investigate whether the effectiveness of our approach is derived from meta-learning, we conduct ablation studies in this experiment. The baseline method is to first train our network model by directly using the Adam optimizer on distortion-specific images, and then fine-tune the model on the training set of authentically distorted images (called Baseline). It is worth noting that baseline method and our method have the same number of network parameters but are trained by two different optimization approaches. The results of all tested images on three authentically distorted IQA databases are summarized in Table 4. From the results, we can see that our MetaIQA method is superior to Baseline method by a large margin on all databases. Compared with the baseline approach, MetaIQA has better generalization performance and can improve the performance of NR-IQA model without changing the network structure. This demonstrates the effectiveness of our method in dealing with the NR-IQA task.

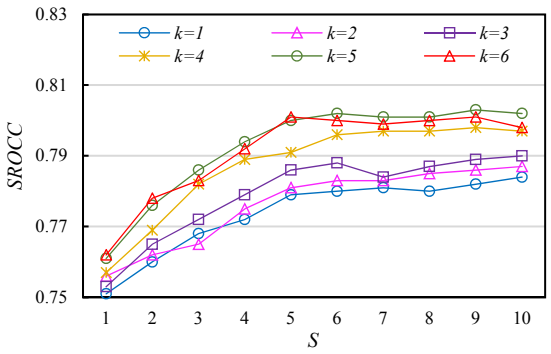

Figure 4. The efficacy of parameters k and S in meta-training on LIVE challenge database measured by SROCC.

4.7. Parameters discussion

Finally, we conduct experiments to discuss the efficacy of two key parameters in the meta-training of our approach, i.e. k to control the number of NR-IQA tasks in a mini-batch and S to control the learning steps of each task. We set k and S to different values and show the SROCC results on LIVE challenge database in Figure 4. The quality evaluation performance of our approach increases with the increase of k and S. If k is larger than 5, the SROCC values of our method drop slightly. When S increases from 1 to 6, the performance of quality evaluation increases dramatically. If S is larger than 6, the SROCC values tend to be stable. Therefore, we set $k = 5$ and $S = 6$ in all the experiments.

5. Conclusion

In this paper, we propose to address the generalization problem of NR-IQA tasks by using meta-learning. We introduce a meta-learning based NR-IQA method with bi-level gradient optimization to learn the shared prior knowledge model of various distortions from a number of NR-IQA tasks, and then fine-tune the prior model on the training data of a NR-IQA task with unknown distortions to obtain the target quality model. Since our model can refine the shared meta-knowledge among various types of distortions when human evaluate image quality, the learned meta-model is easily generalized to unknown distortions. Experiments conducted on five public IQA databases have demonstrated that our approach is superior to the state-of-the-art NR-IQA methods in terms of both generalization ability and evaluation accuracy. In addition, the quality prior model learned from synthetic distortions can also be quickly adapted to the quality assessment of authentically distorted images, which also sheds light on the design of quality evaluation models for real-world applications.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (61771473, 61991451 and 61379143), Natural Science Foundation of Jiangsu Province (BK20181354), the Six Talent Peaks High-level Talents in Jiangsu Province (XYDXX-063).

References

[1] S. Bianco, L. Celona, P. Napoletano, and R. Schettini. On the use of deep learning for blind image quality assessment. *Signal Image Video Process.*, 12(2):355–362, 2018.

[2] 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.*, 27(1):206–219, Jan 2018.

[3] 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.*, 26(1):89–93, Jan. 2019.

[4] G. Ding, W. Chen, S. Zhao, J. Han, and Q. Liu. Real-time scalable visual tracking via quadrangle kernelized correlation filters. *IEEE Trans. Intell. Transp. Syst.*, 19(1):140–150, Jan. 2018.

[5] G. Ding, Y. Guo, K. Chen, C. Chu, J. Han, and Q. Dai. DECODE: deep confidence network for robust image classification. *IEEE Trans. Image Process.*, 28(8):3752–3765, Aug. 2019.

[6] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(2):295–307, Feb. 2016.

[7] C. Finn, P. Abbeel, and S. Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning (ICML)*, pages 1126–1135, Aug. 2017.

[8] X. Gao, F. Gao, D. Tao, and X. Li. Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning. *IEEE Trans. Neural Netw. Learn. Syst.*, 24(12):2013–2026, Dec. 2013.

[9] D. Ghadiyaram and A. C. Bovik. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Trans. Image Process.*, 25(1):372–387, Jan. 2016.

[10] D. Ghadiyaram and A. C. Bovik. Perceptual quality prediction on authentically distorted images using a bag of features approach. *J. Vision*, 17(1):32–32, 2017.

[11] 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.*, 25(11):5293–5303, Nov. 2016.

[12] K. Gu, G. Zhai, X. Yang, and W. Zhang. Using free energy principle for blind image quality assessment. *IEEE Trans. Multimedia*, 17(1):50–63, Jan 2015.

[13] Y. Guo, G. Ding, and J. Han. Robust quantization for general similarity search. *IEEE Trans. Image Process.*, 27(2):949–963, Feb. 2018.

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, Jun. 2016.

[15] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik. Objective quality assessment of multiply distorted images. In *Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 1693–1697, Nov. 2012.

[16] L. Kang, P. Ye, Y. Li, and D. Doermann. Convolutional neural networks for no-reference image quality assessment. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1733–1740, Jun. 2014.

[17] J. Kim and S. Lee. Deep learning of human visual sensitivity in image quality assessment framework. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1969–1977, Jul. 2017.

[18] J. Kim and S. Lee. Fully deep blind image quality predictor. *IEEE J. Sel. Top. Signal Process.*, 11(1):206–220, Feb. 2017.

[19] J. Kim, A. Nguyen, and S. Lee. Deep cnn-based blind image quality predictor. *IEEE Trans. Neural Netw. Learn. Syst.*, 30(1):11–24, Jan. 2019.

[20] 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.*, 34(6):130–141, Nov. 2017.

[21] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference for Learning Representations (ICLR)*, May 2015.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1097–1105, Dec. 2012.

[23] C. Lai and C. Chiu. Using image processing technology for water quality monitoring system. In *International Conference on Machine Learning and Cybernetics (ICMLC)*, volume 4, pages 1856–1861, Jul. 2011.

[24] 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.*, 46(1):39–50, Jan. 2016.

[25] L. Li, H. Zhu, G. Yang, and J. Qian. Referenceless measure of blocking artifacts by tchebichef kernel analysis. *IEEE Signal Process. Lett.*, 21(1):122–125, Jan. 2014.

[26] H. Lin, V. Hosu, and D. Saupe. Koniq-10k: Towards an ecologically valid and large-scale iqa database. *CoRR*, 2018.

[27] H. Lin, V. Hosu, and D. Saupe. Kadid-10k: A large-scale artificially distorted iqa database. In *IEEE International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3, Jun. 2019.

[28] K. Lin and G. Wang. Hallucinated-IQA: No-reference image quality assessment via adversarial learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 732–741, Jun. 2018.

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

[30] 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.*, 27(3):1202–1213, Mar. 2018.

[31] K. Ma, X. Liu, Y. Fang, and E. P. Simoncelli. Blind image quality assessment by learning from multiple annotators. In *IEEE International Conference on Image Processing (ICIP)*, pages 2344–2348, Sep. 2019.

[32] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.*, 21(12):4695–4708, Dec 2012.

[33] A. K. Moorthy and A. C. Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal Process. Lett.*, 17(5):513–516, May 2010.

[34] T. Munkhdalai and H. Yu. Meta networks. In *International Conference on Machine Learning (ICML)*, pages 2554–2563, Aug. 2017.

[35] A. Nichol, J. Achiam, and J. Schulman. On first-order meta-learning algorithms. *CoRR*, 2018.

[36] P. Zhang, W. Zhou, L. Wu, and H. Li. SOM: Semantic obviousness metric for image quality assessment. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2394–2402, Jun. 2015.

[37] 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 *Advances in Neural Information Processing Systems (NIPS) Workshop*, 2017.

[38] 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.*, 30:57–77, 2015.

[39] 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.*, 21(8):3339–3352, Aug. 2012.

[40] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. P. Lillicrap. Meta-learning with memory-augmented neural networks. In *International Conference on Machine Learning (ICML)*, pages 1842–1850, Jun. 2016.

[41] 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.

[42] J. Snell, K. Swersky, and R. S. Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems (NIPS)*, pages 4080–4090, Dec. 2017.

[43] 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 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1199–1208, Jun. 2018.

[44] H. Talebi and P. Milanfar. NIMA: neural image assessment. *IEEE Trans. Image Process.*, 27(8):3998–4011, Aug. 2018.

[45] J. Vanschoren. Meta-learning: A survey. *CoRR*, 2018.

[46] 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.*, 24(1):390–402, Jan. 2015.

[47] G. Wang, Z. Wang, K. Gu, L. Li, Z. Xia, and L. Wu. Blind quality metric of dibr-synthesized images in the discrete wavelet transform domain. *IEEE Trans. Image Process.*, 29:1802–1814, 2020.

[48] Y. Wang and Q. Yao. Few-shot learning: A survey. *CoRR*, 2019.

[49] 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.*, 25(9):4444–4457, Sep. 2016.

[50] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. *IEEE Trans. Image Process.*, 23(11):4850–4862, Nov. 2014.

[51] W. Xue, L. Zhang, and X. Mou. Learning without human scores for blind image quality assessment. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 995–1002, Jun. 2013.

[52] B. Yan, B. Bare, and W. Tan. Naturalness-aware deep no-reference image quality assessment. *IEEE Trans. Multimedia*, 21(10):2603–2615, Oct. 2019.

[53] P. Ye and D. Doermann. No-reference image quality assessment using visual codebooks. *IEEE Trans. Image Process.*, 21(7):3129–3138, Jul. 2012.

[54] P. Ye, J. Kumar, L. Kang, and D. Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1098–1105, Jun. 2012.

[55] H. Zeng, L. Zhang, and A. C. Bovik. A probabilistic quality representation approach to deep blind image quality prediction. *CoRR*, 2017.

[56] L. Zhang, L. Zhang, and A. C. Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Trans. Image Process.*, 24(8):2579–2591, Aug. 2015.

[57] 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.*, 30(1):36–47, Jan 2020.

[58] W. Zhang, K. Ma, and X. Yang. Learning to blindly assess image quality in the laboratory and wild. *CoRR*, 2019.