No-Reference Image Quality Assessment Based on Multi-Task Generative Adversarial Network

YAO MA1, XIBIAO CAI1, FUMING SUN2, SHIJIE HAO3

1School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou, China
2School of Information and Communication, Dalian Minzu University, Dalian, China
3School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China

Corresponding author: X. Cai (e-mail: lgcaixb@163.com)

This work was supported in part by the National Natural Science Foundation of China under grant 61772171 and grant 61572244, in part by the Major Science and Technology Platform Project of the Normal Universities in Liaoning.

Abstract Since human observers are the ultimate receivers of an image, most of the image quality assessment (IQA) methods are based on analysis of the properties and mechanism of the human visual system. However, due to the lack of undistorted images for a reference, the accuracy of the no-reference IQA (NR-IQA) cannot compete with that of the full-reference IQA (FR-IQA). To bridge the performance gap between the FR-IQA and NR-IQA methods, we propose a NR-IQA method based on multi-task generative adversarial network, which attempts to restore dependable hallucinated images to compensate for the missing corresponding reference images. Two tasks, hallucination images and the quality maps are outputted by the generator and are combined with the specific loss to improve the reliability of hallucination images. Besides, two discriminator networks are used to respectively distinguish the undistorted images and hallucination images pairs, quality maps and structural similarity index measurement maps pairs. Finally, the hallucination images and distorted images are input into the IQA network, and quality scores are evaluated based on the differences between them. The superiority of our proposed method is verified by several different experiments on the LIVE datasets, TID2008 datasets, and TID2013 datasets.

Index Terms Generative adversarial network, hallucination images, quality maps, IQA network.

I. INTRODUCTION

Image quality assessment (IQA) is a fundamental task in computer vision and plays an important role in process evaluation, image encoding, and monitoring. Therefore, it is crucial to develop an effective image quality assessment method. The full-reference IQA (FR-IQA) algorithms use all the information on an undistorted reference image to evaluate image quality scores. In contrast, the no-reference IQA (NR-IQA) algorithms evaluate image quality without using any information on an undistorted image. Accordingly, the main advantage of the FR-IQA is that it can quantify visual sensitivity based on a difference between the distorted and reference images, which enables it to adopt the behavior of the human visual system (HVS) effectively. In contrast, due to the lack of the information on a reference image, most of the existing NR-IQA methods mainly try to extract features that can express the HVS process from the statistical characteristics and feature learning. Furthermore, for the HVS is a very complex process, it leads to the NR-IQA methods are more difficult to predict visual quality as effective as the FR-IQA methods. Therefore, the NR-IQA methods can difficulty achieve the same accuracy as the FR-IQA methods in terms of the correlation between subjective and objective quality values.

To bridge the performance gap between the FR-IQA and NR-IQA methods, we propose a NR-IQA method based on multi-task generative adversarial network (GAN), which attempts to restore dependable hallucinated images to compensate for the missing corresponding undistorted images. However, according to the related research work, like Deblur [1], Super-SR [2], DCGAN [3], etc, it can be concluded that the existing GAN technology has low performance in capturing the detailed texture information lost in the distorted image, which makes the generation of high-quality high-definition images become a great challenge, and it becomes even more difficult in the case of the IQA, which has a small amount of data and a high degree of complexity.

To reduce the difficulty of the GAN and improve the
quality of hallucination images, we fully combine the prior knowledge to enable the network to learn in a predetermined direction. Inspired by [4], we introduce multi-tasking into the generator to generate hallucination image and quality map simultaneously. Among the two tasks, the ground truth of the hallucination image is the undistorted image in datasets (in the NR-IQA algorithm, undistorted images can be used in training), and the ground truth of the quality map is obtained by the structural similarity index measurement (SSIM) algorithm [5]. The SSIM is a typical mature FR-IQA algorithm based on image brightness, contrast, and structure, which can represent certain distortion characteristics well. Therefore, we use the quality map as an auxiliary circuit to ensure the two tasks are positively correlated so that the GAN approach can grasp distortion characteristics better.

Because the ground truth of the IQA is a subjective score observed by human beings, we need to take full account of the perceptual characteristics that are in line with human subjectivity. The SSIM algorithm cannot completely simulate human perception from three aspects. Therefore, in addition to learning some perception features from the SSIM map, we integrate specific IQA-loss in the overall network loss. This IQA-loss denotes a result obtained from the feature extraction of the IQA network, which can reflect perception characteristics extracted from the IQA network. Besides, it can make a network not be confined to the three aspects of the SSIM map and can meet the human perception characteristics better. Finally, we use the learned reliable hallucination images to transform the NR-IQA into the FR-IQA problem, so that we can achieve or even surpass the accuracy of the FR-IQA.

The advantages of the proposed method are three-fold.

1. Employing a GAN method to generate dependable hallucination images makes it possible to analyze the characteristics of human subjective visual sensitivity based on the difference between distorted and reference images, so as to solve the problem that the NR-IQA cannot simulate the HSV process from the root.

2. Combining prior knowledge, the generator is converted into a multi-task network to generate hallucination images and quality maps. The quality maps as auxiliary circuit to make the undistorted high-definition images can be restored better, and the generated hallucination images can be more reliable.

3. Adding specific perceptual loss makes it possible to overcome the limitation of the SSIM and address the IQA problem better so that the NR-IQA performance can compete with that of the FR-IQA.

II. RELATED WORKS

A. FR-IQA METHODS

The FR-IQA methods use undistorted reference information, so most of the research work is based on the difference between a distorted image and its corresponding undistorted image. For instance, the SSIM algorithm calculates the structural similarity from three aspects: brightness, contrast, and structure, to judge distortion degree. Recently, many scholars have proposed many improvements on this basis, such as mean structural similarity index measurement (MSSIM) [6], feature similarity index measurement (FSIM) [7], information weighting for SSIM (IW-SSIM) [8], and many others. When introducing auxiliary tasks, we don’t want to make the network structure more complex, so we adopt the traditional feature extraction based on natural statistical characteristics.

As the most classical method, SSIM is not only simple and fast in calculation, but also the benchmark of many subsequent methods. It can be assumed that SSIM has sufficient capacity to represent certain distortion characteristics. Therefore, the introduction of SSIM as an auxiliary task will not increase the complexity, but also play an adequate auxiliary role.

In recent years, feature learning has gradually replaced manual feature extraction. Many FR-IQA methods based on deep learning have been proposed to improve the accuracy to a higher level. For instance, Bosse S. et al. [9] used a convolution layer to extract the features of distorted and undistorted images to obtain the difference between them. The IQA network we present in this work uses a similar approach. Bampis, C. et al. [10] computed the DLM features at four different scales and inputted them into the SVR. Xue W. et al. [11] used the global gradient variation based on the local quality map for overall image quality prediction.

B. NR-IQA METHODS

Due to the lack of a reference image, the NR-IQA methods can only use a distorted image to extract the statistical characteristics of warps caused by distortion. Based on the methods used for feature extraction, the NR-IQA methods can be roughly divided into two categories, those based on natural statistical characteristics and the others based on feature learning. However, changes in natural statistical characteristics can be manifested in many ways; for instance, in [12], in the spatial domain, the regional mutual information of different subsets was calculated based on information lost due to distortion to predict the quality scores. Oszust M. et al. [13] converted the RGB image into the YCbCr color space to extract the local features from the key points. Liu, Y. et al. [14] extracted the statistical measurements from three representative aspects of structure, naturalness, and perception to unsupervised learning. In the wavelet domain, Moorthy A. et al. [15] considered that image distortion could affect the subband coefficients of the wavelet transform, so the generalized Gauss distribution and coefficients of Daubechies wavelet transform were used as distortion characteristics. Sad et al. [16] stated that the degree and type of image distortion had been closely related to discrete cosine transformation (DCT) coefficients, so they extracted features in the DCT domain to predict the quality score.

All the aforementioned algorithms are based on a hypothesis that distortion changes some of the statistical characteristics of natural images, but these characteristics can be difficultly determined comprehensively and accurately manually.
With the development of deep learning, feature learning has gradually replaced manual feature extraction. For instance, Hou et al. [17] proposed a blind IQA model that classifies distortion into five levels for qualitative analysis and predicts numerical scores by using regression. Kim et al. [18] used four classical FR-IQA algorithms to get the quality score of image patches, which was further used as the ground truth in supervised learning. Talebi H. et al. [19] extracted the mean and score deviation values to predict the distribution of human opinion scores by using a convolutional neural network. Due to the lack of enough training data, various data enhancement methods have been used in these works. However, the enhanced datasets introduce a lot of noise, which has a negative impact on accuracy. S. Jia et al. [20,21] computed saliency map for each image and assigned importance value based on its saliency patch for each patch, and can be applied cross High Dynamic Range (HDR) images and Standard Dynamic Range (SDR) domains. Ma et al. [22] collected real-world distorted images and used unsupervised learning to realize the NR-IQA. Liu et al. [23] used a series of image pairs with a known rank order to train the network so that the network had the ability to distinguish image quality. Due to the lack of reference images, unsupervised learning can difficultly quantify the degree of learning distortion as a score, resulting in a low final measurement score.

To address the mentioned problems, inspired by [24], we combine the multi-task GAN method to generate dependable hallucination images and solve the problem that NR-IQA cannot simulate the HSV process from the root so that the effect of NR-IQA can compete with that of the FR-IQA.

III. PROPOSED METHOD

We proposed a NR-IQA method based on multi-task generative adversarial network, which is divided into two steps. First, we train the generator by inputting distorted and undistorted images into the multi-task GAN network, and get the hallucination image and quality map. Second, the hallucination images and distorted images are input into the IQA network, and quality scores are evaluated based on the differences between them. The overall framework of our proposed approach is shown in Fig. 1.

The structure of the proposed multi-task GAN network is shown in Fig. 2, wherein it can be seen that the network consists of a multi-task generator, G, and two discriminators, Dr and Ds, each of which evaluates the generated images from a different task.

A. GENERATOR, G

The generator, G, is made up of three parts: hallucination image, quality map and IQA-loss, which we will describe in detail in this section.

Firstly, as mentioned previously, we use the GAN method to generate dependable hallucination images to compensate for the drawback that there is no available reference image in the NR-IQA, so this is an image-to-image work. Therefore, we adopt the Pix2Pix [25] as a baseline of the multi-tasks GAN approach. In the generator, G, the network adopts U-Net [26] structure, which inputs distorted images and outputs hallucination images. Given a series of distorted images \( I_d \), as well as random noise vector \( z \), to learn their corresponding undistorted reference \( I_r \), we can get the loss that we want to optimize, which is given by:

\[
L_r(I_d, I_r) = L_{GAN}(I_d, I_r) + \lambda L_{L1}(I_d, I_r), \tag{1}
\]

\[
L_{GAN}(G, D_r) = E_{I_d \sim p_{data}(I_d)}[\log D_r(I_d, I_r)] + \frac{1}{2} \int E_{I_d \sim p_{data}(I_d), z \sim p_z}\log(1 - D_r(I_d, G(I_d, z))) dz, \tag{2}
\]

\[
L_{L1}(G) = E_{I_d \sim p_{data}(I_d), z \sim p_z}[||I_r - G(I_d, z)||_1], \tag{3}
\]

where \( L_{GAN}(\cdot) \) means GAN loss [25], which maximizes learning to classify real and synthesized pairs for \( D \) and minimizes learning to fool the discriminator for \( G \). Previous approaches have found it beneficial to mix the GAN objective with a more traditional loss, such as L2 distance. In the actual implementation, we adopt the calculation method of mean squared error (MSE). \( D_r \) denotes a specific discriminator for distinguishing hallucination images and target images, as explained in Section B; \( L_{L1}(\cdot) \) denotes the rebuild loss, showing the distance between true and false, to help the G to learn the characteristics of distortion. In this work, the learning coefficient \( \lambda \) is set to 100 based on the Pix2Pix.

However, to restore the undistorted images, training a GAN structure by learning only hallucination images is not enough to achieve the desired performance because of two reasons. First, there are 24 different types of distortion in the TID2013 dataset. Obviously, different types of distortion have more or less different characteristics. Thus, a GAN structure not only needs to learn the characteristics of each distortion type but also needs to restore hallucination images according to the different distortion types. Second, hallucination images attempt to restore the original high-definition undistorted image, but some of these wrapped textures, structures, and details are irreversible. Even when distortion characteristics are known, it is difficult to reproduce the undistorted image.

Secondly, we extend a single-task network to two tasks and use similarity and difference between the two tasks to...
promote each other. As shown in Fig. 2, in auxiliary circuit, we use the learned hallucination images and the distorted image as input to learning the structural similarity (SSIM) map between them. The similarity between the two tasks is that both are learning characteristics of distortion. The difference is that the hallucination image learning tends to restore some details, edges, textures and other requirements of more detailed work, while the quality map only needs to learn the three aspects of brightness, contrast, and structure, which will narrow the scope of learning and make it rough. These two tasks are mutually reinforcing so that the generator can output satisfactory hallucination images and quality maps simultaneously.

Therefore, similar to the loss in the hallucination image task, we add a loss to the quality map task in the generator. Given a series of distorted images \( I_d \) and the hallucination images \( I'_d \), to learn their corresponding SSIM map \( I_s \), we can get the loss that we want to optimize, which is given by:

\[
L_s(I_d, I_s) = L_{sGAN}(I_d, I_s) + \lambda L_{sL1}(I_d, I_s),
\]

\[
L_{sGAN}(G, D_s) = E_{I_d \sim \text{data}}[\log D_s(I_d, I_s)] + E_{I_d \sim \text{data}}[\log(1 - D_s(G(I_d, I'_d)))],
\]

\[
L_{sL1}(G) = E_{I_d \sim \text{data}}[\| I_s - G(I_d, I'_d) \|_1],
\]

where \( D_s \) denotes a specific discriminator for distinguishing quality maps and target maps, as explained in Section B.

Finally, the IQA method based on deep learning aims to make a network be able to judge the quality score consistent with the human subjective score by feature learning of differently distorted images. Human observation of distorted images not only comes from a difference between the pixels of an undistorted image, but also depends on the overall perception level of the image. Intuitively, the quality map of the auxiliary circuit can be utilized to calculate the perception term. However, because the ground truth of the auxiliary circuit relates to only three aspects, it cannot completely simulate human perception characteristics and has certain limitations for various distortion characteristics. To solve this problem and address the IQA problem better, we integrate specific IQA-loss into the G network.

\[
L_{iqa}(I_d, G(I_d)) = \sum_{i=1}^{N} \frac{1}{H_j} \sum_{j=1}^{W_j} \sum_{y=1}^{H_j} \| \phi_j(G(I_d))_{x,y} - \phi_j(I_r)_{x,y} \|^2.
\]

In Eq. (7), \( \phi_j(x) \) denotes the feature map at the \( j^{th} \) layer of the IQA network, \( W \) and \( H \) represent the dimensions of the feature map, and \( N \) represents the number of feature maps. Since the IQA network is trained for the FR-IQA tasks, the feature map of each layer contains different level perception features of distortion. The activations from the layers of a pre-trained FR-IQA regression network capture the distortion information of the input, which ensures the quality similarity measurement between the output of generator and the ground truth.

### B. DISCRIMINATOR, \( D \)

According to the adversarial theory of GAN method, hallucination images and quality maps are generated simultaneously by multi-task generators, so we need a discriminator to distinguish between the undistorted images and hallucination images, and between the quality maps and SSIM maps. As mentioned earlier, there are many differences between the quality maps and hallucination images, so if only one discriminator is used to distinguish the two groups at the same time, the performance of the discriminator will be degraded. The generator ability to capture various aspects of natural images is proportional to the ability of each discriminator to discern images. Therefore, we decide to use two discriminators to capture different characteristics of distorted images, and they are denoted as \( D_r \) and \( D_s \).
Algorithm 1: No-Reference Image Quality Assessment Based on Multi-Task Generative Adversarial Network

Input: Distorted image \( I_d \)
Output: The predicted quality score of a distorted image \( s \)

Training
1) Get an SSIM map \( I_s \) from a distorted image \( I_d \) and an undistorted reference \( I_r \);
2) Input \( I_d \) into the GAN structure and get hallucination images \( I'_d \) and quality map \( I'_s \) by Eq. (6);
3) Input \( I_d \) and \( I'_r \) into the IQA network to train the network.

Test
1) Input \( I_d \) into the trained generator and get
2) Input \( I_d \) and \( I'_d \) into trained IQA network and get a predicted score \( s \).

C. IQA NETWORK

For the IQA network, we use distorted images and learned hallucination images as datasets, and the human subjectivity score of distorted images is the ground truth, which is typical supervised learning. So, for the test image, the first step is to get the dependable hallucination image by the trained GAN model, and then input both hallucination image and distorted images into the IQA network to regression the quality score. The IQA network is similar to the FR-IQA network presented in [9], which consists of three parts: feature extraction, feature fusion, and regression, as shown in Figure 3. In addition, because the number of image contained in the dataset is insufficient for training, we crop the image to the size of 32*32 without overlap to the augmented dataset. In this network, both distorted images and hallucination images are input into the feature extraction part, including four layers of convolution, while a maxpool layer is placed after every two convolution layers. The fusion layer connects the features of distorted images (f1), hallucination images (f2) and the difference between them (f1-f2) as the final features. Finally, the predicted score is obtained by regression network, which consists of four fully-connected layers.

IV. TRAINING

First, we randomly crop the original distorted image and the corresponding undistorted image to the size of 256x256 and get the SSIM map through the SSIM algorithm as the training datasets of the multi-tasks GAN approach. In the generator network, we use a U-Net [26] with seven down-sampling layers and seven up-sampling layers to generate hallucination images. Then, the input distorted image and the generated hallucination image are used to obtain the quality map with a channel number of 1 through a two-convolution-layer network. In the two discriminators, except for the different number of channels in the input layer, a six-layer down-sampling network is used to distinguish true and false. Because of their different tasks, the representations of kernels within the two networks also toward to preserve different information. For optimization, the adaptive moment estimation optimizer (ADAM) [27] with \( \alpha=0.0002, \beta_1=0.5, \) and \( \beta_2=0.999 \) is employed. All weights are initialized from a zero-centered normal distribution with a standard deviation of 0.02. The GAN network and the IQA network have the same training dataset and test dataset, and the ratio of training set is 0.8, the test set is 0.2.

The proposed NR-IQA algorithm based on the multi-task GAN model is given in Algorithm 1.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATASETS

The following three datasets were used in the experiments, LIVE dataset, TID2008 dataset, and TID2013 dataset.

The LIVE dataset consisted of a total of 779 distorted images, including 29 reference images and five types of distortion. The TID2008 dataset consisted of a total of 1700 distorted images, including 25 reference images and 17 types of distortion; four of these 17 distortion types were common.
to the LIVE dataset, namely gblur, jp2k, jpeg, and wn. The TID2013 dataset represented an extension of the TID2008 dataset, wherein 7 types of distortion and 25 distorted images were added for each type of distortion compared to the TID2008 dataset.

### B. EVALUATION CRITERIA

Following most previous works, three evaluation criteria were adopted: the Pearson linear correlation coefficient (PLCC), Root mean squared error (RMSE), and Spearman rank-order coefficient (SROCC). The SROCC reflected the prediction monotonicity, when was close to 1, high performance of a specific quality measure was indicated. The formula is as follows:

$$SROCC = 1 - \frac{6}{n(n - 1)} \sum_{i=1}^{n} d_i^2/(n( n^2 - 1)).$$  \hspace{1cm} (9)

where \(n\) denotes the number of data pairs in data samples, and \(d_i\) denote the rank in the data sample. The PLCC considered the linear correlation coefficient between the two sets of data, assumed that there are two sets of data for \(X\) and \(Y\), which are the predicted scores and the ground labels in the image quality assessment. The formula is as follows:

$$PLCC = COV(X,Y)/\left(\delta_X \delta_Y\right).$$  \hspace{1cm} (10)

where \(COV()\) denote the covariance of probability theory, \(\delta\) denote standard deviation. The RMSE was used to quantify accuracy. The RMSE value close to 0 indicated a high similarity between predicted score and human subjective score.

### C. WINDOWS SIZE

To obtain the ground truth of the SSIM map, we calculated the structural similarity between the distorted images and the corresponding undistorted images, which required local Gaussian filtering of images. The window size is important because it determines the range of the affected surrounding pixels. Therefore, we compared different window sizes of local Gaussian filter on #01 distortion type in the TID2013 dataset. As shown in Table 1, in the experiments, when size is 5, the value of SROCC and PLCC are the highest, and the value of RMSE is lowest, so a fix window size of 5 was used.

### D. ABLATION STUDY

To investigate the effectiveness of the key components of our method, we conducted the ablation experiments on the TID2008 dataset. The overall results are shown in Table 2. The same training dataset and test dataset and the same IQA network was used for different GANs to ensure that the impacts of different losses can be compared. That is, we input the distorted images and hallucination images generated under different conditions into the same IQA network. The SROCC and PLCC were used as criteria to verify the effectiveness of different loss to the NR-IQA performance; the results are shown in Table 2.

### Table 2. The SROCC and PLCC on TID2008 dataset

<table>
<thead>
<tr>
<th>LOSS</th>
<th>SROCC</th>
<th>PLCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM</td>
<td>0.775</td>
<td>0.809</td>
</tr>
<tr>
<td>FHIM</td>
<td>0.880</td>
<td>0.874</td>
</tr>
<tr>
<td>SCORER</td>
<td>0.886</td>
<td>0.889</td>
</tr>
<tr>
<td>BQME</td>
<td>0.698</td>
<td>0.747</td>
</tr>
<tr>
<td>PM</td>
<td>0.832</td>
<td>0.868</td>
</tr>
<tr>
<td>SDCNN</td>
<td>0.885</td>
<td>0.887</td>
</tr>
<tr>
<td>Lr</td>
<td>0.843</td>
<td>0.859</td>
</tr>
<tr>
<td>Lr + Ls</td>
<td>0.879</td>
<td>0.891</td>
</tr>
<tr>
<td>Lr + Ls + Liqa</td>
<td>0.921</td>
<td>0.928</td>
</tr>
</tbody>
</table>

In Table 2, \(Lr\) denotes the base GAN as baseline, \(Lr + Ls\) denotes that the quality map will be put into the network to form a multi-task GAN; \(Lr + Ls + Liqa\) denotes multi-task GAN combined with specific IQA-loss. It can be seen that \(Ls\) and \(Liqa\) could obviously improve network accuracy. The multi-task GAN \((Lr + Ls)\) approach could learn undistorted images better; the IQA-loss \((Lr + Ls + Liqa)\) could make hallucination images more consistent with human perception and more realistic. Furthermore, the learning ability of the IQA network in quality assessment was studied in three situations, as shown in Figure 4, to compare the RMSE between the predicted score and human subjective score better. It can be seen that both \(Ls\) and \(Liqa\) made the gap between the predicted value and label decrease, and the RMSE of \(Lr + Ls + Liqa\) was the lowest, which indicated the higher ability of the IQA network to evaluate quality score and effectiveness of the SSIM map and the IQA-loss.

We enumerated the hallucination images and quality maps generated by the case of \(Lr + Ls + Liqa\), and the results are shown in Figure 5. In Figure 5, each row represents a different distortion type, and in each row, from left to
**FIGURE 5.** The distorted image, hallucination image, quality map, undistorted image, and SSIM map of our proposed method.
right are distorted image, hallucination image, quality map, undistorted image, and SSIM map. In Figure 5, it can be seen that under the specific loss, the multi-task GAN approach could better grasp the distortion characteristics, and the two tasks were positively correlated and mutually reinforcing. However, due to the various types of distortion, and because each distortion type had its unique characteristics, it was difficult to restore high-definition undistorted images for all distortion types even when the Ls and Liqa were combined. For some distorted images with serious blurring, too much texture details were lost, which made the network powerless. However, the proposed method could still evaluate the quality of these distorted images very well. Although the network could not restore high-definition images, hallucination maps could still grasp the distortion characteristics very well, especially by the Liqa. For instance, as presented in the last row in Figure 5, the quality map was accurately captured where the square effect occurred, so the SROCC value of this type could reach 0.658, which was competitive enough. However, compared with the other types of undistorted images that were easier to learn, the prediction accuracy was much lower.

E. SINGLE-DATASET EXPERIMENTS

The performance of the proposed method was also validated on the TID2013 dataset, and obtained results are given in Table 3. Table 3 shows the comparison results of the proposed method and state-of-the-art NR-IQA methods.

We compared the proposed method with the five representative NR-IQA methods: the HOSA [28], the Rank-IQA [23], the IL-NIQE[29], the SNP-NIQE[14], and the H-IQA [24]. In Table 3, it can be seen that the proposed method performed well for each distortion type. For 15 of a total of 24 types of distortion (over 60% subsets), our method achieved the highest accuracy among all the methods. Except for a few distortion types, the accuracy of the other methods reached almost the same level, which demonstrated the superiority of our proposed method over the other tested methods. For individual distortions, due to the randomness and uncertainty of various noises, the performance on a small number of distortion types such as #15 (Comfort noise) and #20 (Image denoising and non-eccentricity pattern noise) was lower than that of some methods. Specifically, the significant improvements of distortion types #10 (JPEG compression) and #21 (Lossy compression of noisy images) quantitatively demonstrated the effectiveness of our IQA-loss; also, improvements on types #17 (Contrast change) and #18 (Change of color saturation) verified that our auxiliary circuit enhanced the capacity of our generator to capture contrast shift and color change, while improvements on types #4 (Masked noise) and #23 (Chromatic aberrations) verified the ability of our multi-task GAN approach to generate hallucination maps under multiple distortions effectively.

To further verify the effectiveness of our proposed method, we conducted the experiments on the LIVE dataset, and the results are shown in Table 4. We compared the proposed method with the SSIM and seven representative NR-IQA methods: the SCORER [13], the PM [12], IL-NIQE[29], BIECON [18], BPSQM [9], SNP-NIQE[14], and the H-IQA [24]. The comparison results show that our proposed method was highly competitive with state-of-the-art NR-IQA methods. Although the results of our method on LIVE dataset are not the highest, the gap was only two percentage points. The results on TID2013 dataset are significantly better than others. We believe that the amount of data in TID2013 is much larger than that in LIVE dataset, so that our network can fully learn the distortion characteristics. The comparison result fully proves that our network can learn a good accuracy rate when the dataset is sufficient with data enhancement.

F. CROSS-DATASET EXPERIMENTS

In order to verify the generalization ability of our proposed method, we conducted cross-dataset experiments. We tested the multi-task GAN model trained on LIVE, TID2008, and TID2013 across all the three test datasets. Only the four types of distortions that are shared by LIVE, TID2008, and TID2013 are examined in this experiment.

The results are presented in Table 5, where it can be seen that when training and test were performed on the same dataset, the results were the best. When the cross-dataset testing was conducted, especially on the LIVE dataset, the values of SROCC and PLCC decreased significantly, which showed that the generalization ability of the network has to be improved significantly. However, this is a common problem of most NR-IQA algorithms; especially for real-world distorted images, which are often superimposed by multiple distortion types; thus, the network adaptability is still a great challenge.

VI. CONCLUSION

In this paper, a strong multi-tasking GAN approach for the no-reference image quality assessment is presented. First, the similarity and difference between an undistorted image and an SSIM map are used to construct a multi-task generator to generate the dependable hallucination image and quality map. Then, two discriminators are constructed according to different tasks to distinguish hallucination image and undistorted image, and quality map and SSIM map. At the same time, combined with the specific IQA-loss, the reliability of the generated images is enhanced through the adversarial learning, to compensate for the defect that the NR-IQA algorithm has no available undistorted reference in the test, and solve the problem that NR-IQA cannot simulate the HVS process from the root. The experimental results show that the proposed method achieves very good performances both in terms of consistency with human subjective scores and generalization ability. However, in the experiment of evaluating the effect of the method, we didnot divide the dataset according to the image content which might have some influence on the experimental results. And different distortion type has their unique characteristics, our network cannot achieve high accuracy for all distortion types. Therefore, a further in-depth study should be conducted to improve the generality of the...
TABLE 3. The SROCC on TID2013 dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>LIVE</th>
<th>TID2008</th>
<th>TID2013</th>
<th>TID2013</th>
<th>TID2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOSA</td>
<td>0.853</td>
<td>0.625</td>
<td>0.782</td>
<td>0.368</td>
<td>0.905</td>
</tr>
<tr>
<td>RankIQQA</td>
<td>0.667</td>
<td>0.620</td>
<td>0.821</td>
<td>0.365</td>
<td>0.760</td>
</tr>
<tr>
<td>IL-NIQE</td>
<td>0.876</td>
<td>0.816</td>
<td>0.923</td>
<td>0.512</td>
<td>0.808</td>
</tr>
<tr>
<td>SNP-NIQE</td>
<td>0.886</td>
<td>0.733</td>
<td>0.650</td>
<td>0.740</td>
<td>0.873</td>
</tr>
<tr>
<td>H-IQA</td>
<td>0.952</td>
<td>0.890</td>
<td>0.976</td>
<td>0.841</td>
<td>0.911</td>
</tr>
<tr>
<td>Our</td>
<td>0.977</td>
<td>0.957</td>
<td>0.943</td>
<td>0.957</td>
<td>0.985</td>
</tr>
</tbody>
</table>

TABLE 4. The SROCC and PLCC on LIVE dataset and TID2013 dataset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>SSIM</th>
<th>SROCC</th>
<th>SCORER</th>
<th>PM</th>
<th>IL-NIQE</th>
<th>DICEON</th>
<th>BPSQM</th>
<th>SNP-NIQE</th>
<th>H-IQA</th>
<th>Our</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIVE</td>
<td>SROCC</td>
<td>0.948</td>
<td>0.947</td>
<td>0.962</td>
<td>0.902</td>
<td>0.958</td>
<td>0.972</td>
<td>0.908</td>
<td>0.953</td>
<td>0.963</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PLCC</td>
<td>0.945</td>
<td>0.948</td>
<td>0.959</td>
<td>0.906</td>
<td>0.962</td>
<td>0.960</td>
<td>0.907</td>
<td>0.959</td>
<td>0.956</td>
<td></td>
</tr>
<tr>
<td>TID2013</td>
<td>SROCC</td>
<td>0.790</td>
<td>0.856</td>
<td>0.842</td>
<td>0.521</td>
<td>0.717</td>
<td>0.852</td>
<td>0.856</td>
<td>0.879</td>
<td>0.914</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PLCC</td>
<td>0.742</td>
<td>0.870</td>
<td>0.887</td>
<td>0.648</td>
<td>0.762</td>
<td>0.835</td>
<td>0.847</td>
<td>-</td>
<td>0.923</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 5. The SROCC and PLCC on different datasets

<table>
<thead>
<tr>
<th>SROCC</th>
<th>PLCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training dataset</td>
<td>Test dataset</td>
</tr>
<tr>
<td>LIVE</td>
<td>TID2008</td>
</tr>
<tr>
<td>Test dataset</td>
<td>LIVE</td>
</tr>
<tr>
<td>LIVE</td>
<td>0.959</td>
</tr>
<tr>
<td>TID2008</td>
<td>0.730</td>
</tr>
<tr>
<td>TID2013</td>
<td>0.751</td>
</tr>
</tbody>
</table>

proposed method, which will be the subject of our future work.

REFERENCES


Shijie Hao is currently an associate professor with School of Computer and Information Engineering, Hefei University of Technology (HFUT). He received his Ph.D. degree at HFUT in 2012. His research interests include image processing and multimedia content analysis.

Yao Ma is currently pursuing the B.Eng. (Bachelor of Engineering) degree with School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou, China. Her research interests include deep learning, image information processing, and image quality assessment, etc.

Xibiao Cai is currently an associate professor with School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou, China. He received the Ph. D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2014. His current research interests include wireless communication, image content analysis, pattern recognition, and nonlinear filter, etc.

Fuming Sun is currently a professor with School of Information and Communication Engineering, Dalian Minzu University, Dalian, China. He received his Ph.D. degree from the University of Science and Technology of China (USTC), Hefei, China, in 2007. From September 2012 to July 2013, he worked as a visiting scholar in Department of Automation, Tsinghua University. He was a Professor of School of Electronics and Information Engineering in Liaoning University of Technology from 2004 to 2018. His current research interests include content-based image retrieval, image content analysis, and pattern recognition.