IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 23, 2019, accepted November 6, 2019, date of publication December 2, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2957235

Blind Image Quality Assessment of Natural Distorted Image Based on Generative Adversarial Networks

HONGTAO YANG[1,2], PING SHI[1], DIXIU ZHONG[1], DA PAN[1], AND ZEFENG YING[1]
[1]School of Information and Communication Engineering, Communication University of China, Beijing 100024, China
[2]School of Electrical and Information Engineering, Beijing Polytechnic College, Beijing 100042, China

Corresponding author: Ping Shi (shiping@cuc.edu.cn)

This work was supported by in part by the R&D Program of Beijing Municipal Education Commission under Grant KM201910853001, in part by the Fundamental Research Funds for the Central Universities, in part by the High-Quality and Cutting-edge Disciplines Construction Project for Universities in Beijing, and in part by the Beijing Top Young Talents Project under Grant CIT&TCD201704032.

ABSTRACT Most existing image quality assessment (IQA) methods focus on improving the performance of synthetic distorted images. Although these methods perform well on the synthetic distorted IQA database, once they are applied to the natural distorted database, the performance will severely decrease. In this work, we propose a blind image quality assessment based on generative adversarial network (BIQA-GAN) with its advantages of self-generating samples and self-feedback training to improve network performance. Three different BIQA-GAN models are designed according to the target domain of the generator. Comprehensive experiments on popular benchmarks show that our proposed method significantly outperforms the previous state-of-the-art methods for authentically distorted images, which also has good performances on synthetic distorted benchmarks.

INDEX TERMS Generative adversarial networks, deep learning, image quality assessment, no-reference/blind image quality assessment, natural distorted image.

I. INTRODUCTION

With the development of advanced technologies, visual media has shown explosive growth. Thus, image quality assessment (IQA) becomes one of the hot topics in the field of image processing and computer vision, which refers to analyzing the characteristics of images and predicting the perceptual quality of a distorted image.The main purpose of IQA is to use appropriate evaluation criteria to make the evaluation results most consistent with human subjective evaluation. Overall, IQA include subjective IQA and objective IQA, depending on whether or not people are involved. Objective IQA could be classified into full-reference IQA (FR-IQA) [1]–[3], reduced-reference IQA (RR-IQA) [4], and no-reference IQA (NR-IQA) [5]–[17]depending on applying the reference image. It is difficult and infeasible in practical applications to obtain an ideal reference image. Therefore, blind image quality assessment (BIQA) is more realistic for the real world environment and is more challenging to make good performances.

Various IQA methods recently are data-driven, which depend on the databases. Most of the methods utilized synthetic distorted image databases [18]–[21], which are obtained through the ideal distortions and digital simulations from the reference image. The degrees and parameters of the synthetic distorted image are set artificially, and each distorted image only has a single type and a single level of distortion, which is an idealized image that does not exist in real life. For example, Live dataset [18] is designed by simulating JP2K compression, JPEG compression, additive White Gaussian Noise, Gaussian blur and simulated Fast Fading to the reference image independently. However, various authentic distortions exist in the natural images. Natural distorted images are different from synthetic distorted images. In terms of the content of the image, as the synthetic distorted images are always obtained by importing distortions to the reference images, the content in these datasets is monotonous. For the natural distorted images, the content is rich and colorful. In terms of distortions of the image, synthetic distorted

The associate editor coordinating the review of this manuscript and approving it for publication was Chaker Larabi.

images contain only a few specific types of distortions and almost all of them are single type distortions. But natural distorted images cover almost all the types and levels of distortions, which are interwoven with each other and could not be separated. Examples include low-light noises, over-exposed impairments, compression errors, and new types of distortions resulted from a mixture of different types. In addition, some other types of distortions such as geometric distortion are not often considered. This is a challenging work, so it is necessary to build an IQA model to better evaluate the authentically distorted image quality.

Recently, some efforts have been made to exploit the IQA on natural distorted images [5], [22], [23]. Experiments on LIVE Challenge (LIVE In the Wild Image Quality Challenge Database) database show that the correlation coefficient between the predicted score and the subjective score is only about 0.7, which is far from excellent performance. Most of the existing IQA models [6], [16], [22], [24]–[27] are designed based on the synthetic distorted images, which have poor generalization performance when transfer to natural distorted images.

Unlike synthetic distortion images, natural distortion images are directly captured from the actual scene, and there are no reference images, therefore, the quality prediction of natural distorted images can only be achieved in a no-reference IQA manner. Generative Adversarial Networks (GAN) [28] consists of a pair of competing network structures called generator (G) and discriminator (D) respectively, which can learn deep features without sufficient labeled training data. The ultimate goal of IQA is to get the mapping from distorted images to quality-related information, and GAN is to get the mapping from the feature vector of latent space to the target domain. Both IQA and GAN have similarities and correlations. Inspired by this, BIQA-GAN is designed and adversarial learning mechanism between G and D is utilized to get mapping from distorted image space to quality-related space. For traditional methods, it is difficult to extract the features of natural images; and for the previous CNN models, it is difficult to generate a high-quality quality-related information which is similar to label. However, the models based on GAN could take advantage of its strong fitting capability through adversarial learning between G and D to learn the features of images, which could better fit natural distorted images.

In this paper, we first design No-reference Quality map generation model based on GAN (NQM-GAN). Similar to GAN [28], the G in NQM-GAN is to predict the image quality score, and its target domain is the quality map related to the quality score. For the loss function, MAE is added to the loss function of the G, which could penalize the pixel error between the output of the G and its corresponding Ground Truth, besides, it could prevent gradient vanishing due to the influence of D. Through experiments, it was found that the use of the target domain of MOS map with the same pixel value could not make the discriminator converge to the expected distribution, thus affect the prediction performance of G. To solve this problem, we subsequently proposed two improved structures, which are No-reference saliency Weighted Quality score Map generation model based on GAN (NWQM-GAN) and No-reference Quality score generation model based on GAN (NQS-GAN). Finally, to demonstrate the strong competitiveness of our approach, comprehensive experiments are conducted on existing IQA benchmarks including LIVE [18], TID2013 [20] and LIVE Challenge [29], [30]. The results show that the proposed NQS-GAN significantly outperforms the previous state-of-the-art methods on three publicly available benchmarks. It could be proved that through adversarial learning, the mapping from distorted image space to the quality score space can be well realized, thus significantly improves the prediction precision and robustness. Besides, the models have very good generalization performances, which not only has good performance on authentically distorted images, but also on synthetic distorted datasets.

Our main contributions of this work are summarized into three folds:

(1) Natural distorted images are authentically distorted images,this is the first time that GAN related method applied to the natural distorted image quality assessment.

(2) As deep learning is data-driven, shortage of images in IQA datasets is a serious problem. In our proposed BIQA-GAN method, through adversarial learning, D could assist G to fit the distribution of input data, which does not need to rely solely on the generation of high-quality quality-related labels. It can better predict the score even if the dataset is small.

(3) According to the different target domain of the generator, pixel-level quality map, weighted saliency quality map and quality score are separately generated by NQM-GAN, NWQM-GAN and NQS-GAN, which are novel structures of BIQA-GAN.

II. RELATED WORK

A. IMAGE QUALITY ASSESSMENT

Generally speaking, the research in objective IQA could be carried out in three perspectives, which are research based on Human Vision System (HVS), research based on Natural Scene Statistics (NSS) and research based on the self-learning IQA model by data-driven.

In the early stage of IQA, it is basically based on the first two perspectives.Wang *et al.* [31] propose SSIM algorithm, considering that HVS is relatively suitable for extracting scene structure information. SSIM can be used to perceive the image quality through the measurement of structural information variation, which realizes the similarity measurement by comparing with the reference image in the pixel brightness, contrast and structure. In the following studies, many IQA models are designed based on SSIM, such as MS-SSIM proposed by combining multi-scale to improve the model performance. Zhang *et al.* [32] propose FSIM considering that HVS mainly focus on low-level features, such as image edges, colors and contours. FSIM takes phase

consistency as the basic feature, as it is proved that HVS could obtain effective information related to phase consistency. Besides, phase consistency is not affected by image contrast, but image local contrast is related to the perceived quality, so FSIM applies image gradient amplitude as the second feature.There are two assumptions for the application of NSS in IQA. One is that high-quality natural images always obey some statistical laws related to perception, the other is the image distortion would disturb these statistical laws. Therefore, image quality can be evaluated through quantitatively analyzing models. Mittal *et al.* [6] proposed BRISQUE through measurement of spatial domain statistical properties, which extracts a total of 36 spatial domain normalized features through two different scales of the image to realize the distortion type classification and score regression. DIIVINE [24] extracts the statistical correlation features of sub-band distribution, orientation, scale space in the wavelet transform domain.

Both IQA by simulating HVS and NSS rely on hand-design features, which is not easy to achieve satisfactory results. The performance of CNN-based IQA model is often better than that of the models based on analyzing HVS and NSS. With the development of deep learning, learning from distorted images automatically is becoming more and more popular. CORNIA [33] obtain feature codes from image block sets and finally get quality score by max-pooling strategy. The studies in recent years show that deep convolution neural network (CNN) has been applied more and more in synthetic distorted IQA based on its powerful capability of data processing and feature extraction. Kang *et al.* [7] apply a 6-layers CNN to implement a simple score regression model, which is also modified to a multi-task CNN [8], where the network learns both distortion type and quality score simultaneously; Bianco *et al.* [9] replace the last classification layer of AlexNet with the linear regression layer, fine tune the network parameters to obtain the quality score prediction model. Kim and Lee [10] propose an idea that training a CNN to replicate a conventional FR-IQA and design a blind image evaluator based on a convolution neural network (BIECON), which applies a CNN to estimate a patch score map and use one hidden layer to regress the extracted patch-wise features into a subjective score. Kim and Lee [2] also construct a deep IQA (DeepQA) model based on visual perception related features, which first calculate error map through comparing the reference with the distorted image, and then the error map combined with distorted image are forwarded to the network to generate sensitivity map which could reflect the features of the local details of the distorted image, After multiplication with the error map, it is regressed to the subjective score. Pan *et al.* [16] design a blind predicting similar quality map (BPSQ) for IQA, which applies a FR metric to calculate similar graphs of image patches, and then uses an U-shaped CNN to generate a similar graph, which is used to train a CNN regression model to predict the score of the image patch. Lin and Wang [17] propose a Hallucinated-IQA framework for NR-IQA via adversarial learning, which jointly optimizes quality-aware generative network, hallucination-guided quality regression network, and IQA-discriminator in an end-to-end manner. They generate a hallucinated reference, which is forwarded to the regression network with distorted image to learn perceptual discrepancy. The performance is excellent on the synthetic distorted images, but not mentioned on natural distorted images.

CNN has been applied more and more in IQA, but most of these models are utilized in the synthetic distorted IQA database, once applied to the natural distorted IQA database, the performance reduced dramatically. So the research in the field of natural distorted image still has many problems to deal with, which is also the issue we focus on.

B. NATURAL DISTORTED IMAGE QUALITY ASSESSMENT

Similarly, most of the studies on natural distorted IQA are also based on the modeling and analysis of HVS and NSS. To study complex multi-distortions of natural distorted images, Ghadiyaram and Bovik [22] propose FRIQUEE algorithm, which combines with 330 distortions related to NSS features from different transformation domains in the LMS, CIELAB and RGB color space to capture the damage caused by various types of distortions. Yang *et al.* [5] apply BRISQUE [6] to extract the four orientation parameters of local normalization coefficient distribution, along with blurriness, contrast, dynamic range, and color information, to predict the quality of natural distorted image. Ying *et al.* [23] propose a joint dictionary NR-IQA method, which first extract the aesthetic features and NSS features from natural distorted images, then acquire the feature dictionary and quality dictionary through joint dictionary learning, and finally calculate the image quality score according to the feature dictionary and quality dictionary.

The IQA based on HVS and NSS is to design corresponding features to predict the quality score, which relates to distortions with human perception and the disturbance with NSS. For the natural distorted image, distortions interact with each other, which are hard to be dealt with separately. Therefore, how to obtain representative complex multi-distorted features is the core content for natural distorted images.

C. GENERATIVE ADVERSARIAL NETWORK

Generative Adversarial Networks (GAN) [28] has played an important role in the machine learning field for the capability to learn high-dimensional, complex real data distribution. Besides, GAN and their variants do not depend on any assumptions about the distribution, but could get real-like samples from latent space. This could lead GAN to various applications, such as Pix2pix [34], PAN [35], CycleGAN [36], DiscoGAN [37] in image translation; SRGAN [38] in super resolution; SeGAN [39], Perceptual GAN [40] for small object detection in object detection; Coupled GAN [41] in joint image generation; VGAN [42],

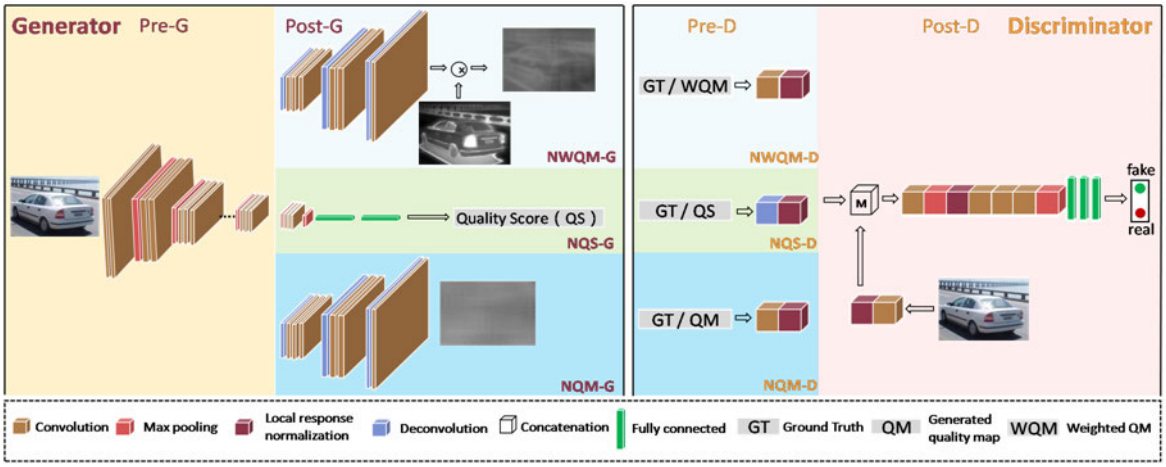


FIGURE 1. The general architecture of BIQA-GAN models. G is for generating quality-related information. ResNet101 is truncated and utilized as the pre-structure of G. Different post-structures of G are designed according to the different quality-related information. D is used for a binary classifier. Different pre-structures of D are designed to fit for different output of G. Same post-structures of D are similar with AlexNet.The output of G and the Ground Truth are alternately forwarded to D for binary classification.

Pose-GAN [43], MoCoGAN [44] in video generation; Stack GAN [45], TAC-GAN [46] in text to image; SD-GAN [47], SL-GAN [48], DR-GAN [49], AGEGAN [50] in changing facial attributes; GAN has shown its excellent fitting capability to target domain distribution.

III. OUR APPROACH

A. GENERAL ARCHITECTURE OF BIQA-GAN

In the paper, three BIQA-GAN models are proposed, which are NQM-GAN, NWQM-GAN and NQS-GAN. The general architecture is illustrated in FIGURE 1.

The G of BIQA-GAN is used for generating quality-related information for IQA. ResNet101 [51] is truncated to the previous pool1 of the softmax layer, and utilize it as the pre-structure of G, which is called ResNet101-pool1. Different post-structures of G are designed in these independent models according to the different quality-related information to be generated, which are quality map, the weighted saliency quality map and the quality score generated by NQM-GAN, NWQM-GAN and NQS-GAN separately.

The D of BIQA-GAN is used for a binary classifier. Different pre-structures of D are designed to fit for different output of G. The post-structures of D in BIQA-GAN are the same, parallel branch is processed separately and merge into the subsequent single processing branch, whose structure is consistent with AlexNet after the first group of convolution, the last 1000 softmax output is replaced with binary logistic outputs. The output of G and the Ground Truth are alternately forwarded to D for binary classification discrimination. At the same time, the classification error is fed back to G. In the process, an adversarial training between G and D is constituted to guide G to generate better quality-related information. Considering the correlations between the quality score and the distorted image, auxiliary classification discrimination branch is applied with the corresponding distorted image as input.

B. PROPOSED NQM-GAN MODEL

Quality map [16] corresponds to the objective quality score of each pixel in the distorted image, which is the reflection of the objective image quality in pixel-level. In order to get excellent performance, a novel NQM-GAN is first designed to use adversarial strategy to generate good quality map. The general structure of NQM-GAN is shown in FIGURE 2.

G in GAN [28] is to map the feature representation to the natural image spatial domain, while G in NQM-GAN is to map the distorted image to quality score spatial domain. Therefore, the input of G in NQM-GAN is no longer a random vector initialized into Gaussian white noise, but the distorted image itself. The output of G is the quality map corresponding to the objective quality score of each pixel in the distorted image.

The structure of G in NQM-GAN is shown in FIGURE 3. The 3*224*224 distorted patch is forwarded to ResNet101-pool1 to get 2048*14*14 feature map. After a series of deconvolution and convolution layers processing, 1*224*224 quality map is obtained. G does not change the image dimensions, which only realizes the dimension reduction through the convolution layers and the dimension expansion through deconvolution layers. Each value in the quality map corresponds to the objective quality score of each pixel in the distorted image. In the prediction stage, the final quality score is obtained by applying average pooling strategy to the quality map, which is depicted in Eqn.(1), where $G_\theta(I_d)$ is the quality map generated by G; H and W are the height and width of the quality map.

$$S_{NQM} = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} G_\theta(I_d)(i,j) \tag{1}$$

D is depicted in the bottom of the FIGURE 2, which is modified based on AlexNet. Considering the correlation between the quality map and the corresponding distorted

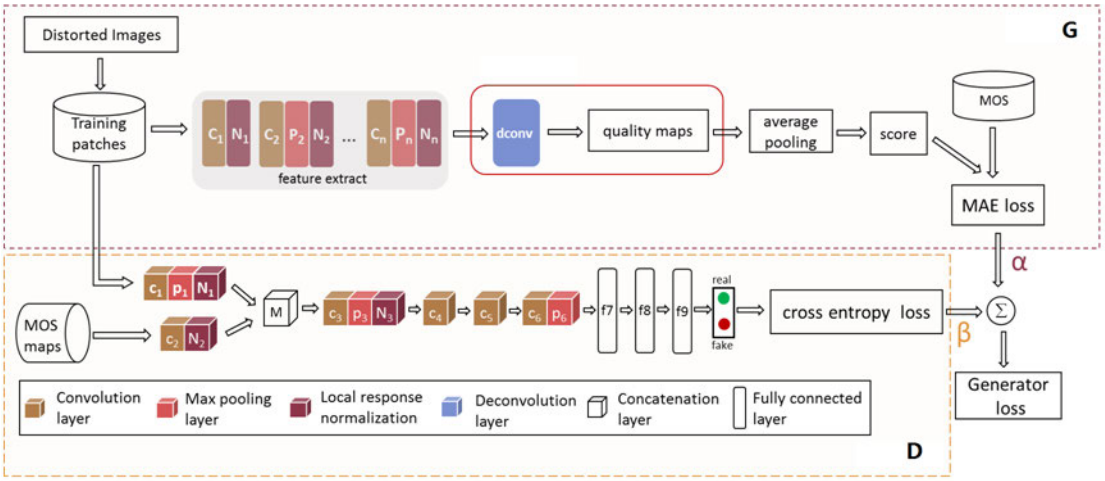


FIGURE 2. The general structure of NQM-GAN. For G, ResNet101-pool1 is utilized to extract features and deconvolution is used to generate quality map. For D, parallel branch is processed separately and merge into the subsequent single processing branch, whose structure is consistent with AlexNet after the first group of convolution, the last 1000 softmax output is replaced with binary logistic outputs.

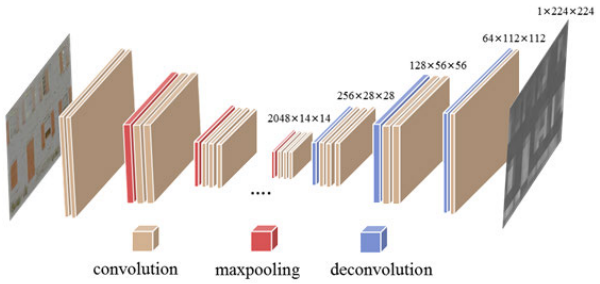


FIGURE 3. The structure of the generator in NQM-GAN.

image, both of them are forwarded to D to realize the binary classification. Therefore, the pre-structure of D has two branches, one is for the distorted image, the other is for the generated quality map or Ground Truth alternately. Each branch is processed separately and merge into one single branch, which is consistent with AlexNet after the first group of convolution. It is worth noting there is no subjective pixel-level quality map as Ground Truth, which only has MOS annotation. In order to obtain the Ground Truth corresponding to the quality map, the MOS value is tiled together to get the equal-pixel-value grayscale MOS map consistent with the dimension of the distorted image.

According to the experiments of NQM-GAN on LIVE Challenge in the following section, applying equal-pixel-value MOS map as the target domain could not make D converge to the expected distribution, thus affect the prediction of G. To solve this problem, the structure is modified and NWQM-GAN and NQS-GAN are proposed.

C. PROPOSED NWQM-GAN MODEL

Not all the areas of the image have the same amount of quality related information. Flat areas such as the sky may be sensitive to some distortions and insensitive to others. Distortions in some prominent areas may be more detrimental to perceived quality. In order to better conform to the quality perception of humans, different areas with different contents should be treated separately and assigned different weights, so as to obtain the predicted quality score more consistent with the subjective perception. Due to the saliency area on human perception, the weighted strategy is utilized to the quality map to get the weighted saliency quality map, thus NWQM-GAN model is designed. The Ground Truth is also modified to the weighted MOS map which is obtained by the multiplication of the MOS map and the saliency map together. Saliency map is obtained by using the saliency detection method proposed by Margolin [52]. This avoids the problem of using equal-pixel-value MOS map as the target domain.

The structure of NWQM-GAN is depicted in the top of FIGURE 1. NWQM-GAN has similar structures with NQM-GAN depicted in FIGURE 2. However, both the post-structure of G and the pre-structure of D apply the weighted saliency strategy to quality map and MOS map separately. The weighted saliency quality map and weighted MOS map are forwarded to D alternately for binary classification, simultaneously the error information is fed back to G. Besides, the sum pooling is utilized to the weighed quality map to calculate the final quality score instead of the average pooling in NQM-GAN. Eqn.(2) depicts the process, where $G_\theta(I_d)$ represents the weighted saliency quality map generated by G, H and W are the height and width of the weighted saliency quality map.

$$S_{NWQM} = \sum_{i=1}^{H} \sum_{j=1}^{W} G_\theta(I_d)(i,j) \tag{2}$$

D. PROPOSED NQS-GAN MODEL

The datasets such as LIVE Challenge only have MOS annotations as labels. Besides, the goal of IQA is to predict

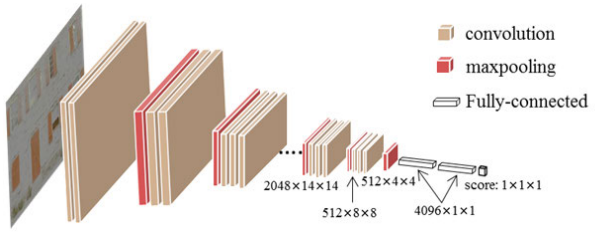


FIGURE 4. The structure of the generator in NQS-GAN.

the score. Although most of GAN related models generate multi-dimensional image, there is no relevant theory proving that it could only generate multi-dimensional data. Therefore, G is modified to generate the score instead of pixel-level quality map.

NQS-GAN is depicted in the middle of FIGURE 1. For the branch of G, since it is not necessary to generate the pixel-level quality map, deconvolution and convolution layers in NQM-GAN are replaced with the fully connected layers to regress a score. As is shown in FIGURE 4, the 224*224 distorted patch is forwarded to convolution and pooling layers to obtain a 4*4 feature map, then through a simple score regression network including three fully connected layers to get a score.

For the branch of D, the structure is also consistent with that of D in the NQM-GAN. Before forwarding to D, the deconvolution module is added to expand dimensions of quality score or MOS value to merge with the distorted image processing branch. The output of G is a score, which also avoids the problem of using equal-pixel-value MOS map as the target domain.

E. MODIFICATION OF LOSS FUNCTION

Loss function of BIQA-GAN is composed of the loss of G and the loss of D. In BIQA-GAN, the task of D is a binary classifier, which maps the input to the probability space which belongs to the real sample distribution. However, the real sample is not a natural image, but the gray-scale MOS map (saliency weighted MOS map or MOS value) corresponding to the distorted image. In this section, the NQM-GAN is taken as an example to introduce the loss function. The real sample in the NQM-GAN corresponds to the gray-scale MOS map. For the other two improved models NWQM-GAN and NQS-GAN, it is only necessary to replace the real sample with saliency weighted MOS map or MOS value, the rest are the same.

The optimization goal of D is to minimize the classification loss. The loss function of D is the binary classification cross entropy loss which is shown in Eqn.(3), where I_d represents the input distorted image; $G_\theta(I_d)$ represents the generated quality map;X_r represents the Ground Truth corresponding to $G_\theta(I_d)$; y represents the label of D, if the input is $G_\theta(I_d)$, y is 0;if the input is the Ground Truth,y is 1.

$$l_D = \min_X[-y\log(D_w(I_d, X_r)) - (1-y)\log(1 - D_w(I_d, G_\theta(I_d)))] \quad (3)$$

G aims at generating quality map. To let the quality map close to the corresponding Ground Truth pixel by pixel, the loss function of G not only includes the classification error fed back from D, but also the mean absolute error (MAE) which is used for pixel error penalty. As is shown in Eqn.(4), where α and β is the parameter to balance feedback error from D and MAE loss. The optimization goal of G is to minimize the pixel error between quality map and MOS map, meanwhile maximize correct classification error of D.

$$l_G = \min_G \max_D[\alpha \cdot \|X_r - G_\theta(I_d)\| + \beta \cdot [-\log(1 - D_w(I_d, G_\theta(I_d)))]] \quad (4)$$

Compared with the loss function of GAN [28], it can be seen that the loss function of D does not change in BIQA-GAN, while the loss function of G adds a MAE loss. The addition of MAE loss can further limit the distance between quality map and MOS map. In addition to penalizing pixel error, MAE ensures that G would not produce severe vanishing gradient when the performance of D reaches saturation. In GAN [28], the gradient of G is only provided by D, which is easy to be affected by D, thus resulting in vanishing gradient and stopping learning. For our proposed loss function of G, even if D fails to provide the classification error gradient, there still exists MAE loss to ensure its gradient not zero.

F. MODEL TRAINING PROCESS

The BIQA-GAN applies back propagation algorithm for random batches. N is the sample numbers for each input. The training algorithm of NQM-GAN is implemented below as the representative.

G. IMPLEMENTATION DETAILS

All the proposed networks are implemented in MXNet. To evaluate the performances of the proposed models, most experiments are carried out on LIVE Challenge. In the experiment, each database is randomly divided into 80% for training and 20% for testing, which ensures that the images in test sets never exists in training sets. For example, LIVE Challenge contains a total of 1,162 natural distorted images, where 80% of the dataset (nearly 930 images) are utilized as the training set, and the remaining 20% (nearly 232 images) as the test set. In order to augment the numbers of the training samples, each distorted image is split to 224*224 image patches with the stride of 40. For each patch clipped from the same distorted image, the MOS (or DMOS) value is set with the same MOS(or DMOS) value with the original distorted image. For optimization, the adaptive moment estimation optimizer (ADAM) is applied. Considering the local contrast normalization would have influence on the contrast variation distortion, which is common in natural distorted images, the mean subtraction method is utilized for data preprocessing, and ImageNet pretrained initialization is utilized for ResNet101 structure in G. Since all the patches of the distorted image are treated equally, the final predicted score

Algorithm 1 Training Process of NQM-GAN

Begin:
Initialize model parameters
For number of training epochs **do**:
a)Prepare the corresponding subjective MOS maps $\{ x_{1r}, x_{2r}, \cdots, x_{Nr} \}$
b)Sample the input distorted patches $\{ I_1, I_2, \cdots, I_N \}$
c)Input image patches to G to generate quality maps$\{ x_{1g}, x_{2g}, \cdots, x_{Ng} \}$
d)Update D:
1)Input distorted image patches, MOS maps into D to calculate classification error and stochastic gradient.

$$grad_{d1} = \nabla \theta_d \log (D(I_d, X_r)) \tag{5}$$

2)Input distorted image patches,quality maps into D to calculate classification error and stochastic gradient.

$$grad_{d2} = \nabla \theta_d [- \log (1 - D(I_d, X_g))] \tag{6}$$

3)Update network parameters according to $grad_{d1}$ and $grad_{d2}$ based on gradient descent method.
e)Update G:
1)Calculate MAE loss according to quality maps in c) and MOS maps.
2)Input distorted patches and quality maps into D, and calculate the feedback classification error.
3)Calculate the stochastic gradient of G.

$$grad_g = \nabla \theta_d [\left\| X_r - X_g \right\|_1 + \lambda \log (D(I_d, X_g))] \tag{7}$$

4)Update G according to $grad_g$ based on gradient descent method.
End for
Until End of training epochs

of the image is the mean value of all the predicted scores of patches in the image.

IV. EXPERIMENTS

A. DATASETS AND METRICS

Existing classical IQA databases include LIVE [18], TID2008 [19], TID2013 [20], CSIQ [21] and LIVE Challenge [29], [30]. The first four databases are synthetic distorted image datasets, which contain a small number of high-quality reference images, and the related distorted images. LIVE Challenge is authentically distorted image dataset. Most of our experiments are carried out on LIVE Challenge. Comparison experiments are also implemented on LIVE and TID2013.

1) LIVE DATABASE

The LIVE database [18] contains 779 distorted images distorted by 5 distortion types derived from 29 reference images.

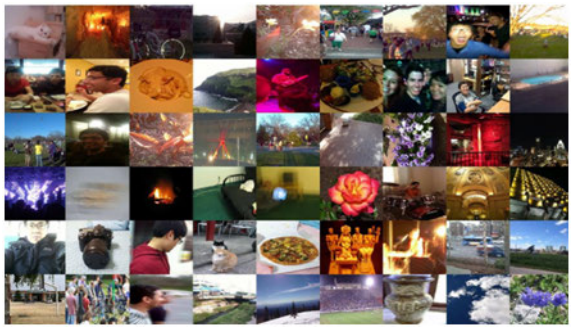


FIGURE 5. Samples from LIVE challenge.

The 5 distortion types are JPEG compression (JPEG), Fast-Fading (FF), White Noise (WN), Gaussian blur (GBLUR) and JPEG2000 compression (JP2K). A differential mean opinion score (DMOS) is assigned to each distorted image, approximately in the range from 0 to 100. A higher DMOS indicate lower perceptual quality.

2) TID2013 DATABASE

The TID2013 database [20] contains 25 reference images and 3000 distorted images with 24 distortion types. Each image is associated with an MOS in the range from 0 to 9. The first 17 distortion types in TID2013 are shared with TID2008 database. Owing to more distortion types and images, TID2013 is more challenging.

3) LIVE CHALLENGE DATABASE

LIVE Challenge [29], [30] contains 1,162 natural distortion images. The database contains a total of over 350000 subjective scores evaluated by a total of 8100 unique observers using Amazon online crowd sourcing marking system. FIGURE 5 show the samples from LIVE Challenge.

To evaluate the performance of our models, three correlation criterions are applied in our experiments, which are Spearman Rank Order Correlation Coefficient (SROCC), Pearson Linear Correlation Coefficient (PLCC), and Root Mean Square Error (RMSE).

SROCC is the index to measure whether the prediction score of the model is consistent with the trend of the subjective score, which is shown in Eqn.(8), where N is the total number of distorted images, d_n represents the sorting difference between the subjective score of the n-th image and the prediction score, larger SROCC value indicates more consistence with the human subjective evaluation.

$$SROCC = 1 - \frac{6 \sum_n d_n^2}{N(N^2 - 1)} \tag{8}$$

PLCC is the index to measure the linear correlation between the subjective evaluation and objective evaluation after nonlinear regression, which is depicted in Eqn.(9), where S_p represents the predicted quality score, and S_m represents the subjective score of the image, larger PLCC value indicates more consistent with the human subjective

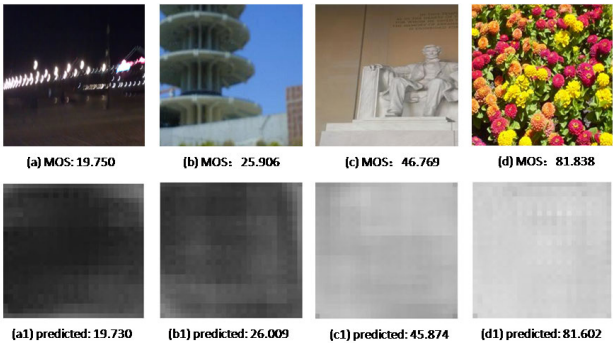


FIGURE 6. Natural distorted images and the quality maps generated by NQM-GAN. (a)-(d) are natural distorted images in LIVE Challenge with MOS value below; (a1)-(d1) are quality maps corresponding to the images in (a)-(d) with predicted score below.

evaluation.

$$PLCC = 1 - \frac{\sum_i (S_p - S_m)(\hat{S_p} - \hat{S_m})}{\sqrt{\sum_i (S_p - S_m)^2 \sqrt{(\hat{S_p} - \hat{S_m})^2}}} \tag{9}$$

RMSE is the index to measure the accuracy of the predictions after the nonlinear regression, which is shown in Eqn.(10), where S_p represents the predicted quality score, and S_m represents the subjective score of the image. A smaller RMSE value indicates more consistent with the human subjective evaluation.

$$RMSE = \sqrt{\frac{\sum_i (S_p - S_m)^2}{n - 1}} \tag{10}$$

B. QUALITY MAP AND WEIGHTED SALIENCY QUALITY MAP

1) QUALITY MAP

The samples of quality maps generated by NQM-GAN are shown in FIGURE 6. It is noted that the pixel value distribution is relatively concentrated in the quality maps. The reason is that the ground truth corresponding to the quality map is set to be the equal-pixel-value grayscale MOS map, so the target domain of NQM-GAN is the equal-pixel-value grayscale MOS map. For the quality map, little difference could be found in pixel level, presenting the equivalent gray distribution.

2) WEIGHTED SALIENCY QUALITY MAP

Saliency maps obtained by saliency detection method [52] and the weighted saliency quality maps generated by NWQM-GAN are shown in FIGURE 7. The brightness distribution of the weighted saliency quality map generated by NWQM-GAN is basically consistent with its corresponding saliency map obtained by saliency detection method [52], but in some areas it is not reasonable. For example, the bike in the darkness in FIGURE 7(a), street lights are saliency area in (a1). so it is with (a2). This is clearly not consistent with the strategy of applying saliency.

To judge the performances of NWQM-GAN influenced by undistorted reference image generated weighted saliency map, comparison experiments are carried out on LIVE and TID2013. From FIGURE 8, the weighted saliency quality map generated by NWQM-GAN could better reflect the degree of the distortion. In subjective perception, the distortion degree in (a4) is inferior to that in (a1), the brightness of the corresponding weighted saliency quality map in (a4) is significantly lower than that in (a1), which is in accordance with the corresponding image.

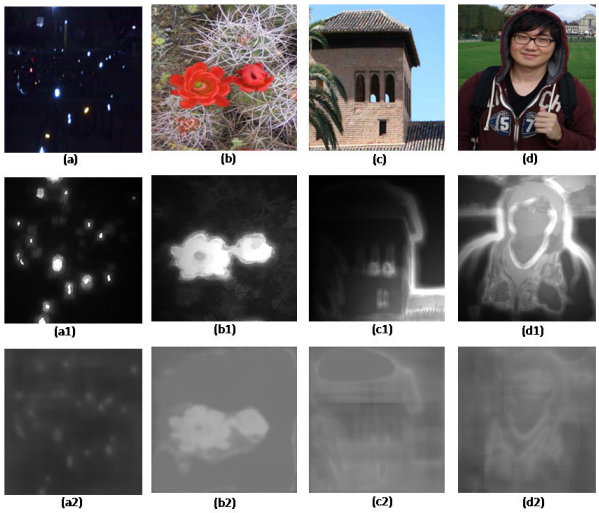


FIGURE 7. Natural distorted images and weighted saliency quality maps generated by NWQM-GAN. (a)-(d) are natural distorted images in LIVE Challenge; (a1)-(d1) are saliency maps obtained by saliency detection method [52]; (a2)-(d2) are the weighted saliency quality maps generated by NWQM-GAN.

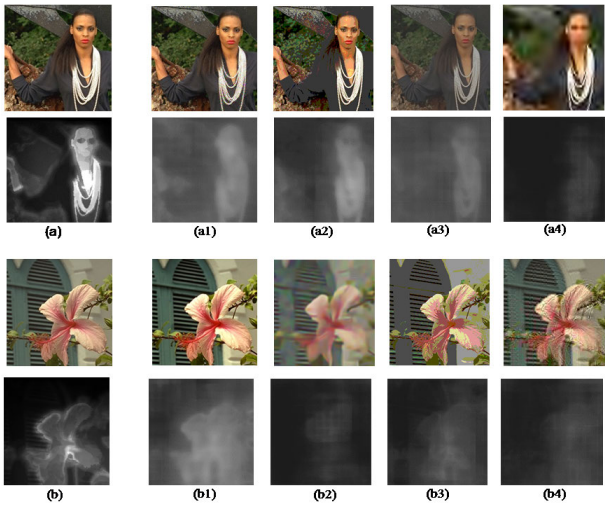


FIGURE 8. The weighted saliency quality maps generated by NWQM-GAN on LIVE and TID2013. (a) is undistorted reference image in LIVE and its corresponding weighted saliency quality map; (a1)-(a4) are different types of synthetic distorted images and their corresponding weighted saliency quality map; (b) is undistorted reference image in TID2013 and its corresponding weighted saliency quality map, (b1)-(b4) are different types of synthetic distorted images and their corresponding weighted saliency quality map.

C. ANALYSIS ON DIFFERENT DATASETS

As can be seen from TABLE 1, the correlation coefficients of NWQM-GAN in LIVE and TID2013 do not have

TABLE 1. The performances of the proposed models on different databases.

Methods	LIVE		TID2013		LIVE Challenge	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
NQM-GAN	0.926	0.953	0.861	0.890	0.810	0.840
NWQM-GAN	0.922	0.943	0.872	0.902	0.786	0.792
NQS-GAN	**0.985**	**0.986**	**0.901**	**0.908**	**0.869**	**0.893**

obvious gap with that of NQM-GAN, but in LIVE Challenge, the correlation coefficients of NWQM-GAN is obviously smaller than that of NQM-GAN. The reason is that the weighted saliency quality maps in LIVE and TID2013 are directly obtained based on the undistorted reference image. Weighted saliency quality maps in LIVE Challenge could only be obtained by the distorted images. Thus, using saliency weighted map which has been affected by the distorted image is bound to decline the performance. This is the reason why the performance of NWQM-GAN is worse than that of NQM-GAN on LIVE Challenge. Related literatures [53]–[55] about saliency in IQA have also proved that the weighted saliency quality maps should be obtained from the undistorted reference image directly. More reasonable quality weight combining the content and structure information of the image should be taken into account to get a better weighted saliency quality map.

Extensive experiments based on the proposed models are carried out on LIVE, TID2013, LIVE Challenge. Compared with NQM-GAN on LIVE Challenge, NQS-GAN achieves 7.3% and 6.3% improvement in SROCC and PLCC respectively. The performance of NQS-GAN is proved to be very effective to predict the score not only on synthetic distorted databases, but also on authentically distorted databases.

D. ANALYSIS BASED ON NQS-GAN

As is shown in Eqn.(4), α and β are the parameters of balancing the feedback error of D and MAE loss. To get a better NQS-GAN model, comparison experiments are carried out on LIVE Challenge utilizing different α and β. Different parameters represent different models, which are trained separately. As can be seen from TABLE 2, the best model is achieved when α is 0.05 and β is 1.0.

TABLE 2. Experimental results of NQS-GAN utilizing different parameters (α and β) to balance feedback error of D and MAE loss on LIVE Challenge. The bold indicates the best results in SROCC, PLCC and RMSE.

NQS-GAN	α=0.01 β =1.0	α=0.02 β =1.0	α=0.05 β=1.0	α=0.1 β =1.0	α=1.0 β =1.0	α=1.0 β =10.0
SROCC	0.8442	0.8667	**0.8693**	0.8506	0.8640	0.8657
PLCC	0.8822	0.8913	0.8929	0.8823	0.8906	**0.8936**
RMSE	10.965	9.244	**9.150**	10.950	10.000	10.124

E. PERFORMANCE COMPARISON

The implementation details of the proposed models have already illustrated in III(G). The dataset is randomly splited to 80% training set and 20% test set, and the median values of correlation coefficients are chosen from the results. The comparison performances of the proposed models and the existing superior IQA models on LIVE, TID2013 and LIVE Challenge are shown in TABLE 3. The proposed BIQA-GAN models are compared with 6 classical or state-of-the-art FR-IQA methods (DeepQA [2], SSIM [31], FSIM [32], GMSD [56], MDSI [57], WaDIQaM-FR [58]), 2 state-of-the-art RR-IQA methods (SDM [59],RIQMC [60]) and 18 classical or state-of-the-art NR-IQA methods (SDM [59], QAC [63],DIIVINE [24], BRISQUE [6], BLIINDS-II [61], NIQE [25], S3-index [26], C-DIIVINE [27], FRIQUEE [22], Luping [5], HFD-BIQA [64], R^3 [65], DIQaM-NR [58], BIECON [10], BPSQM [16], Hallucinated-IQA [17]). The top performances for FR and NR metrics are separately highlighted in bold. From TABLE 3, NQS-GAN has superior performance to state-of-the-art NR-IQA methods, even better than DeepQA [2] on LIVE, which is the state-of-the-art FR model. NQS-GAN also achieves a remarkable improvement against Hallucinated-IQA [17] on TID2013. NQM-GAN and NWQM-GAN also have competitive performance comparing with other models on LIVE Challenge.

DIIVINE [24], BRISQUE [6], BLIINDS-II [61], NIQE [25], S3-index [26], C-DIIVINE [27], FRIQUEE [22], BPSQM [16] are the part of the classical NR-IQA models. Although these models have superior performances in synthetic distorted datasets, when applied to the natural distorted datasets, the performances are greatly reduced. This indicates that the natural distortion is far more complex than the synthetic distortion. When these methods designed on synthetic distorted datasets transfer to natural distorted datasets, generalization performances would get worse. For LIVE Challenge, it can be seen that the performances of the proposed models are significantly better than the models designed on the synthetic distorted images, even better than FRIQUEE [22] and Luping [5], which are proposed specially for natural distorted datasets. Compared with correlation coefficient of BPSQM [16] on LIVE Challenge, NQS-GAN achieves almost 20% and 24% improvement in SROCC and PLCC respectively, which are up to 0.869 and 0.893, highly consistent with the human subjective perception. It is obvious that adversarial learning is competitive for IQA, especially for natural distorted images.

F. STATISTICAL SIGNIFICANCE TEST

Statistical significance test is the experiment to judge whether the performances among different models are significant. In the experiment, t-test is conducted based on the values of SROCC among different models, which are obtained from 80%train-20%test splits. Each pair of the models is compared on Live Challenge database, the results are listed in Table 4. NQS-GAN is statistically superior to the other no-reference algorithms we evaluated. Our proposed NQM-GAN and NWQM-GAN are also competitive comparing with other models.

G. SCATTER PLOTS

Comparison of scatter plots of subjective MOS against predicted scores obtained from proposed models on Live

TABLE 3. The comparison results of existing models and proposed models on LIVE, TID2013 and the LIVE Challenge. The top performance for FR and NR metrics are highlighted in bold. '-NA-' indicates that the value is not provided in the corresponding paper or the corresponding source code is not available.

Type	Metric	LIVE		TID2013		LIVE Challenge	
		SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
FR	SSIM [31]	0.948	0.945	0.637	0.691	-NA-	-NA-
	FSIM [32]	0.963	0.960	0.802	0.859	-NA-	-NA-
	GMSD [56]	0.959	0.938	0.801	0.854	-NA-	-NA-
	MDSI [57]	0.967	0.966	0.890	0.909	-NA-	-NA-
	WaDIQaM-FR [58]	0.970	0.980	**0.940**	0.946	-NA-	-NA-
	DeepQA [2]	**0.981**	**0.982**	0.939	**0.947**	-NA-	-NA-
RR	SDM [59]	0.936	0.933	-NA-	-NA-	-NA-	-NA-
	RIQMC [60]	-NA-	-NA-	0.804	0.865	-NA-	-NA-
NR	BIQI [62]	0.726	0.740	-NA-	-NA-	0.519	0.540
	QAC [63]	0.868	0.863	-NA-	-NA-	0.298	0.318
	DIIVINE [24]	0.916	0.917	0.549	0.654	0.509	0.558
	BRISQUE [6]	0.940	0.942	0.573	0.651	0.602	0.610
	BLIINDS-II [61]	0.920	0.923	0.536	0.628	0.405	0.450
	NIQE [25]	0.914	0.915	0.317	0.426	0.421	0.478
	S3-index [26]	0.944	0.943	-NA-	-NA-	0.310	0.320
	C-DIIVINE [27]	0.944	0.947	-NA-	-NA-	0.630	0.660
	FRIQUEE [22]	0.948	0.962	-NA-	-NA-	0.682	0.706
	Luping [5]	-NA-	-NA-	-NA-	-NA-	0.683	0.706
	HFD-BIQA [64]	0.951	0.948	0.764	0.681	-NA-	-NA-
	R^3 [65]	0.960	0.966	-NA-	-NA-	0.631	0.642
	DIQaM-NR [58]	0.960	0.972	0.835	0.855	0.601	0.606
	WaDIQaM-NR [58]	0.963	0.954	0.761	0.787	0.680	0.671
	BIECON [10]	0.961	0.962	0.721	0.761	-NA-	-NA-
	BPSQM-Fg [16]	0.971	0.961	0.828	0.856	0.719	**0.723**
	BPSQM-Fg-MD [16]	0.973	0.963	0.862	**0.885**	**0.724**	0.719
	Hallucinated-IQA [17]	**0.982**	**0.982**	**0.879**	-NA-	-NA-	-NA-
Proposed NR	**NQM-GAN**	0.926	0.953	0.861	0.890	0.810	0.840
	NWQM-GAN	0.922	0.943	0.872	0.902	0.786	0.792
	NQS-GAN	**0.985**	**0.986**	**0.901**	**0.908**	**0.869**	**0.893**

TABLE 4. Statistical significance test results on Live Challenge database. "1/0/−" respectively represents the model in row is statistically better than/worse than/indistinguishable with the model in column with 0.95 confidence.

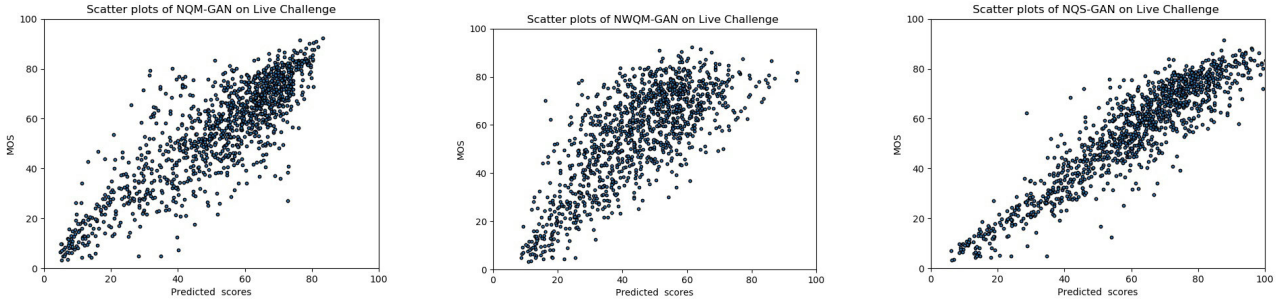
	NIQE	DIIVINE	BRISQUE	BPSQM-Fg	BPSQM-Fg-MD	NQM-GAN	NWQM-GAN	NQS-GAN
NIQE	-	0	0	0	0	0	0	0
DIIVINE	1	-	0	0	0	0	0	0
BRISQUE	1	1	-	0	0	0	0	0
BPSQM-Fg	1	1	1	-	-	0	0	0
BPSQM-Fg-MD	1	1	1	-	-	0	0	0
NQM-GAN	1	1	1	1	1	-	1	0
NWQM-GAN	1	1	1	1	1	0	-	0
NQS-GAN	1	1	1	1	1	1	1	-

Challenge database is shown in FIGURE 9. It can be seen that the predicted image quality scores show a nearly linear relationship with MOS values for the proposed models. However, the scatter plot of NQM-GAN is more compact than that of NWQM-GAN. Objective scores predicted by NQS-GAN is more correlated with subjective ratings than NQM-GAN and NWQM-GAN, where most samples compactly gather around the diagonal line. The scatter plots in FIGURE 9 are consistent with the comparison results of our proposed models in Table.3.

H. CONVERGENCE PERFORMANCE ANALYSIS

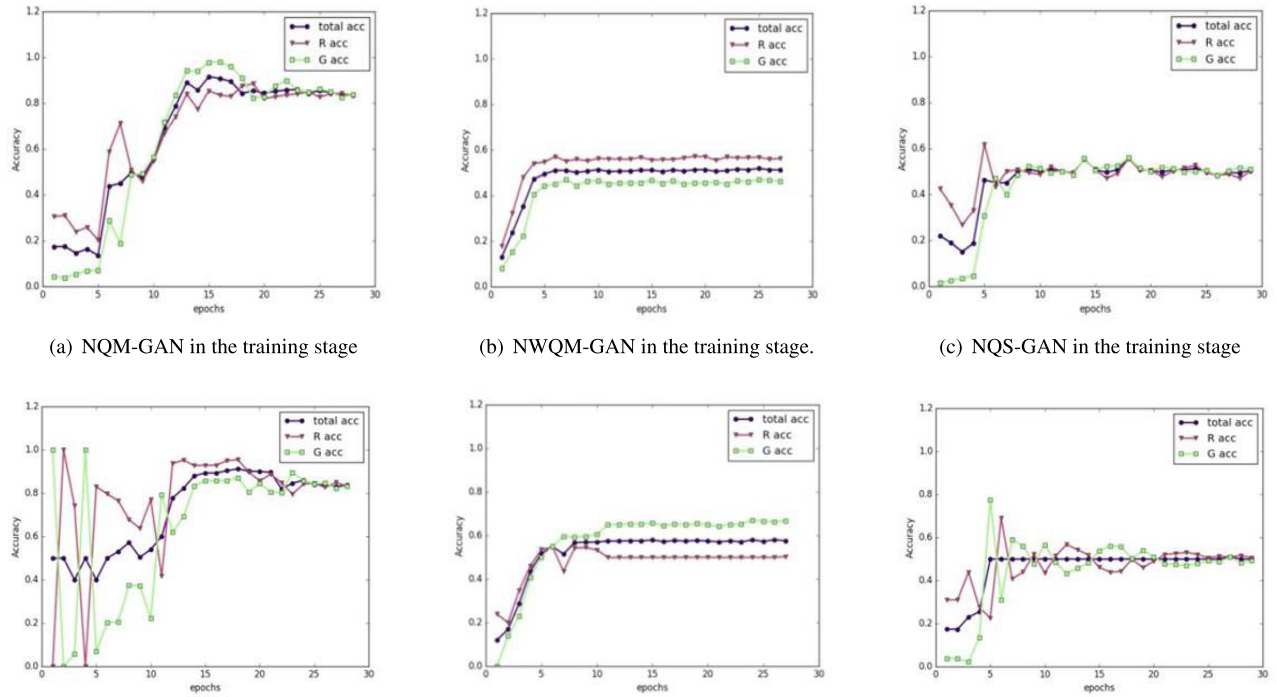
In order to analyze the influences of different target domains, the performances of G and D in proposed BIQA-GAN models are compared on LIVE Challenge. FIGURE 10 illustrates the classification accuracy curves of D in the training and testing stages for different models. The blue dot, red triangle and green square curves represent the classification accuracy for all samples, for real samples and for generated samples respectively. FIGURE 11 and FIGURE 12 depict the MAE of G in the training and testing stage on LIVE Challenge. MAE is the distance between the generated sample and the real sample, which reflects the prediction performance of the model.

From the adversarial learning mechanism in GAN, the goal of G is to generate samples to cheat D, and the final state of D is expected to classify all input samples randomly, thus the classification accuracy should be about 0.5. FIGURE 10 (a) (d) show that the performance of NQM-GAN deviates from the expected state in both training and testing stages. FIGURE 10 (b) (e) show that the classification accuracy of NWQM-GAN for either real or generated samples could converge to a stable state, but have the distance with 0.5, which indicates neither G nor D could converge to the expected distribution. FIGURE 10 (c) (f) illustrate the classification accuracy of NQS-GAN could conform to the



(a) Scatter plots of NQM-GAN (b) Scatter plots of NWQM-GAN (c) Scatter plots of NQS-GAN

FIGURE 9. Comparison of scatter plots of subjective MOS against predicted scores obtained from proposed models on Live Challenge database. The x-axis is the predicted score and the y-axis is the MOS value.



(a) NQM-GAN in the training stage (b) NWQM-GAN in the training stage. (c) NQS-GAN in the training stage

(d) NQM-GAN in the testing stage (e) NWQM-GAN in the testing stage (f) NQS-GAN in the testing stage

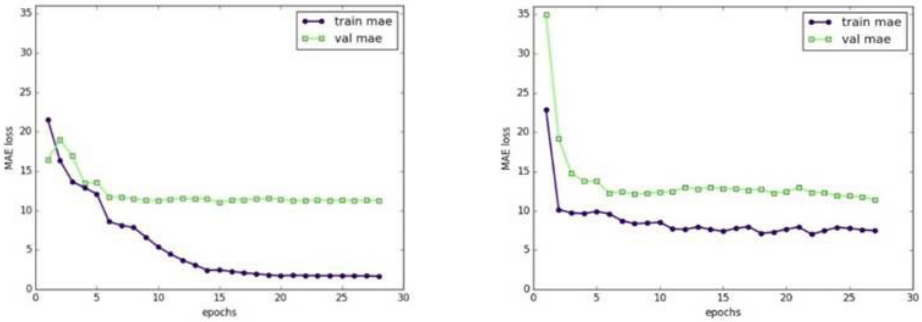
FIGURE 10. Classification accuracy of D in the training and testing stage of the proposed BIQA-GAN models.

expected state in both training and testing stages, which further proves the excellent performance in the proposed models. TABLE 3 shows that the correlation coefficients in NQS-GAN are significantly better than the other proposed models, which could be illustrated in FIGURE 10. The reason is that the convergence distribution of D in NQM-GAN and NWQM-GAN deviates from the expected distribution, which affects the prediction of the G. Further, the consistency between the predicted scores and the subjective scores decreases. As for NQS-GAN, the classification accuracy of D converges to nearly 0.5 for all samples, which is consistent with expected distribution. So validation MAE of G in NQS-GAN is lower than that in NQM-GAN and NWQM-GAN, which could be depicted in FIGURE 11 and FIGURE 12. Therefore, NQS-GAN could better predict the score.

I. COMPUTATIONAL COMPLEXITY ANALYSIS

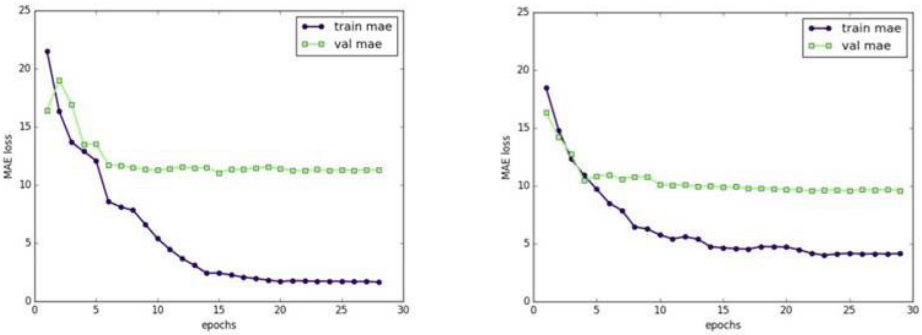
The image of size 500*500 in Live Challenge is split to 224*224 image patches with the stride of 40. Therefore, there are altogether 49 patches forwarding to the predicted network. The experiment is conducted on a PC with Intel Core i7-5930K CPU with 6 processor of speed 3.5 GHz*12, 62.8GiB Memory, NVIDIA GeForce GTX TITAN X GPU and Ubuntu14.04 64-bit. All the metrics were implemented with Pycharm. The program is running on only 1 GPU.

IQA-CNN [7] is the first attempt to implement CNN to IQA. BPSQM [16] is an excellent model with the state of the art performance in IQA at the time of publishing. The computational complexity comparison experiments are carried out among them. From Table 5, it can be seen that the computational time of NQM-GAN and NWQM-GAN are



(a) MAE curves of G in NQM-GAN. (b) MAE curves of G in NWQM-GAN.

FIGURE 11. Comparison of MAE of G in NQM-GAN and NWQM-GAN.



(a) MAE curves of G in NQM-GAN. (b) MAE curves of G in NQS-GAN.

FIGURE 12. Comparison of MAE curves of G in NQM-GAN and NQS-GAN.

TABLE 5. Computational complexity comparison on Live Challenge.

IQA Model	Time(sec/image)
IQA-CNN [7]	0.0704
BPSQM [16]	0.2519
NQM-GAN	1.3084
NWQM-GAN	1.4371
NQS-GAN	0.5872

very close. The reason is that the structures of the predicting networks are similar. As NWQM-GAN is a little complex in the multiplication of the MOS map and the saliency map, the time consumption is a litter longer. Comparing with them, NQS-GAN require less computing time, the reason for that is the post-structure of G in NQS-GAN is more simple than the post-structure of G in NQM-GAN and NWQM-GAN whose post-structure is a series of deconvolution and convolution layers stacking together. The structure differences could be seen from FIGURE 3 and FIGURE 4. NQS-GAN not only has excellent performance, but the computation time is less. Comparing with BPSQM [16], the proposed models are a little longer because of the different network structure. However, the computational complexity of all proposed models is acceptable.

V. CONCLUSION

In this paper, a natural distorted NR-IQA method was proposed in a new perspective, which is the first time that GAN related method applied to the natural distorted image quality assessment. Three different models were designed based on different target domains. The proposed BIQA-GAN models do not require any extra annotations or prior knowledge for training, only with its advantages of self-generating samples and self-feedback training to improve performances. Extensive comparison experiments on LIVE, TID2013 and LIVE Challenge could demonstrate that NQS-GAN has competitive performance on NR-IQA task, which has been proved to be superior to state-of-the-art models.Optimizing our proposed network to predict complex distorted image quality is a potential direction for future work.

REFERENCES

[1] Y. Liang, J. Wang, X. Wan, Y. Gong, and N. Zheng, "Image quality assessment using similar scene as reference," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016.

[2] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1676–1684.

[3] X. Min, G. Zhai, K. Gu, Y. Zhu, J. Zhou, G. Guo, X. Yang, X. Guan, and W. Zhang, "Quality evaluation of image dehazing methods using synthetic hazy images," *IEEE Trans. Multimedia*, to be published.

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

[5] L. Yang, H. Du, J. Xu, and Y. Liu, "Blind image quality assessment on authentically distorted images with perceptual features," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2042–2046.

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

[7] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1733–1740.
[8] K. Le, Y. Peng, L. Yi, and D. Doermann, "Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 2791–2795.
[9] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal Image Video Process.*, vol. 12, no. 2, pp. 355–362, Feb. 2018.
[10] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 206–220, Feb. 2017.
[11] X. Min, K. Ma, K. Gu, G. Zhai, Z. Wang, and W. Lin, "Unified blind quality assessment of compressed natural, graphic, and screen content images," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5462–5474, Nov. 2017.
[12] J. Wu, Z. Xia, H. Li, K. Sun, K. Gu, and H. Lu, "No-reference image quality assessment with center-surround based natural scene statistics," *Multimedia Tools Appl.*, vol. 77, no. 16, pp. 20731–20751, 2018.
[13] J. Wu, H. Li, Z. Xia, and Z. Xia, "Screen content image quality assessment based on the most preferred structure feature," *J. Electron. Imag.*, vol. 27, no. 3, 2018, Art. no. 033025.
[14] X. Min, G. Zhai, K. Gu, X. Yang, and X. Guan, "Objective quality evaluation of dehazed images," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2879–2882, Aug. 2019.
[15] J. Wu, Z. Xia, H. Zhang, and H. Li, "Blind quality assessment for screen content images by combining local and global features," *Digit. Signal Process.*, vol. 91, pp. 31–40, Aug. 2019.
[16] D. Pan, P. Shi, M. Hou, Z. Ying, S. Fu, and Y. Zhang, "Blind predicting similar quality map for image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6373–6382.
[17] K.-Y. Lin and G. Wang, "Hallucinated-IQA: No-reference image quality assessment via adversarial learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 732–741.
[18] L. C. H. R. Sheikh, Z. Wang, and A. C. Bovik. *Live Image Quality Assessment Database Release 2*. [Online]. Available: http://live.ece.utexas.edu/research/quality
[19] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008—A database for evaluation of full-reference visual quality assessment metrics," *Adv. Mod. Radioelectron.*, vol. 10, pp. 30–45, May 2009.
[20] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and J. C. C. Kuo, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. IEEE Eur. Workshop Vis. Inf. Process. (EUVIP)*, Jun. 2013, pp. 106–111.
[21] E. C. Larson and D. Chandler. (2010). *Categorical Image Quality (CSIQ) Database*. [Online]. Available: http://vision.okstate.edu/csiq
[22] D. Ghadiyaram and A. C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *J. Vis.*, vol. 17, no. 1, p. 32, 2016.
[23] G. Ying, F. Zhen-Qi, Y. Ying, and S. Feng, "No-reference quality assessment of authentically distorted images based on joint dictionary," *J. Optoelectron. Lase*, vol. 29, no. 1, pp. 105–112, 2018.
[24] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
[25] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2012.
[26] C. T. Vu, T. D. Phan, and D. M. Chandler, "S_3: A spectral and spatial measure of local perceived sharpness in natural images," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 934–945, Mar. 2012.
[27] Y. Zhang, A. K. Moorthy, D. M. Chandler, and A. C. Bovik, "C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes," *Signal Process., Image Commun.*, vol. 29, no. 7, pp. 725–747, Aug. 2014.
[28] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[29] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, Jan. 2016.
[30] D. Ghadiyaram and A. C. Bovik. (2015). *Live in the Wild Image Quality Challenge Database*. Accessed: Mar. 2017. [Online]. Available: http://live.ece.utexas.edu/research/ChallengeDB/index.html
[31] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[32] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[33] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1098–1105.
[34] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.
[35] C. Wang, C. Xu, C. Wang, and D. Tao, "Perceptual adversarial networks for image-to-image transformation," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4066–4079, Aug. 2018.
[36] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2223–2232.
[37] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 1857–1865.
[38] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4681–4690.
[39] K. Ehsani, R. Mottaghi, and A. Farhadi, "SeGAN: Segmenting and generating the invisible," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6144–6153.
[40] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1222–1230.
[41] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 469–477.
[42] C. Vondrick, H. Pirsiavash, and A. Torralba, "Generating videos with scene dynamics," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 613–621.
[43] J. Walker, K. Marino, A. Gupta, and M. Hebert, "The pose knows: Video forecasting by generating pose futures," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 3332–3341.
[44] S. Tulyakov, M.-Y. Liu, X. Yang, and J. Kautz, "MoCoGAN: Decomposing motion and content for video generation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1526–1535.
[45] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Belongie, "Stacked generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5077–5086.
[46] A. Dash, J. C. B. Gamboa, S. Ahmed, M. Liwicki, and M. Z. Afzal, "TAC-GAN-text conditioned auxiliary classifier generative adversarial network," 2017, *arXiv:1703.06412*. [Online]. Available: https://arxiv.org/abs/1703.06412
[47] C. Donahue, Z. C. Lipton, A. Balsubramani, and J. McAuley, "Semantically decomposing the latent spaces of generative adversarial networks," 2017, *arXiv:1705.07904*. [Online]. Available: https://arxiv.org/abs/1705.07904
[48] W. Yin, Y. Fu, L. Sigal, and X. Xue, "Semi-latent GAN: Learning to generate and modify facial images from attributes," 2017, *arXiv:1704.02166*. [Online]. Available: https://arxiv.org/abs/1704.02166
[49] L. Q. Tran, X. Yin, and X. Liu, "Representation learning by rotating your faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 12, pp. 3007–3021, Dec. 2019.
[50] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adversarial networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2089–2093.
[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[52] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1139–1146.
[53] J. Y. Lin, T. J. Liu, W. Lin, and C.-C. J. Kuo, "Visual-saliency-enhanced image quality assessment indices," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, Oct. 2013, pp. 1–4.
[54] Z. Zhang, J. Zhang, X. Wang, Q. Guan, and S. Chen, "Image quality assessment based on structural saliency," in *Proc. 19th Int. Conf. Digital Signal Process.*, Aug. 2014, pp. 492–496.

[55] Q. Zhou, X. Liu, L. Zhang, W. Zhao, and Y. Chen, "Saliency-based image quality assessment metric," in *Proc. 3rd Int. Conf. Syst. Inform. (ICSAI)*, Nov. 2016, pp. 918–924.
[56] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.
[57] H. Z. Nafchi, A. Shahkolaei, R. Hedjam, and M. Cheriet, "Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator," *IEEE Access*, vol. 4, pp. 5579–5590, 2016.
[58] 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 Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.
[59] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new reduced-reference image quality assessment using structural degradation model," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 1095–1098.
[60] K. Gu, G. T. Zhai, and M. Lin, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2015.
[61] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.
[62] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, Mar. 2010.
[63] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 995–1002.
[64] J. Wu, J. Zeng, Y. Liu, G. Shi, and W. Lin, "Hierarchical feature degradation based blind image quality assessment," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2017, pp. 510–517.
[65] Q. Wu, H. Li, Z. Wang, F. Meng, B. Luo, W. Li, and K. N. Ngan, "Blind image quality assessment based on rank-order regularized regression," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2490–2504, Nov. 2017.

HONGTAO YANG received the master's degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2010. He is currently pursuing the Ph.D. degree with the College of Information Engineering, Communication University of China. He is also a Teacher with the School of Electrical and Information Engineering, Beijing Polytechnic College. His main research interests include machine learning and image quality assessment.

PING SHI received the Ph.D. degree from the Communication University of China (CUC), in 2011. She is currently a Professor with the School of Information and Communication Engineering, CUC. She is also the Head of the Intelligent Media Computing Group. Her current research interests are in deep learning, intelligent media computing, and image quality assessment.

DIXIU ZHONG received the B.E. and master's degrees from the Communication University of China, Beijing, China in 2015 and 2018, respectively. Her research fields include deep learning and computer vision.

DA PAN received the B.E. and Ph.D. degrees from the Communication University of China, Beijing, China, in 2012 and 2018, respectively. His current research interests include image quality assessment and digital image processing.

ZEFENG YING is currently pursuing the Ph.D. degree with the College of Information Engineering, Communication University of China. His research fields include deep learning and image quality assessment.