Semi-Supervised Authentically Distorted Image Quality Assessment with Consistency-Preserving Dual-Branch Convolutional Neural Network

Guanghui Yue, *Member*, *IEEE*, Di Cheng, Leida Li, *Member*, *IEEE*, Tianwei Zhou, Hantao Liu, *Member*, *IEEE*, and Tianfu Wang

***Abstract*—Recently, convolutional neural networks (CNNs) have provided a favoured prospect for authentically distorted image quality assessment (IQA). For good performance, most existing CNN-based methods rely on a large amount of labeled data for training, which is time-consuming and cumbersome to collect. By simultaneously exploiting few labeled data and many unlabeled data, we make a pioneering attempt to propose a semi-supervised framework (termed SSLIQA) with consistency-preserving dual-branch CNN for authentically distorted IQA in this paper. The proposed SSLIQA introduces a consistency-preserving strategy and transfers two kinds of consistency knowledge from the teacher branch to the student branch. Concretely, SSLIQA utilizes the sample prediction consistency to train the student to mimic output activations of individual examples represented by the teacher. Considering that subjects often refer to previous analogous cases to make scoring decisions, SSLIQA computes the semantic relation among different samples in a batch and encourages the consistency of sample semantic relation between two branches to explore extra quality-related information. Benefiting from the consistency-preserving strategy, we can exploit numerous unlabeled data to improve network's effectiveness and generalization. Experimental results on three authentically distorted IQA databases show that the proposed SSLIQA is stably effective under different student-teacher combinations and different labeled-to-unlabeled data ratios. In addition, it points out a new way on how to achieve higher performance with a smaller network.**

***Index Terms*—Image quality assessment, authentical distortion, consistency-preserving, semi-supervised.**

I. Introduction

NOWADAYS, digital images are ubiquitous in almost every aspect of daily life due to the popularity of portable digital cameras and rapid development of social media. On a routine day, several billion images are uploaded and shared on social media platforms such as Google, Flicker, Twitter, Facebook, Instagram, etc. However, the real-world images captured by inferior devices and naive photographers can undergo multiple distortions, e.g., noise corruption, diverse blurs, under-/over-exposure, compression artifacts, bringing negative visual quality to the viewing audiences [1]. Therefore, to maintain the quality of service, it is highly in-demand to design effective quality assessment methods for real-world captured images with authentic distortions [2].

Since no pristine information is available, the authentically distorted image quality assessment (IQA) can only be operated in a no-reference (NR) manner. In the past decades, tremendous efforts have been devoted to designing NR-IQA methods [3]–[8]. Generally speaking, existing NR-IQA methods can be roughly categorized as conventional NR-IQA and convolutional neural network (CNN) based NR-IQA.

Conventional NR-IQA proposes to summarize the distortion characteristics with handcrafted features and to map them to a quality score. Early works applied a two-stage paradigm that extracted statistical features to identify distortion first, followed by distortion-specific quality integration [9]–[11]. Later, more general methods were reported to comprehensively evaluate images degraded by different kinds of distortions. Such methods empirically selected features from diverse aspects, such as natural scene statistic (NSS) [12], structure and luminance analyses [13], gradient variation [14], codebook construction [15], and free energy modeling [16], to distinguish the distorted images from the high-quality ones without distortion identification. Despite remarkable successes achieved, the aforementioned methods were accomplished in evaluating synthesized distortions, yet showed limited ability in representing the essence of complex authentic distortions. Recently, several attempts were made to simplify the authentic distortions to a group of hybrid synthesized distortions and explored effective features to represent them [17], [18]. For more favoured performance, extensive features were required to capture the distortions in an all-rounded way [19]. Although these attempts have successfully promoted the research of authentically distorted IQA to a certain extent, the performance grows slowly, exposing the limitation of representing authentic distortions by handcrafted features.

This work was supported in part by National Natural Science Foundation of China (Nos. 62001302, 62103286), in part by Guangdong Basic and Applied Basic Research Foundation (Nos. 2019A1515111205, 2021A1515011348, 2019A1515110401), in part by Natural Science Foundation of Shenzhen (No. JCYJ20190808145011259), in part by Shenzhen Science and Technology Program (No. RCBS20200714114920379), in part by Tencent Rhinoceros Birds - Scientific Research Foundation for Young Teachers of Shenzhen University, and in part by Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under Grant VRLAB2021C05. (Corresponding author: Tianwei Zhou.)

G. Yue, D. Cheng, and T. Wang are with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, the School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China. (email: yueguanghui@szu.edu.cn; chengdigogogo@outlook.com; tfwang@szu.edu.cn.)

L. Li is with the School of Artificial Intelligence, Xidian University, Xi'an 710071, China. (e-mail: ldli@xidian.edu.cn.)

T. Zhou is with the College of Management, Shenzhen University, Shenzhen 518060, China. (e-mail: tianwei@szu.edu.cn.)

H. Liu is with the School of Computer Science and Informatics, Cardiff University, Cardiff CF243AA, U.K (e-mail: liuh35@cardiff.ac.uk).

Compared with the conventional NR-IQA, CNN-based NR-IQA is conducive to obtaining better performance due to the strong ability in learning and fusing discriminative features [20]–[23]. In spite of this, this kind of methods requires a great amount of labeled images for stable and satisfactory performance. Unfortunately, existing IQA databases are commonly small-scale, which is propitious for properly training an IQA model. To address such a challenge, a pioneering work conducted by Kang *et al*. divided the image into multiple small pathes and directly labeled them with the quality score of the whole image [20]. In this way, the training set was greatly expanded. Inspired by this, more works were conducted by either redesigning the network structure, changing the patch weighting strategy, or utilizing the proxy label [21], [24]–[26]. It has been widely acknowledged that both local and global information are important for achieving a comprehensive visual quality [27]–[30]. Only utilizing local pathes as the network's inputs will lose the global information. To tackle this issue, subsequent scholars implemented the cropping and resizing operations on the whole image to support larger inputs for the network [31]–[33]. Inspired by [34], efforts were made to split the complex authentic distortion into several simple synthesized distortions and to label the generated image with relative quality [35]. Through controlling the distortion degree, a large image sets could be generated to train the network by using the relative image quality. However, the distortion decomposion was subjective and highly dependent on designer's experience.

Overall, CNN-based NR-IQA methods receive growing favor from scholars with advantages of less experience dependence and automatic quality-related feature extraction/fusion. They evade the limitations of conventional NR-IQA methods and gradually become the promising alternatives for authentically distorted IQA. Nevertheless, such methods are data-hungry. How to felicitously address the conflict between the small scale of IQA databases and the large scale of training data required is still under discussion. Current solutions are not ideal. On the one hand, since the authentic distortions are not uniformly distributed in the spatial domain, it is unsuitable to arrange each local patch with the same quality as the whole image. Also, it is impossible to obtain the proxy labels via full-reference (FR) IQA methods as there is no high-quality counterparts of authentically distorted images. On the other hand, since the authentic distortions are usually complex, it is hard to completely model the authentic distortions by a group of synthesized ones. Utilizing the synthesized images may lead to the learned model sub-optimal as there are some differences between the synthesized images and real-world authentically distorted images.

In this study, we propose a novel semi-supervised framework, termed SSLIQA, for authentically distorted IQA. Our SSLIQA follows an asymmetric parallel dual-branch structure and trains the network in a semi-supervised learning manner by preserving the consistency between two branches. It has two kinds of advantages compared with previous works and handles the challenges in the authentically distorted IQA task well. First, benefiting from the consistency-preserving strategy, it can simultaneously exploit few labeled data and many unlabeled data with real-world authentic distortions to train the network, thereby improving the network's effectiveness and generalization. This avoids the drawbacks raised by using the synthesized images for training in previous works. One highlight in the consistency-preserving strategy is that, inspired by the subjective scoring behaviors, we introduce a batch-level consistency loss to consider the semantic relation between images within each batch. Such a loss is rarely reported in previous IQA methods and contributes to improving the performance, as discussed in Section IV-D3. Second, thanks to the asymmetric parallel dual-branch structure, the small-size student branch can mimic the behaviors of the large-size teacher branch well. In this way, we only require the small-size student in the inference stage and achieve higher performance with a smaller network. Generally, a smaller IQA network with higher performance is more favoured for the computationally limited platforms. However, such a problem has been rarely discussed in previous works. Extensive experiments on three public authentically distorted IQA databases demonstrate the superiority of the proposed SSLIQA against twelve state-of-the-art NR-IQA methods. The contributions of this study are summarized below.

- To cope with the challenge of limited labeled data, we propose a pioneering semi-supervised NR-IQA framework with dual branches to exploit images with and without human annotations for evaluating the quality of authentically distorted images. With the assistance of unlabeled images, the proposed framework can achieve higher prediction accuracy and generalization compared to the competing methods.
- In view of the subjective scoring behaviors, we propose a consistency-preserving strategy to transfer quality-related knowledge from the teacher branch to the student branch by explicitly encouraging both sample-level and batch-level consistencies between two branches. The sample-level consistency enforces the student to mimic output activations of individual examples represented by the teacher. The batch-level consistency matches the sample semantic relation in a given batch between two branches to explore extra quality-related information.
- We propose a new asymmetric parallel structure by incorporating a large teacher branch and a small student branch in a NR-IQA framework. Through collaborative training of two branches, the performance of student is improved and approaches to that of teacher. Contrary to current NR-IQA methods that usually ignore the amount of network parameters but blindly chase better performance, our solution provides us a new way on how to achieve higher performance with a smaller network.

The remainder of this article is organized as follows. Section II briefly reviews the related works in NR-IQA. Section III details the proposed SSLIQA for authentically distorted images. Section IV introduces the experiments in detail, including the experimental settings, results and discussions. Finally, Section V draws the concluding remarks of this study.

II. Related Work

A. Conventional NR-IQA Methods

In the past decades, how to design effective NR-IQA methods has been widely discussed by the IQA community. Several efforts have been made and many conventional NR-IQA methods have been proposed [4], [5]. One basic consensus among these methods is to effectively analyze and represent the characteristics of diverse distortions by handcrafted features and to make the final decision about image quality based on these features.

Generally, conventional NR-IQA methods can be divided into specific-purpose and general-purpose methods. The former knows the distortion type in advance, and puts forward some features that can represent the distortion intensity. For instance, Li *et al*. [36] utilized the Tchebichef moments to score the the blocking artifacts. Considering that the blurriness usually degrades the edge information, Wang *et al*. [37] fitted the distribution of image gradient magnitudes to model the properties of blurred images. Yue *et al*. [38] introduced a blurriness assessment metric via analysis of the local binary pattern features. Driven by the observation that the contrast change usually brings under-/over-exposure, utilizing the entropy to measure the local or global information is widely used in the contrast change IQA task [39]–[41].

The general-purpose methods aim to evaluate various distortions in a unified way, which brings great challenges for quality-aware feature selection. In recent literature, it has been widely acknowledged that the synthesized distortions can be reflected by the destruction from NSS. In early works, NSS features were extracted from diverse transform domains, such as the discrete cosine transform domain [10], [42] and the wavelet transform domain [43]. To avoid the computational burden costed in domain transformation, later works tended to extract NSS features in the spatial domain [12], [44]. Since authentic distortions are more complex than synthesized distortions, mining NSS features from more domains, e.g., color space, gradient space, would be conducive to achieving higher performance [19]. Apart from NSS features, scholars also attempted to construct visual codebooks to store quality-aware features. For example, Ye *et al*. [45] exploited a large amount of local patches to construct the codebook by means of unsupervised learning. In the inference stage, the quality of query image could be directly obtained without assuming any specific types of distortions. Similarity, Jiang *et al*. [15] simultaneously constructed two visual dictionaries by keeping the relationship between local patches and its qualities, yielding a better prediction accuracy.

In summary, conventional NR-IQA methods have proven effective in evaluating synthetically distorted images. The effectiveness of such works highly depends on the selection of the handcrafted features. However, their further implementation in authentically distorted IQA is restricted by many obstacles. On the one hand, since authentic distortions are quite complex, it is hard to completely represent the distortion characteristics by limited handcrafted features derived from specific prior knowledge. On the other hand, blindly expanding the number of extracted features will increase the training difficulties during IQA model generation and be easy to lead to over-fitting. To this end, it is highly desired to propose more advanced methods with high accuracy but less experience dependence.

B. CNN-Based NR-IQA Methods

In recent years, CNN-based methods have shown greater advantages than conventional NR-IQA methods and gradually become promising alternatives for tackling the challenging NR-IQA task [21]. How to cope with the insufficient labeled data for network training is one of the key challenges that hinder the further development of CNN-based methods.

One straightforward idea is directly dividing the image into multiple patches and labeling each patch with the quality of the whole image [20]. Obviously, such an idea ignores the fact that the image quality varies in different regions. In view of this, Kim *et al*. [46] applied FR-IQA methods to form proxy labels for local patches, and integrated the obtained local scores to generate the whole image quality. Different from [46], Bosse *et al*. [47] proposed a deeper network and utilized a weighting strategy to integrate local scores. Similarly, Jiang *et al*. [48] proposed to fuse the local scores with an adaptive weighting method by considering the effect of image patch contents. Considering that not all image patches are useful for training the CNN model, they also proposed a strategy to select effective patches by calculating the Euclidean distance between subjective score of the whole image and predicted scores of patches [49]. Since distortion easily damages the image structure information, Pan *et al*. [24] also utilized the gradient information as the network's input apart from the local patches. In view of that different distortions affect the image appearance differently, Jiang *et al*. [50] took the distortion classification task as an auxiliary task of quality prediction to improve network's representation ability for more accurate prediction. Although dividing the image into blocks can effectively increase the number of training samples, it inevitably leads to the loss of global information.

Later works applied the network with large input as the feature extractor to explore both local and global information [2]. To expand the number of training samples, the original image was randomly cropped many times. For example, Golestaneh *et al*. [33] randomly selected 50 patches with the size of 224×224 from an image and took the mean value of all 50 patches' prediction scores as the image's prediction score. To obtain accurate IQA performance, they considered the relative distance information between the images within each batch during network training. In addition, the resizing operation was also adopted for global information preservation. For instance, Wu *et al*. [31] unified the input of network as 300×300 and generated a great number of training data by annotating the synthesized distorted images with FR-IQA methods. Considering that the effectiveness of such an annotation strategy is restricted by the FR-IQA methods selected, Liu *et al*. [34] applied the natural order of the relative quality to train the network. The relative quality can be easily obtained by processing one image with different distortion levels. Given that ranking the samples from specific distortion types is

beneficial for obtaining effective representations for the IQA task, Sun *et al*. [51] learned a new network using a large number of synthesized distorted images, named GraphIQA, in which each synthesized distortion is represented as a graph. Before predicting the quality of an image with authentic distortions, the learned network should be finetuned on an authentically distorted image database. Nonetheless, these methods mainly work effectively when the query images are with the similar distortions as the training data. In practice, authentically distorted images contain quite complex distortions. It is inappropriate to mimic their characteristics by a simple distortion. In view of this, Ou *et al*. [35] split the authentic distortions into a group of synthesized ones and generated many synthesized data for training the IQA network.

In summary, the conflict between small-scale labeled data available and large-scale data required for training remains the focus of discussion in designing CNN-based NR-IQA methods. Although many data enhancement solutions have been proposed, they are unsuitable for tackling the data insufficiency issue in authentically distorted IQA. On the one hand, as the authentically distorted image does not have pristine reference, we cannot utilize existing FR-IQA methods to generate the proxy label. On the other hand, since the authentic distortions are usually complex, it is hard to completely model the authentic distortions by a group of synthesized ones. Therefore, how to design a more advanced framework that can effectively solve the authentically distorted IQA task with limited labeled data becomes our main concern. In this paper, we propose a new framework by using few label data and many unlabeled data for network training. Different from the recently reported fully-supervised GraphIQA, the proposed framework is based on semi-supervised learning. In addition, it does not require any prior knowledge of distortion types and only train the network in one stage without finetuning. To effectively use the unlabeled data, the proposed framework has two parallel branches and computes both sample-level and batch-level consistencies between two branches. During the calculation of batch-level consistency, we consider the semantic similarity between the images in each batch and compute the difference between similarity matrices of two branches. This is quite different from the strategy introduced by Golestaneh *et al*. [33], which used the triplet loss to help the network learn the relative ranking between the images within each batch.

III. The Proposed SSLIQA Method

In this section, we introduce the proposed semi-supervised quality assessment framework (i.e., SSLIQA) for authentically distorted images in detail. Since our SSLIQA exploits images with and without human annotations, we first show the problem setting and describe the framework of SSLIQA in an overview. Then, we detail the proposed consistency-preserving strategy, including the sample-level consistency and batch-level consistency between the teacher and the student. Finally, we present the loss function and introduce how to simultaneously use labeled and unlabeled data for training.

A. Problem Setting and Framework

Let $D^L = \{x_i, y_i\}_{i=1}^{N_L}$ denote the IQA database consisting of N_L labeled images, where x_i and y_i are the i-th image and its corresponding subjective quality score, respectively. Supervised CNN-based NR-IQA aims to train a network (i.e., find a mapping function $f(x_i)$ parameterized by Θ_f) to predict the quality score $\widetilde{y_i}$ for approximating y_i. To fully train the network with limited labeled data, existing methods usually transfer (x_i, y_i) into $\{x_{i,k}, y_{i,k}\}_{k=1}^{N_k}$ via various operations like cropping, flipping, and distortion synthesis, thereby increasing the labeled data N_k times. For diverse operations, the label $y_{i,k}$ is set as either the quality score (i.e., y_i) of the whole image or the proxy value from the existing FR-IQA methods. However, neither approach is suitable to the authentically distorted IQA task, as discussed in Section II-B.

In this study, inspired by the fact that vast amounts of unlabeled data with authentic distortions can be easily collected, we consider a more encouraging and practically feasible semi-supervised learning manner, i.e., training NR-IQA model using labeled data as well as unlabeled data. The unlabeled data are expected to provide extra knowledge of distortion representation and content understanding, targeting at better training the network and further improving the network performance in terms of both effectiveness and generalization. Let $D^U = \{x_j\}_{j=1}^{N_U}$ denote the unlabeled set consisting of N_U unlabeled images. Now the problem becomes how to find a mapping function $f(x)$ with the participation of both labeled set D^L and unlabeled set D^U.

Fig. 1 shows the overall framework of the proposed SSLIQA, where the NR-IQA task is completed in an end-to-end manner. Specifically, SSLIQA holds an asymmetric parallel dual-branch structure, consisting of a teacher branch and a student branch. In each branch, we set a feature encoder, which is built upon the pre-trained CNN model (e.g., AlexNet [52], ResNet [53]), followed by a quality regressor. The encoder aims to generate the quality-related features, based on which the regressor is able to predict the quality. In the training stage, the data $x = (x^L, x^U)$ are first fed into two branches simultaneously, where x^L and x^U are the labeled data and unlabeled data, respectively. For x^L, the supervised objective $\mathcal{L}_\kappa^S$ ($\kappa \in \{t, s\}$) is used to separately enforce each branch to learn quality-related features in the encoder. For both x^L and x^U, a sample-level consistency $\mathcal{L}_s^U$ is applied to align the predictions from two branches. Second, for an input mini-batch with B samples, we compute the semantic relation among different samples at each branch and apply a batch-level consistency $\mathcal{L}_b^U$ to align sample semantic relations from two branches for exploring extra quality-related information. Finally, by collaboratively exploiting x^L and x^U, we are able to learn a satisfactory NR-IQA network with the constraints of $\mathcal{L}_t^S$, $\mathcal{L}_s^S$, $\mathcal{L}_s^U$, and $\mathcal{L}_b^U$. Note that, in contrast to $\mathcal{L}_t^S$ and $\mathcal{L}_s^S$, $\mathcal{L}_s^U$ and $\mathcal{L}_b^U$ are unsupervised loss as no human annotations is required during calculation. In the inference stage, the quality score of a query image can be obtained by feeding it into the student branch.

One of the highlights of our SSLIQA is the asymmetric network structure. As shown in Fig. 1, the feature encoder

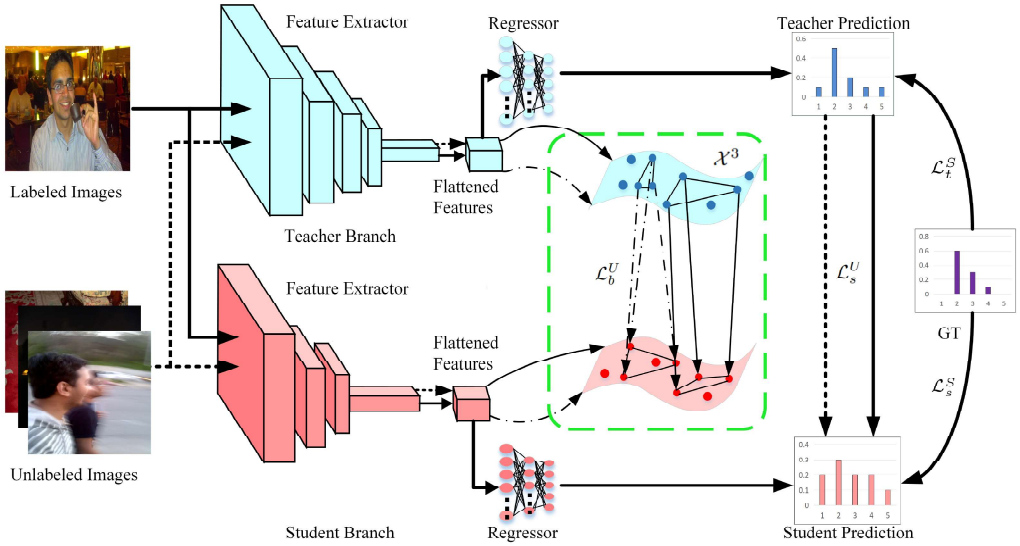


Fig. 1. The overview of our proposed SSLIQA framework, which holds an asymmetric parallel dual-branch structure. SSLIQA consists of a teacher branch and a student branch, and applies a consistency-preserving strategy to exploit images with and without labels simultaneously. Each branch has a supervised constraint $\mathcal{L}_\kappa^S$ ($\kappa \in \{t, s\}$) from the ground truth (GT) to help the network learn effective feature representations. The consistency-preserving strategy is composed of a sample-level consistency constraint $\mathcal{L}_s^U$ and a batch-level consistency constraint $\mathcal{L}_b^U$. Here, the solid and dotted lines indicate labeled and unlabeled data flows, respectively.

(i.e., backbone) is different between two branches. There are three aspects of motivation behind this. First, since authentic distortions are not easily synthesized to obtain quality ranking information, it is inappropriate to compare the relative quality between two branches with the same backbone, as suggested by [34]. Second, since the backbone of each branch is pre-trained on ImageNet, it would learn more semantic information for classification, instead of quality-related information, to make the network converge quickly if we set a symmetric parallel structure. Third, a smaller IQA network with higher performance is more favoured in practice as most portable smart devices with digital cameras are with limited computational ability. By setting an asymmetric parallel structure, we can boost the performance of the small student branch to approach to that of the large teacher branch, and only utilize the well-trained student branch to test a query image for reducing the computational requirements. More importantly, this strategy makes our network very flexible to be extended by changing the backbone of the teacher branch, as discussed in Section IV-D.

B. Consistency-Preserving Strategy

As discussed in Section III-A, the core idea of our proposed SSLIQA is to enforce the consistency between two branches. In this way, images with and without human annotations can be exploited for training a NR-IQA network. After carefully considering the subjective scoring behaviors, we propose a consistency-preserving strategy, which consists of a sample-level consistency and a batch-level consistency, to better tackle the IQA task in this study. In what follows, we will introduce each kind of consistency in detail.

1) Sample-Level Consistency: Recent progress on semi-supervised learning shows that encouraging the prediction consistency of teacher and student would boost the student's performance [54]. This intuitively inspires us to compute the distance (e.g., mean absolute error) between outputs of teacher and student in our SSLIQA. One derived question is what kind of output we should choose. In existing literature, NR-IQA methods, especially the one designed for synthetically distorted images, usually take a single scalar quality score as the ground truth to train the network based on the assumption that viewers usually reach a consensus on the visual quality of one image. By doing so, the trained network can only feed back the image quality with a scalar value. Although intuitive, a single scalar value usually fails to completely reveal the real quality of one authentically distorted image.

During the subjective experiment, an authentically distorted image will probably confuse reviewers due to its complex distortions, thereby receiving divergent opinion scores from different viewers. To illustrate this, we show two examples selected from the KonIQ-10K database [55] in Fig. 2, where the left side displays two images with the similar mean opinion scores (MOSs), and the right side presents their corresponding histogram of subjective rating scores from different viewers. As can be seen, compared with the rating scores of the second image, those of the first image are more divergent, almost spanning all the five scales. This indicates that the rating score distribution (RSD) contains more scoring information (e.g., scoring uncertainty) than a single scale quality score. With this observation, we propose to use the RSD as the output of each branch. For one image, we have two outputs from the proposed SSLIQA. The sample-level consistency $\mathcal{L}_s^U$ can be defined as:

$$\mathcal{L}_s^U = \frac{1}{B^L}\sum_{i=1}^{B^L} \|\widetilde{y}_t^i - \widetilde{y}_s^i\|_2^2 + \frac{1}{B^U}\sum_{j=1}^{B^U} \|\widetilde{y}_t^j - \widetilde{y}_s^j\|_2^2, \quad (1)$$

where B^L and B^U denote the number of labeled data and the number of unlabeled data in a mini-batch, respectively. $\widetilde{y}_t^\kappa \in \mathbb{R}^{1\times P}$ and $\widetilde{y}_s^\kappa \in \mathbb{R}^{1\times P}$ ($\kappa \in \{i, j\}$) respectively indicate the output of teacher and the output of student, where P refers to the maximum scoring scale. By minimizing $\mathcal{L}_s^U$ during the training process, the network would enforce the student to mimic output activations of individual examples represented by the teacher.

Fig. 2. Examples from the KonIQ-10K database. They have similar MOSs (1.804 vs. 1.848), but are with different rating score distributions.

2) Batch-Level Consistency: Recall from the recent study on subjective quality assessment of authentically distorted images [55]–[57], we often refer to previous analogous cases (either similar in contents or similar in distortions) to make scoring decisions for the current image. In addition, the scoring decisions are also affected by the image content, which helps the viewers understand the connection between image and distortion [58]. Based on these observations, we propose to model such scoring behaviors in the proposed NR-IQA network. Concretely, we take the output $F \in \mathbb{R}^{C\times H\times W}$ of the last convolutional layer of the backbone as the semantic information, where C is the channel number, and H and W are the spatial dimension of the feature map. For samples in a mini-batch, we first compute the semantic relation among a triplet of samples along each branch in an angle-wise way. Taking the student branch as an example, the semantic relation $\psi(s_m, s_n, s_h)$ can be calculated as:

$$\psi(s_m, s_n, s_h) = \cos\angle s_m s_n s_h = \langle d_{m,n}, d_{h,n}\rangle, \tag{2}$$

where

$$d_{m,n} = \frac{s_m - s_n}{\|s_m - s_n\|_2}, \tag{3}$$

$$d_{h,n} = \frac{s_h - s_n}{\|s_h - s_n\|_2}, \tag{4}$$

where $s_\kappa = \mathcal{F}_\kappa^s \in \mathbb{R}^{C_s\times 1\times 1}$ denotes the semantic feature of the κ-th sample x_κ ($\kappa \in \{m, n, h\}$) in a mini-batch. C_s is the channel number of the semantic feature obtained from the student branch. Here, we transfer $F_\kappa^s \in \mathbb{R}^{C_s\times H_s\times W_s}$ into $\mathcal{F}_\kappa^s$ by using the global averaging pooling operation to reduce the computational complexity. According to Eqs. (2)-(4), we can also obtain the semantic relation $\psi(t_m, t_n, t_h)$ of the teacher branch. Thanks to the angle-wise similarity used in Eq. (2), we can subtly tackle the dimension mismatch problem between semantic features from two branches.

In the asymmetric parallel network, encouraging the semantic relation consistency is conducive to learning more robust quality-related feature representation under different encoders. With this assumption, we apply a batch-level consistency $\mathcal{L}_b^U$, defined as:

$$\mathcal{L}_b^U = \frac{1}{|\mathcal{X}^3|}\sum_{(x_m,x_n,x_h)\in\mathcal{X}^3} \mathcal{S}_{L_1}(\psi(s_m, s_n, s_h) - \psi(t_m, t_n, t_h)), \tag{5}$$

where $\mathcal{X}^3 = \{(x_m, x_n, x_h)|m \neq n \neq h\}$ is a set of 3-tuples of distinct samples. $\mathcal{S}_{L_1}(\cdot)$ is the smooth L1 loss [59]. During network training, we simply use all possible tuples from samples in a given mini-batch. One advantages of utilizing batch-level consistency is that we do not need to provide any human annotations for supervision, as shown by Eq. (5). By exploiting more unlabeled data, the batch-level consistency can better train the student to form the same semantic relation with that of the teacher, promoting extracting additional semantic information for performance improvement.

C. Loss Function

The proposed SSLIQA aims to simultaneously exploit labeled and unlabeled data to train a model for effectively solving the authentically distorted IQA task. As shown in Fig. 1, SSLIQA adopts an asymmetric parallel dual-branch structure. For each branch, we utilize the labeled data to optimize it for the quality-related feature extraction. Specifically, given the mini-batch with B^L samples, the supervised loss L^S is the linear combination of two components:

$$\mathcal{L}^S = \underbrace{\frac{1}{B^L}\sum_{i=1}^{B^L}\|\widetilde{y}_t^i - y^i\|_2^2}_{\mathcal{L}_t^S} + \underbrace{\frac{1}{B^L}\sum_{i=1}^{B^L}\|\widetilde{y}_s^i - y^i\|_2^2}_{\mathcal{L}_s^S}, \tag{6}$$

where $y^i \in \mathbb{R}^{1\times P}$ is the ground truth (i.e., the rating score distribution) of the i-th images. The first item $\mathcal{L}_t^S$ and second item $\mathcal{L}_s^S$ are mean square error losses calculated from the teacher branch and the student branch, respectively.

To transfer quality-related knowledge, we propose a consistency-preserving strategy that explicitly encourages both sample-level and batch-level consistencies between two branches, as discussed in Section III-B. Since this strategy does not require any human labels for training, images with and without human annotations can all participate in network training. Concretely, the unsupervised loss L^U linearly combines the sample-level consistency $\mathcal{L}_s^U$ and the batch-level consistency $\mathcal{L}_b^U$:

$$\mathcal{L}^U = \mathcal{L}_s^U + \beta\cdot\mathcal{L}_b^U, \tag{7}$$

where β is a hyperparameter to balance $\mathcal{L}_s^U$ and $\mathcal{L}_b^U$. In this study, we empirically set it as 100. Finally, the total loss of the proposed SSLIQA is given by:

$$\mathcal{L} = \mathcal{L}^S + \lambda\cdot\mathcal{L}^U, \tag{8}$$

where λ is the trade-off weight between the supervised loss $\mathcal{L}^S$ and unsupervised loss $\mathcal{L}^U$. In this study, we apply a Sigmoid warming up function, i.e., $\lambda(t) = e^{-5(1-t/T)^2}$, to

control the value of trade-off weight λ. Here, t and T denote the current epoch and the maximum epoch, respectively. It is clear that, the λ value would gradually ramp-up from 0 to 1 during network training. Such a design could guarantee that the training loss would not be dominated by $\mathcal{L}^U$ at the beginning of network training when the consistency targets for unlabeled data are unreliable.

IV. Experiments and Results

A. Experimental Settings

1) Databases: In this study, we choose three authentically distorted IQA databases for evaluating and comparing the performance of our method and others, including KonIQ-10K [55], LIVE-C [57], and NNID [56]. KonIQ-10K includes 10,073 authentically distorted images with the spatial resolution of 1024×768. The images have balanced distributions of content, sharpness and brightness, covering complex authentic distortions involved in real-world photography. For each image, it provides more than 120 ratings, and the subjective quality score is reported in the form of MOS as well as RSD. The MOS values fall into the range of [1, 5]. LIVE-C consists of 1,162 authentically distorted images. Each image is collected by a mobile device without introducing any synthesized distortions beyond those occurring during capture, processing, and storage. The MOS value of each image ranges from 0 to 100. NNID is a natural night-time IQA database, containing a total of 2,240 images with 448 different image contents. Each image is captured by one of three photographic equipment at night, and its quality is reported in the form of MOS with the range of [0,1]. Since only KonIQ-10K provides the RSD, we mainly conduct experiments on it and use the remainder two database in the cross-validation experiments. Details of these databases are briefly summarized in Table I.

TABLE I
SUMMARY OF AUTHENTICALLY DISTORTED IQA DATABASES RELATING TO SCENE, IMAGE NUMBER, SUBJECTIVE SCORE TYPE AND SCORE RANGE.

Database	Scene	Number	Score Type	Score Range
KonIQ-10K	daytime & nighttime	10,073	MOS & RSD	[1,5]
LIVE-C	daytime & nighttime	1,162	MOS	[0,100]
NNID	nighttime	2,240	MOS	[0,1]

2) Evaluation Criteria: Four commonly adopted and widely acknowledged criteria by the IQA community are employed, including Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-order Correlation Coefficient (SRCC), Kendall Rank-order Correlation Coefficient (KRCC), and Root Mean Squared Error (RMSE). Among them, PLCC and RMSE are used for measuring the prediction accuracy, while SRCC and KRCC are used for evaluating the prediction monotonicity [56], [60]. Generally, a superior IQA method has higher values of PLCC, SRCC, and KRCC, and a smaller value of RMSE.

B. Implementation Details

In our SSLIQA, we adopt the AlexNet [52] and ResNet101 [53] pre-trained on ImageNet as the backbone (i.e., feature extractor) of the student branch and teacher branch, respectively. For the output of each backbone, we utilize the global average pooling to shrink its dimension and feed the resulted features into the quality regressor, which consists of a group of fully connected (FC) layers. Specifically, we respectively set four and two FC layers for the teacher and student as their backbones provide outputs with different dimensions. After the last FC layer, a Softmax function is applied to generate the predicted quality $\widetilde{y} \in \mathbb{R}^{1\times P}$.

We implement the proposed SSLIQA on the Pytorch library, and conduct the experiments on a workstation equipped with two Intel XEON 4210R CPUs and one NVIDIA RTX3090 GPU. The network is trained by using the Adam optimizer. The learning rate is initialized as 2e-4 and decayed with a 0.5 after every two epochs. We totally train 10 epoches for the NR-IQA task, and the ramp-up epoch T is set as 10. The batch size is set to 64, including 16 labeled images and 48 unlabeled images. Following the data enhancement strategies in previous works [33], we randomly flip the image along the horizontal direction. Meanwhile, we resize the image into 512×384 and randomly crop 10 sub-images with the resolution of 224×224. In the inference stage, 10 sub-images are randomly selected from a query image. For each sub-image, its quality score can be obtained by only feeding it into the student branch. By averaging 10 scores obtained, we can finally get the quality score of the query image.

C. Performance Comparison

In this subsection, we first compare our SSLIQA with two categories of recently reported NR-IQA methods. The first category contains four conventional NR-IQA methods, including NIQE [44], BRISQUE [12], GWH-GLBP [18], and SSEQ [11]. The second category contains seven CNN-based NR-IQA methods, including CNNIQA [20], WaDIQaM [47], PAQ-2-PIQ [61], NSSADNN [62], MetaIQA [22], MB-CNN [24], GraphIQA [51], and one Transformer-based NR-IQA method, namely MUSIQ-single [63]. For our SSLIQA, we randomly partition the KonIQ-10K database into training and testing sets by selecting 8,000 and 2,000 images, respectively. Two thirds of images in the training set discards their labels and serves as the unlabeled data during network training. In other words, we have 2,000 labeled images and 6,000 unlabeled images. The performance of our SSLIQA is reported on the testing set. For each competing method (except NIQE), we retrain the NR-IQA model on the 2,000 labeled images and test it on the testing set. Since NIQE is an opinion-unaware method, we directly test it on the testing set. Following the common practice in NR-IQA [24], we repeat the random split procedure 10 times, and take the median value of each evaluation criterion as the result.

Table II summarizes the experimental results on the KonIQ-10K database in terms of four evaluation criteria. From the table, we can see that: 1) The general performance of conventional NR-IQA methods is fairly unsatisfactory. Among the four methods, BRISQUE performs the best, but only achieves 0.581, 0.541, 0.373, and 0.447 in PLCC, SRCC, KRCC, and RMSE, respectively. NIQE obtains the worst results, in which

TABLE II
PERFORMANCE COMPARISONS ON THE KONIQ-10K DATABASE. FOR CONVENIENCE, THE BEST RESULT OF EACH EVALUATION CRITERION IS HIGHLIGHTED IN **BOLDFACE**.

	Methods	PLCC	SRCC	KRCC	RMSE
Conventional NR-IQA	NIQE [44]	0.300	0.276	0.186	0.524
	BRISQUE [12]	0.581	0.541	0.373	0.447
	GWH-GLBP [18]	0.557	0.502	0.344	0.459
	SSEQ [11]	0.326	0.303	0.206	0.518
CNN-based NR-IQA	CNNIQA [20]	0.654	0.635	0.446	0.449
	WaDIQaM [47]	0.665	0.644	0.459	0.411
	PAQ-2-PIQ [61]	0.728	0.718	0.524	0.369
	NSSADNN [62]	0.595	0.549	0.382	0.464
	MetaIQA [22]	0.860	0.826	0.635	0.279
	MB-CNN [24]	0.609	0.600	0.416	0.465
	GraphIQA [51]	0.862	**0.845**	**0.652**	0.280
	MUSIQ-single [63]	0.858	0.835	0.643	0.283
	SSLIQA (Ours)	**0.867**	0.841	**0.652**	**0.274**

both PLCC and SRCC are hard to reach 0.3. 2) Compared with conventional NR-IQA methods, CNN-based methods are more powerful in tackling the authentically distorted IQA task. For instance, NSSADNN, although performing the worst among the seven methods, still holds a comparable performance against BRISQUE. MetaIQA gets favoured results with a PLCC of 0.860, a SRCC of 0.826, a KRCC of 0.635, and a RMSE of 0.279. Nevertheless, there is still much room for performance improvement. 3) In contrast to these competing methods, the proposed SSLIQA holds considerable performance advantages and exhibits superior effectiveness. Specifically, it respectively achieves 0.867 in PLCC, 0.841 in SRCC, 0.652 in KRCC, and 0.274 in RMSE, and accordingly surpasses MetaIQA 0.7%, 1.5%, 1.7%, and 0.5% on these evaluation criteria. Meanwhile, it outperforms the runner-up (GraphIQA) 0.5% in PLCC and 0.6 % in RMSE, while is slightly inferior to it in SRCC. These indicate that our SSLIQA is more competent for the authentically IQA task.

Potential reasons about above results are given as follows. First, the key points of conventional methods lie in the handcrafted features selected. Generally speaking, one straightforward way of existing methods is to explore effective features via the analysis of structure, texture, naturalness, etc., as these image attributes are usually changed with the introduction of distortions. Empirically, statistical features (e.g., NSS, local binary patterns) can, to some extent, measure the distortion type or degree, because the synthetical distortions are regular in the spatial and transform domains [12], [18]. However, the authentic distortions are usually very complex. Traditional handcrafted features are hard to fully quantify such distortions even if we increase the number of features. Second, CNN-based methods advantage in quality-related feature extraction and fusion in an automatic way, instead of in an experience-based approach. Therefore, it is acceptable that CNN-based methods are more competent than conventional methods in the authentically distorted IQA task. In spite of this, most existing methods, e.g., CNNIQA, WaDIQaM, NSSADNN, and MB-CNN, are plagued by global information negligence, which is caused by dividing the image into small patches (with the size of 32×32) to expand the training data. As a result, they only obtain mediocre performance. In contrast, MetaIQA preserves the original size of the input image, achieving better results. Last but not the least, to cope with the problem of insufficient labeled data, our SSLIQA randomly crops the image multiple times with a relative large size (224×224) to preserve the global information. Moreover, SSLIQA simultaneously exploits both labeled data and unlabeled data and trains the IQA model in a semi-supervised manner. It is convenient to mine extra knowledge of distortion representation and content understanding from the unlabeled data, and to further improve the network performance.

To demonstrate the experimental results more intuitively, we show the scatter plots of MOSs versus predicted scores generated by the NR-IQA methods on the KonIQ-10K database in Fig. 3. From Fig. 3, it is clear that the blue points by the proposed method distribute around the red fitting curve more closely than the other NR-IQA methods. This indicates that the objective scores delivered by our method are more consistent with subjective scores.

Generally, a good NR-IQA method should not only achieve good performance on one database but also perform stably on unseen databases. To investigate the generalization ability of our SSLIQA, we further conduct a group of cross-validation experiments. More concretely, we revisit the learned NR-IQA models on the KonIQ-10K database with the default settings described in Section IV-C, and directly test them on the LIVE-C database and the NNID database without any fine-tuning, respectively. Similar to the main experiments, we also resize each testing image in NNID into 512×384 and randomly crop 10 sub-images with the resolution of 224×224. Whereas, we keep the original size of each image in LIVE-C. Obviously, the prediction value of each testing sub-image will be within the range of [1, 5] when using the IQA model learned on KonIQ-10K. To calculate RMSE, we should map the prediction value into the MOS range of the target database (i.e., LIVE-C or NNID) using a nonlinear logistic function. According to the recommendation of video quality experts group [64], we use the four-parametric logistic function in this study. Since testing image in NNID is cropped 10 times, its overall quality score is calculated as the mean of prediction values of 10 sub-images. Considering the low performance of conventional NR-IQA methods, we do not investigate their generalization ability here. Table III summarizes the comparison results. As can be seen, similar to the results in Table II, SSLIQA and MetaIQA hold the leading advantages than other methods across two datasets. More specifically, our SSLIQA obtains better results than MetaIQA in SRCC, KRCC, and RMSE on LIVE-C, while is slightly inferior to MetaIQA in all four evaluation criteria on NNID. One possible reason about this is that KonIQ-10K only has a small part of nighttime images, while NNID consists of nighttime images. There is large domain shift between the two databases, which hinders our method's understanding of nighttime images. To validate this argument, we replace 1,000 daytime unlabeled images by 1,000 nighttime unlabeled images while keep other images unchanged and retrain our SSLIQA. In this way, more nighttime images are included in the training stage. Experimental results show that our SSLIQA has a better performance (PLCC=0.783, SRC-

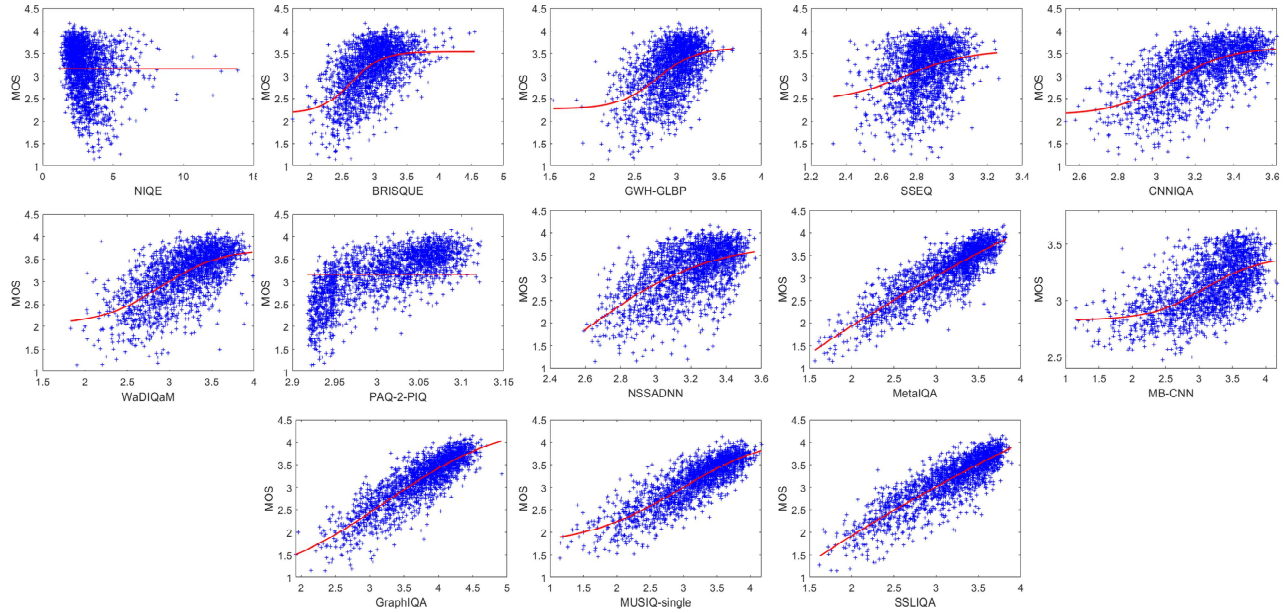


Fig. 3. Scatter plots of subjective scores (i.e., MOSs) versus objective scores predicted by NR-IQA methods on the KonIQ-10K database.

C=0.779, KRCC=0.580, RMSE=0.105) on the NNID database than before. This indicates that our SSLIQA can improve its effectiveness in evaluating nighttime images by using more nighttime images during network training. Compared to our SSLIQA, MetaIQA adopts a strategy of learning to learn, which more advantages in quickly adapting to new data with such large domain shift. Therefore, MetaIQA has an advantage over our SSLIQA in achieving better performance on NNID. Nevertheless, as shown by the results on LIVE-C, SSLIQA still performs better than MetaIQA when the testing set has the similar domain as that of the training set. In addition, the proposed SSLIQA is ahead of all other competing methods on two databases in all evaluation criteria by a large margin. These indicates the superiority of our SSLIQA against others in generalization.

TABLE III
RESULTS OF CROSS-VALIDATION EXPERIMENTS ON THE LIVEC AND NNID DATABASE. THE BEST RESULT OF EACH EVALUATION CRITERION IS HIGHLIGHTED IN **BOLDFACE**.

Methods	LIVE-C				NNID			
	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
CNNIQA [20]	0.513	0.485	0.333	17.601	0.597	0.584	0.420	0.136
WaDIQaM [47]	0.538	0.535	0.369	17.111	0.704	0.702	0.509	0.122
PAQ-2-PIQ [61]	0.528	0.506	0.345	16.954	0.715	0.712	0.514	0.118
NSSADNN [62]	0.439	0.426	0.291	18.437	0.733	0.731	0.536	0.114
MetaIQA [22]	**0.706**	0.676	0.482	14.377	**0.784**	**0.781**	**0.583**	**0.105**
MB-CNN [24]	0.481	0.459	0.313	17.915	0.553	0.548	0.386	0.140
GraphIQA [51]	0.619	0.591	0.414	15.948	0.728	0.727	0.529	0.116
MUSIQ-single [63]	0.538	0.512	0.355	17.104	0.521	0.488	0.338	0.145
SSLIQA (Ours)	**0.706**	**0.695**	**0.497**	**14.371**	0.771	0.770	0.571	0.108

D. Ablation Study

As described in Section III, our SSLIQA follows an asymmetric parallel dual-branch structure, and exploits both labeled and unlabeled images to cope with the authentically distorted NR-IQA task in a semi-supervised manner. Experimental results in Section IV-C demonstrates its good performance in effectiveness and generalization. In this subsection, we further conduct ablation experiments to investigate the robustness of our SSLIQA and to discuss the positive role of our network design concept. All experiments are conducted on the KonIQ-10K database with the same settings as the main experiment described in Section IV-C.

1) Robustness Analysis under Different Student-Teacher Combinations: Firstly, we fix the student branch as AlexNet [52], and select diverse classical networks, e.g., Inception-ResnetV2 [65], DenseNet121 [66], and ResNet101 [53] as the teacher branch, respectively. Both the regressors in the student and the teacher consists of several FC layers. Here, the regressor of the student is set to (256→5), where the number denotes the amount of neural nodes in the associated FC layer. Since the last block of these teachers provide outputs with different dimensions, their regressors are respectively set to (1536→512→256→5), (1024→512→256→5), and (2048→1024→512→256→5). Table IV lists the results under different student-teacher combinations. From the upper part of the table, there exists obvious performance gap (approximately 4% in PLCC and SRCC) between the student (i.e., AlexNet) and any of the remaining networks. This is mainly because AlexNet has a simpler structure than others, thereby showing limited power in the IQA task. In spite of this, when incorporating it with each of these networks (as the teacher) using our semi-supervised learning strategy, we can obviously increase its performance, as shown in the lower part of Table IV. Due to the relatively good performance, we chose the combination of AlexNet and ResNet101 as the result of our main experiment, as shown in Table II.

Secondly, we select one recently reported NR-IQA network

TABLE IV
ABLATION EXPERIMENTS ON DIFFERENT STUDENT-TEACHER COMBINATIONS. THE LOWER PART OF THIS TABLE SHOWS THE STUDENT'S PERFORMANCE IN CASE OF INCORPORATING IT WITH DIFFERENT TEACHERS. HERE, WE TAKE ALEXNET AS THE STUDENT.

Methods	Baseline			
	PLCC	SRCC	KRCC	RMSE
AlexNet [52]	0.835	0.808	0.612	0.302
InceptionResnetV2 [65]	0.875	0.846	0.657	0.266
DenseNet121 [66]	0.881	0.856	0.669	0.260
ResNet101 [53]	0.875	0.848	0.662	0.267
TReS [33]	0.879	0.861	0.676	0.262
Teachers	Semi-supervised NR-IQA			
	PLCC	SRCC	KRCC	RMSE
InceptionResnetV2 [65]	0.856	0.831	0.639	0.283
DenseNet121 [66]	0.864	0.839	0.650	0.276
ResNet101 [53]	0.867	0.841	0.652	0.274
TReS [33]	0.849	0.821	0.629	0.290

(TReS [33]) as the teacher to explore whether our method has the potential to combine with the existing NR-IQA networks. Here, we remove the relative ranking constraint of original TReS as it could burden the semi-supervised learning. As can be seen from the upper part of Table IV, compared with other baselines, TReS obtains superior performance as it considers both low-level spatial details and high-level semantic concepts for distortion understanding, instead of only the high-level semantic concepts. As shown in the last line, similar to these classical networks, it also boosts the performance of the student when playing the role of the teacher in the proposed semi-supervised training strategy. An interesting observation is that, although more obvious superiority is observed as a baseline, TReS does not help the student (i.e., AlexNet) achieve better performance compared to other networks after the semi-supervised training. One potential reason is that TReS uses the Transformer, which captures long-range dependencies and global information, to extract features from the input image. In contrast, AlexNet utilizes CNN, which mainly captures local information due to the limited receptive field of the convolution operation, to extract features from the input image. Since the focuses of Transformer and CNN are different during feature extraction, the CNN-based student may be hard to fully mimic the behavior of the Transformer-based teacher. Nevertheless, the teacher still contributes to improving the student's performance.

TABLE V
ABLATION EXPERIMENTS ON DIFFERENT STUDENT-TEACHER COMBINATIONS. THE LOWER PART OF THIS TABLE SHOWS THE STUDENT'S PERFORMANCE IN CASE OF INCORPORATING IT WITH THE TEACHER RESNET101.

Methods	Baseline			
	PLCC	SRCC	KRCC	RMSE
MobileNetV3 [67]	0.869	0.842	0.650	0.274
EfficientNetB0 [68]	0.883	0.856	0.669	0.259
Students	Semi-supervised NR-IQA			
	PLCC	SRCC	KRCC	RMSE
MobileNetV3 [67]	0.875	0.856	0.667	0.271
EfficientNetB0 [68]	0.888	0.868	0.684	0.255

Finally, we further keep the teacher (ResNet101) unchanged while replace the student (AlexNet) by two recently reported lightweight networks (MobileNetV3 [67] and EfficientNetB0 [68]) to investigate whether the performance of our method can be further improved. Table V shows the results. As seen, the teacher still boosts the performance of the student. Moreover, compared the combination of AlexNet+ResNet101, the performance of our method can be improved when replacing AlexNet by either of these two lightweight networks as the student. Overall, our proposed SSLIQA has strong robustness under different student-teacher combinations.

2) Robustness Analysis under Different Ratios of Labeled-to-Unlabeled Data: Generally, unlabeled data could provide extra information to boost the network learning. Here, we further investigate the performance change of our SSLIQA under different ratios of labeled-to-unlabeled data. The experimental results are shown in Table VI. For convenient understanding, we denote each SSLIQA method in the form of "student + teacher". Compared with the baseline results reported in the first row of Table IV, there exists an obvious performance gain no matter what proportion of unlabeled data is included. Moreover, it is clear that, overall, almost all four criteria increase gradually with the increase of the number of unlabeled images. These demonstrate that the unlabeled images play a positive role in achieving good IQA performance, and our SSLIQA is robust to different settings of labeled-to-unlabeled data.

3) Effectiveness of Sample-level Consistency and Batch-level Consistency: As described in Section III-B, SSLIQA proposes a consistency-preserving strategy, consisting of sample-level consistency and batch-level consistency, to exploit unlabeled data. Here, we make more in-depth investigations to explore the performance variations when removing each of them from SSLIQA. Table VII tabulates the experimental results. For convenient expression, we denote these two operations as "w/o Sample-level Consistency" and "w/o Batch-level Consistency", respectively. Through comparing the results in Table VI and Table VII, it obviously leads to a negative effect no matter which module is removed. Compared with sample-level consistency, batch-level consistency has greater influence on the final results. For instance, there would be approximately a 0.2% decrement of PLCC and a 0.2% decrement of SRCC when removing the sample-level consistency, and be approximately a 0.7% decrement of PLCC and a 0.8% decrement of SRCC when removing the batch-level consistency regarding the combination of AlexNet + InceptionResnetV2. This indicates that both two kinds of consistency play a positive role in the authentically distorted IQA task, and their natural cooperation helps the proposed SSLIQA work well.

4) Benefits from the Asymmetric Parallel Dual-branch Structure: In recent years, it has become a mainstream trend to utilize the classical classification network as the backbone of a NR-IQA model. For better performance, larger networks are especially favoured and welcome. However, this would increase the number of network parameters and put forward greater requirements for hardware devices. In our daily life, portable smart devices with digital cameras (e.g., phone, iPad) are popular tools for capturing digital images. A smaller model with higher performance is more desired in this situation. To

TABLE VI
ABLATION EXPERIMENTS ON DIFFERENT ON RATIOS OF LABELED-TO-UNLABELED DATA. THE BEST RESULT OF EACH EVALUATION CRITERION IS HIGHLIGHTED IN **BOLDFACE**.

Methods	Labeled:Unlabeled = 1:1				Labeled:Unlabeled = 1:2				Labeled:Unlabeled = 1:3			
	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
AlexNet + InceptionResnetV2	0.852	0.828	0.636	0.287	0.854	0.832	0.640	0.286	0.856	0.831	0.639	0.283
AlexNet + DenseNet121	0.859	0.834	0.643	0.280	**0.864**	**0.839**	**0.649**	**0.277**	0.864	0.839	0.650	0.276
AlexNet + TReS	0.844	0.813	0.620	0.295	0.847	0.817	0.627	0.291	0.849	0.821	0.629	0.292
AlexNet + ResNet101	**0.860**	**0.836**	**0.646**	**0.279**	0.863	0.837	0.648	0.278	**0.867**	**0.841**	**0.652**	**0.274**

TABLE VII
ABLATION EXPERIMENTS ON THE CONSISTENCY-PRESERVING STRATEGY.

Methods	w/o Sample-level Consistency			
	PLCC	SRCC	KRCC	RMSE
AlexNet + InceptionResnetV2	0.854	0.829	0.637	0.284
AlexNet + DenseNet121	0.863	0.838	0.648	0.277
AlexNet + TReS	0.844	0.813	0.620	0.291
AlexNet + ResNet101	0.865	0.843	0.654	0.274
Methods	w/o Batch-level Consistency			
	PLCC	SRCC	KRCC	RMSE
AlexNet + InceptionResnetV2	0.849	0.823	0.631	0.290
AlexNet + DenseNet121	0.853	0.827	0.635	0.286
AlexNet + TReS	0.856	0.831	0.639	0.285
AlexNet + ResNet101	0.856	0.830	0.638	0.284

TABLE VIII
COMPARISONS IN PARAMETER QUANTITY AND PERFORMANCE INCREMENT ($\triangle$PLCC AND $\triangle$SRCC). THE FIRST FOUR ROWS PRESENT THE RESULTS OF THE ENTIRE NETWORK UNDER DIFFERENT STUDENT + TEACHER COMBINATIONS IN THE TRAINING STAGE, RESPECTIVELY. THE LAST ROW SHOWS THE RESULTS OF THE STUDENT NETWORK (ALEXNET) IN THE TESTING STAGE.

Methods	Parameters	Parameter Ratio	$\triangle$PLCC	$\triangle$SRCC
AlexNet + InceptionResnetV2	57.7M	4.50%	2.50%	2.85%
AlexNet + DenseNet121	10.1M	32.90%	3.47%	3.84%
AlexNet + TReS	155.0M	1.60%	1.68%	1.61%
AlexNet + ResNet101	47.8M	5.50%	3.83%	4.09%
AlexNet	2.5M	100%	-	-

illustrate the advantage of the proposed SSLIQA in model size saving, we first respectively report its parameter numbers under different teacher settings during the training stage and the testing stage in Table VIII. As shown in the first column, different from the training stage, the network has very few parameters during the testing stage. For instance, its parameters are only 4.50% and 5.50% of AlexNet + InceptionResnetV2 and AlexNet + ResNet101, respectively. This is because our SSLIQA holds an asymmetric parallel dual-branch structure and enforces the student (AlexNet) to mimic the behavior of the teacher (i.e., InceptionResnetV2, DenseNet121, TReS, or ResNet101) during the training stage. Under such a configuration, our network has a relatively large parameter number. After network training, however, only the small-size student (i.e., AlexNet), is used in the testing stage, thereby reducing the model parameters greatly. Next, we show the performance increment brought by the proposed asymmetric dual-branch framework. Compared with its baseline (i.e., AlexNet in Table IV), the student's performance is greatly improved (see the last two columns of Table VIII) when incorporating it with a teacher by utilizing our SSLIQA. For example, the PLCC increment ($\triangle$PLCC) is more than 1.68% and the SRCC increment ($\triangle$SRCC) is more than 1.61% no matter which teacher is selected. In brief, our SSLIQA provides us a new solution on how to achieve higher IQA performance with a smaller network.

V. CONCLUSION

In this paper, we propose a semi-supervised NR-IQA framework, termed SSLIQA, for authentically distorted images. SSLIQA adopts an asymmetric parallel dual-branch structure, and its success lies in simultaneously exploiting both labeled and unlabeled images with the assistance of a consistency-preserving strategy. Concretely, such a strategy, inspired by the subjective scoring behaviors, enforces the student to mimic activations of the teacher, and helps to explore the intrinsic relation between images. Extensive experiments and ablation studies demonstrate that our SSLIQA is superior to twelve state-of-the-art NR-IQA methods with considerable effectiveness and generalization. Moreover, benefiting from the consistency-preserving strategy and the asymmetric network structure, our SSLIQA can effectively exploit the unlabeled data to achieve higher IQA performance with a smaller network. This points to an interesting avenue for future work.

REFERENCES

[1] Q. Jiang, Z. Peng, S. Yang, and F. Shao, "Authentically distorted image quality assessment by learning from empirical score distributions," *IEEE Signal Processing Letters*, vol. 26, no. 12, pp. 1867–1871, 2019.

[2] L. Li, T. Song, J. Wu, W. Dong, J. Qian, and G. Shi, "Blind image quality index for authentic distortions with local and global deep feature aggregation," *IEEE Transactions on Circuits and Systems for Video Technology*, accepted, DOI: 10.1109/TCSVT.2021.3112197, 2021.

[3] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 4, pp. 529–539, 2009.

[4] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[5] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2008–2024, 2013.

[6] W. Zhou, L. Yu, Y. Zhou, W. Qiu, M.-W. Wu, and T. Luo, "Local and global feature learning for blind quality evaluation of screen content and natural scene images," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2086–2095, 2018.

[7] G. Yue, C. Hou, K. Gu, T. Zhou, and H. Liu, "No-reference quality evaluator of transparently encrypted images," *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2184–2194, 2019.

[8] J. Xu, W. Zhou, and Z. Chen, "Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 5, pp. 1724–1737, 2021.

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

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

[11] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communication*, vol. 29, no. 8, pp. 856–863, 2014.

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

[13] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2457–2469, 2016.

[14] W. Zhou, L. Yu, W. Qiu, Y. Zhou, and M. Wu, "Local gradient patterns (lgp): An effective local-statistical-feature extraction scheme for no-reference image quality assessment," *Information Sciences*, vol. 397, pp. 1–14, 2017.

[15] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2035–2048, 2018.

[16] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 50–63, 2015.

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

[18] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 541–545, 2016.

[19] D. Ghadiyaram and A. C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *Journal of Vision*, vol. 17, no. 1, pp. 32–32, 2017.

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

[21] 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 Processing Magazine*, vol. 34, no. 6, pp. 130–141, 2017.

[22] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "Metaiqa: Deep meta-learning for no-reference image quality assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 143–14 152.

[23] W. Zhou, Q. Jiang, Y. Wang, Z. Chen, and W. Li, "Blind quality assessment for image superresolution using deep two-stream convolutional networks," *Information Sciences*, vol. 528, pp. 205–218, 2020.

[24] Z. Pan, F. Yuan, X. Wang, L. Xu, S. Xiao, and S. Kwong, "No-reference image quality assessment via multi-branch convolutional neural networks," *IEEE Transactions on Artificial Intelligence*, accepted, DOI: 10.1109/TAI.2022.3146804, 2022.

[25] S. Li, J. Xue, and Y. Han, "No-reference stereoscopic image quality assessment based on local to global feature regression," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2019, pp. 448–453.

[26] L. Shen, X. Chen, Z. Pan, K. Fan, F. Li, and J. Lei, "No-reference stereoscopic image quality assessment based on global and local content characteristics," *Neurocomputing*, vol. 424, pp. 132–142, 2021.

[27] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011006, 2010.

[28] S.-H. Bae and M. Kim, "A novel image quality assessment with globally and locally consilient visual quality perception," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2392–2406, 2016.

[29] L. Li, Y. Zhou, K. Gu, W. Lin, and S. Wang, "Quality assessment of dibr-synthesized images by measuring local geometric distortions and global sharpness," *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 914–926, 2017.

[30] G. Yue, C. Hou, K. Gu, T. Zhou, and G. Zhai, "Combining local and global measures for dibr-synthesized image quality evaluation," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 2075–2088, 2019.

[31] 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 Transactions on Image Processing*, vol. 29, pp. 7414–7426, 2020.

[32] G. Yue, C. Hou, W. Yan, L. K. Choi, T. Zhou, and Y. Hou, "Blind quality assessment for screen content images via convolutional neural network," *Digital Signal Processing*, vol. 91, pp. 21–30, 2019.

[33] S. A. Golestaneh, S. Dadsetan, and K. M. Kitani, "No-reference image quality assessment via transformers, relative ranking, and self-consistency," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1220–1230.

[34] X. Liu, J. Van De Weijer, and A. D. Bagdanov, "Rankiqa: Learning from rankings for no-reference image quality assessment," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1040–1049.

[35] F.-Z. Ou, Y.-G. Wang, J. Li, G. Zhu, and S. Kwong, "A novel rank learning based no-reference image quality assessment method," *IEEE Transactions on Multimedia*, accepted, DOI: 10.1109/TMM.2021.3114551, 2021.

[36] L. Li, H. Zhu, G. Yang, and J. Qian, "Referenceless measure of blocking artifacts by tchebichef kernel analysis," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 122–125, 2013.

[37] S. Wang, C. Deng, B. Zhao, G.-B. Huang, and B. Wang, "Gradient-based no-reference image blur assessment using extreme learning machine," *Neurocomputing*, vol. 174, pp. 310–321, 2016.

[38] G. Yue, C. Hou, K. Gu, and N. Ling, "No reference image blurriness assessment with local binary patterns," *Journal of Visual Communication and Image Representation*, vol. 49, pp. 382–391, 2017.

[39] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 838–842, 2015.

[40] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Transactions on Cybernetics*, vol. 47, no. 12, pp. 4559–4565, 2017.

[41] H. Z. Nafchi and M. Cheriet, "Efficient no-reference quality assessment and classification model for contrast distorted images," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 518–523, 2018.

[42] M. A. Saad, A. C. Bovik, and C. Charrier, "A dct statistics-based blind image quality index," *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 583–586, 2010.

[43] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[44] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[45] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3129–3138, 2012.

[46] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 206–220, 2017.

[47] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, 2017.

[48] X. Jiang, L. Shen, Q. Ding, L. Zheng, and P. An, "Screen content image quality assessment based on convolutional neural networks," *Journal of Visual Communication and Image Representation*, vol. 67, p. 102745, 2020.

[49] X. Jiang, L. Shen, G. Feng, L. Yu, and P. An, "An optimized cnn-based quality assessment model for screen content image," *Signal Processing: Image Communication*, vol. 94, p. 116181, 2021.

[50] X. Jiang, L. Shen, L. Yu, M. Jiang, and G. Feng, "No-reference screen content image quality assessment based on multi-region features," *Neurocomputing*, vol. 386, pp. 30–41, 2020.

[51] S. Sun, T. Yu, J. Xu, W. Zhou, and Z. Chen, "Graphiqa: Learning distortion graph representations for blind image quality assessment," *IEEE Transactions on Multimedia*, accepted, DOI: 10.1109/TMM.2022.3152942, 2022.

[52] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[54] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," in *International Conference on Learning Representations*, 2017.

[55] H. Lin, V. Hosu, and D. Saupe, "Koniq-10k: Towards an ecologically valid and large-scale iqa database," *arXiv preprint arXiv:1803.08489*, 2018.
[56] T. Xiang, Y. Yang, and S. Guo, "Blind night-time image quality assessment: Subjective and objective approaches," *IEEE Transactions on Multimedia*, vol. 22, no. 5, pp. 1259–1272, 2020.
[57] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372–387, 2016.
[58] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, and Y. Zhang, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3667–3676.
[59] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.
[60] Y. Zhou, Y. Sun, L. Li, K. Gu, and Y. Fang, "Omnidirectional image quality assessment by distortion discrimination assisted multi-stream network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 1767–1777, 2022.
[61] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, and A. Bovik, "From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3575–3585.
[62] B. Yan, B. Bare, and W. Tan, "Naturalness-aware deep no-reference image quality assessment," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2603–2615, 2019.
[63] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "Musiq: Multi-scale image quality transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5148–5157.
[64] V. Q. E. Group *et al.*, "Final report from the video quality experts group on the validation of objective models of video quality assessment," in *VQEG meeting, Ottawa, Canada, March, 2000*, 2000.
[65] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on Artificial Intelligence*, 2017, pp. 4278–4284.
[66] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
[67] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1314–1324.
[68] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*. PMLR, 2019, pp. 6105–6114.

Guanghui Yue received the B.S. degree in communication engineering from Tianjin University in 2014, and the Ph.D. degree in information and communication engineering from Tianjin University, Tianjin, China, in 2019. He was a joint Ph.D. student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, from September 2017 to January 2019.

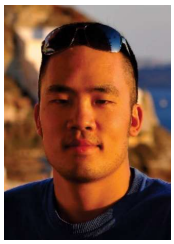
He is currently an Assistant Professor with the School of Biomedical Engineering, Health Science Center, Shenzhen University. His research interests include medical image analysis, bioelectrical signal processing, image quality assessment, 3D image visual discomfort prediction, pattern recognition, and machine learning.

Di Cheng received the B.S. degree in Electronic Information Engineering from Wuhan University of Science and Technology, Hubei Province, China, in 2020. Currently, he is pursuing the master's degree in biomedical engineering at Shenzhen University, China. His research interests include image quality assessment, medical image analysis, image restoration, and deep learning.

Leida Li (M'14) received the B.S. and Ph.D. degrees from Xidian University, Xian, 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, Kaohsiung, Taiwan. From 2014 to 2015, he was a Visiting Research Fellow with the Rapid-Rich Object Search Laboratory, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he was a Senior Research Fellow from 2016 to 2017. He is currently a Professor with the School of Information and Control Engineering, China University of Mining and Technology, China, and also with the School of Artificial Intelligence, Xidian University, China. His research interests include multimedia quality assessment, affective computing, information hiding, and image forensics. He has served as an SPC for IJCAI 2019-2020, the Session Chair for ICMR in 2019 and PCM in 2015, and the TPC for AAAI in 2019, ACM MM 2019-2020, ACM MM-Asia in 2019, ACII in 2019, and PCM in 2016. He is currently an Associate Editor of the *Journal of Visual Communication and Image Representation* and the *EURASIP Journal on Image and Video Processing*.

Tianwei Zhou received the B.S. degree in automation from Tianjin University in 2014 and the Ph.D. degree in control science and engineering from Tianjin University, Tianjin, China, in 2019. She was a joint Ph.D. student with the Department of Electrical & Computer Engineering, National University of Singapore from August 2017 to August 2018.

She is currently an Assistant Professor with the College of Management, Shenzhen University. Her current research interests include event-triggered control, intelligent scheduling, image processing, and medical image analysis.



Hantao Liu received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011. He is currently an Associate Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. He is an Associate Editor of the IEEE Transactions on Human Machine Systems and the IEEE Transactions on Multimedia.

Tianfu Wang received the Ph.D. degree in biomedical engineering from Sichuan University, Chengdu, China, in 1997. He is currently a Professor with Shenzhen University, Shenzhen, China. His current research interests include ultrasound image analysis, medical image processing, pattern recognition, and medical imaging.