IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 1, 2019, accepted July 19, 2019, date of publication July 24, 2019, date of current version August 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2930707

Siamese-Network-Based Learning to Rank for No-Reference 2D and 3D Image Quality Assessment

YUZHEN NIU[1,2], DONG HUANG[1], YIQING SHI[1], AND XIAO KE[1,2]

[1]Fujian Key Laboratory of Network Computing and Intelligent Information Processing, College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China

[2]Key Laboratory of Spatial Data Mining and Information Sharing, Ministry of Education, Fuzhou 350116, China

Corresponding author: Xiao Ke (kex@fzu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61672158 and Grant 61502105, in part by the Technology Guidance Project of Fujian Province under Grant 2017H0015, in part by the Industry-Academy Cooperation Project under Grant 2017H6008, in part by the Natural Science Foundation of Fujian Province under Grant 2019J02006 and Grant 2018J01795, and in part by the Fujian Collaborative Innovation Center for Big Data Application in Governments.

ABSTRACT 2D image quality assessment (IQA) and stereoscopic 3D IQA are considered as two different tasks in the literature. In this paper, we present an index for both no-reference 2D and 3D IQA. We propose to transform the IQA task into a task of quality comparison between images. By generating image pairs, the amount of training data reaches the square of the original amount of data, effectively solving the lacking of training samples. We also propose a learning to rank model using Siamese convolutional neural networks (LRSN) for quality comparison. The presented LRSN has two branches that have the same structure, share weights with each other, and take two image patches as inputs. The goal of LRSN is learning to rank the quality scores between the two input image patches. The relative quality score of a test image is obtained by first comparing its image patches with many image patches of other images and counts the number of times that its image patches are ranked superior to other patches. The experimental results on three 2D (LIVE, CSIQ, and TID2013) and three 3D (LIVE 3D Phase-I, LIVE 3D Phase-II, and NBU) IQA databases demonstrate that the proposed LRSN model works well for both 2D and 3D IQA and outperforms the state-of-the-art no-reference 2D and 3D IQA metrics.

INDEX TERMS No-reference image quality assessment, stereoscopic image quality assessment, Siamese convolutional neural networks, learning to rank.

I. INTRODUCTION

Digital images are usually distorted during acquisition, compression, and transmission. The distortions usually reduce the fidelity of the images. Therefore, image quality assessment (IQA) has been a topic of intense research in the fields of multimedia, image processing, and computer vision. In this paper, we use IQA refers to distortion related image fidelity quality assessment.

IQA metrics can be divided into two categories: subjective IQA and objective IQA. Subjective IQA is conducted by people. The process of subjective IQA is often complex, expensive, and time-consuming. Therefore, subjective IQA is difficult to apply in real applications, especially in real-time systems. Objective IQA aims to simulate the subjective perception of the human visual system (HVS) by means of mathematical models, machine learning, etc. Objective IQA can be divided into three types: no-reference, reduced-reference, and full-reference. Full-reference IQA requires the original image as a comparison to evaluate the quality of the distorted image. Reduced-reference IQA only needs partial information of the original image. No-reference IQA does not need the information of the original image which is usually unavailable in real applications. So no-reference IQA has broader application prospects than full-reference IQA and reduced-reference IQA. Therefore, we focus on no-reference IQA in this paper.

Early objective IQA metrics rely on hand-crafted HVS related features. Because of the limited understanding of HVS, the performance of conventional IQA metrics shows

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

a significant gap to the subjective perception. In recent years, deep learning, like Convolutional Neural Networks (CNN), has been successfully applied to many image processing and compute vision tasks, such as image recognition, super-resolution, and face recognition. Many researchers have used deep learning for 2D [1]–[3] and 3D [4] no-reference IQA. The IQA metrics based on deep learning can automatically extract features without relying on the understanding of HVS. Moreover, the IQA metrics based on deep learning have effectively reduced the gap between objective IQA and subjective perception.

A 3D image consists of a pair of monocular views, i.e. left and right views, taken by two cameras to simulate human binocular vision. The main factors that affect 3D image quality include fidelity, aesthetics, and visual comfort. Distortion-free and highly appealing 3D images may still be considered to be of visually low-quality if they have low visual comforts. Some 3D image aesthetics and visual comfort assessment metrics have been presented. Aesthetics assessment and visual comfort assessment metrics focus on aspects different from image fidelity quality assessment. In this paper, we focus on image fidelity quality assessment and use IQA for image fidelity quality assessment when without introducing ambiguity. Please refer to the survey [5] for more details on aesthetics assessment and visual comfort assessment.

Compared with objective 2D IQA, stereoscopic 3D IQA is more complicated, and it is necessary to consider the interaction between the left and right views. 3D IQA metrics can be mainly divided into three categories according to the type of information they utilize: 1) metrics that only consider left and right views; 2) metrics that consider depth/disparity information; and 3) metrics that consider binocular characteristics. In the literature, 2D IQA and 3D IQA are considered as two different tasks. In this paper, we represent a 3D image using a 2D image and present an index for both 2D and 3D no-reference IQA.

Despite the performance improvements achieved by the CNN-based IQA models, lacking training samples is one of the challenges for CNN-based objective IQA [5]–[7]. Existing CNN-based no-reference IQA models solve this problem by two types of methods. The first type of methods [1], [2] divides the image into a number of image patches, and assigns the subjective mean opinion score (MOS) of an image to its image patches. However, using MOS as the quality of each image patch is questionable because different image patches have different image contents and image content influences the quality. The second type of methods [3] uses full-reference IQA metrics to compute the quality score which serves as the MOS of an image. The drawback of this type of methods is that the performance of the no-reference IQA models directly depends on the performance of the full-reference IQA metrics used.

In this paper, we present a no-reference IQA model based on learning to rank method using Siamese Convolutional Neural Networks (LRSN) for both 2D and 3D images. In order to solve the problem of lacking training samples, we propose to transform the IQA task into a task of quality comparison between image patches. By generating pairs of image patches, the amount of training data reaches the square of the original amount of data. Furthermore, the pairwise comparison is a very effective way to obtain image quality scores. It is more in line with the perception of image quality of the HVS to compare and rank the distorted image with the reference image or two distorted images [8].

To solve the problem of quality difference between different image patches caused by varying image contents, we first sort the image patches according to their standard deviations. Subsequently, a number of image patches ranked in the middle are selected as the representative image patches for training, thereby reducing the influence of an uneven image quality distribution. We use Siamese CNN (SCNN) to achieve learning to rank for no-reference IQA. The proposed SCNN consists of two branches that have the same structure, share weights with each other, and take two image patches as inputs. Because the two branches share weights, the parameter size of the proposed SCNN is reduced to be comparable to a single-branch network.

We evaluate the proposed LRSN model on three 2D IQA databases and three 3D IQA databases. The experimental results show that the proposed LRSN model outperforms state-of-the-art 2D and 3D no-reference IQA metrics. In particular, the performance on the TID2013 database and the LIVE 3D Phase-II database is significantly improved.

The main contributions of this paper are as follows. 1) We propose to transform the IQA task into a task of quality comparison between image patches. The generated image patch pairs effectively solve the problem of lacking training samples. 2) We propose a new no-reference IQA model (LRSN) for no-reference IQA based on Siamese CNN. The presented Siamese CNN has two branches that have the same structure, share weights, and take two image patches as inputs. 3) The proposed LRSN model works well for both 2D and 3D no-reference IQA. Experimental results show that the proposed LRSC model achieves superior performance to the state-of-the-art 2D and 3D no-reference IQA models.

The rest of the paper is organized as follows. Section II provides related works on objective no-reference 2D and 3D IQA metric. In Section III, we present detailed description of the proposed LRSN model. In Section IV, the experimental results are presented. Section V concludes the paper.

II. RELATED WORKS

Unlike full-reference IQA and reduced-reference IQA, no-reference IQA does not use the information from the reference image. Therefore, no-reference IQA has more practical prospects than full-reference IQA and reduced-reference IQA in real applications without reference images. No-reference IQA has become a topic of intense research in the past decade. We describe related works on 2D and 3D IQA metrics in subsections II-A and II-B, respectively.

A. 2D IMAGE QUALITY ASSESSMENT

Many natural scene statistic based no-reference IQA metrics [9]–[12] are proposed to assess the quality of images distorted by various distortion types. Mittal *et al.* [9] proposed a natural image quality evaluator (NIQE), which is presented based on the construction of a quality aware collection of statistical features. Zhang *et al.* [10] proposed integrated local NIQE (ILNIQE) as an improvement of NIQE by using five types of natural scene statistic features to learn a multivariate Gaussian model of pristine images. The blind image spatial quality evaluator (BRISQUE) [11] extracts natural scene statistic features from a statistical model of locally normalized luminance coefficients in the spatial domain and demonstrates that these features correlate well with human judgments of quality. The blind image integrity notator using discrete cosine transform statistics (BLIINDS-II) [12] is a fast single-stage framework that relies on a statistical model of local discrete cosine transform coefficients. Discrete cosine transform features are extracted from the natural scene statistic model and thereafter fed to a Bayesian probabilistic inference model to evaluate image quality.

Some machine-learning-based no-reference IQA models have also been presented [13], [14]. Xue *et al.* [13] proposed a quality-aware clustering (QAC) based no-reference IQA model, which can learn a set of quality-aware centroids and estimate the quality level of image patches as a codebook. Ye *et al.* [14] proposed a blind learning model of image quality using synthetic scores (BLISS) which combines multiple full-reference measures into a single synthetic score.

In recent years, CNN has been applied to IQA. Kang *et al.* [1] applied a shallow CNN to no-reference IQA, which achieved an improved performance than the previous no-reference IQA models based on hand-crafted features. Zeng *et al.* [15] used a pretrained ResNet to extract features and fine-tune the network to learn the probability representation of a distorted image instead of its IQA score. Bosse *et al.* [2] proposed a no-reference IQA model based on deep CNN. In addition, they also adjusted the network to handle the full-reference IQA task. Kim and Lee [3] pretrained a model using the local score of a full-reference IQA index as ground truth, and subsequently fine-tuned the model using MOSs, whose performance depends on the performance of the selected full-reference IQA index. Ma *et al.* [16] proposed a quality-discriminable image pair inferred quality (dipIQ) index that uses a large number of image pairs for training. The premise of this model is that the distortion type and level of each distorted image are known. However, in practical applications without a reference image, the distortion type and level are unknown and sometimes are difficult to compute accurately. Ma *et al.* [17] proposed a multi-task end-to-end optimized deep neural network (MEON) for no-reference IQA. The training of MEON includes two steps: a distortion-type identification sub-network is first trained and thereafter a quality prediction sub-network is trained starting with the pre-trained early layers and the outputs of the first sub-network.

B. 3D IMAGE QUALITY ASSESSMENT

Compared with objective 2D IQA, 3D IQA is more complicated, and it is necessary to consider the influence of distortions on left view, right view, and left and right eye parallax. Many binocular perception-based metrics have been proposed to improve the performance of no-reference 3D IQA metrics by incorporating binocular perception. Zhou *et al.* [18] proposed two binocular combinations of stimuli, which were generated by the eye-weighting model and the contrast-gain control model, and then used the extreme learning machine for quality prediction. Shao *et al.* [19] proposed a framework for 3D no-reference IQA using joint sparse representation. The feature-prior and feature-distribution are combined to formulate a stereoscopic 3D quality prediction. Shao *et al.* [20] proposed a 3D no-reference IQA method, which transfers the information from the source feature domain to its target quality domain by dictionary learning.

Some depth perception-based metrics assess the image quality based on the disparity map or synthesized cyclopean image. Chen *et al.* [21] proposed a 3D no-reference IQA model that combines 2D and 3D features extracted from a stereoscopic image to estimate the perceptual quality. Jiang *et al.* [22] proposed an index based on deep non-negativity constrained sparse autoencoder with the input of the cyclopean image, left, and right views.

Some 3D IQA methods based on difference perception use the difference between the left and right views to assess image quality. Shen *et al.* [23] proposed combining the spatial frequency information and statistic feature extracted from the cyclopean and difference map to represent the binocular characteristic and asymmetric information of a stereoscopic image. Zhang *et al.* [4] proposed a CNN-based 3D no-reference IQA model, which considers the difference image as the representation of the depth and distortion in a stereoscopic image.

III. PROPOSED METHOD

In this section, we describe the proposed LRSN model that based on preference learning using SCNN. The framework of the proposed model is shown in Fig. 1. Firstly, we give the process of generating 2D and 3D image pairs in subsection III-A. Then, we describe the structure of SCNN proposed in this paper. Finally, we give the image quality score prediction method.

A. IMAGE PAIR GENERATION

1) IMAGE PREPARATION

For a 2D distorted image, inspired by the paper [11], the proposed LRSN model first performs local contrast normalization on the image. Local contrast normalization not only alleviates the saturation problem usually caused by using sigmoid neurons in CNN, but also makes the network more robust to changes in brightness and contrast. Given intensity image $I(i, j)$, the formulas for calculating the normalized

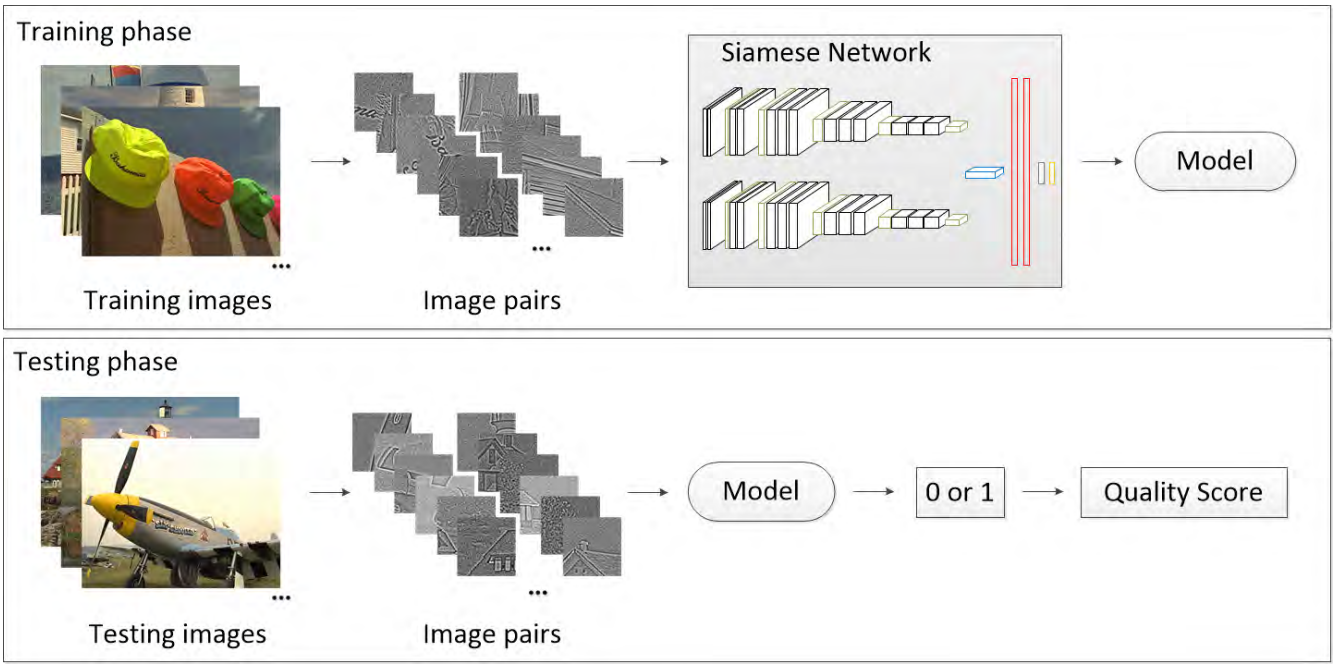


FIGURE 1. Framework of the proposed LRSN model.

value $\hat{I}(i,j)$ are as follows:

$$\hat{I} = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C}, \tag{1}$$

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} I_{k,l}(i,j), \tag{2}$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} (I_{k,l}(i,j) - \mu(i,j))^2}, \tag{3}$$

where C is a constant used to prevent the denominator from being zero, K and L are the width and height of the normalization window. The paper [11] proves that using a smaller normalized window can achieve good performance, so we set $K = L = 3$. $\omega_{k,l}$ is a 2D circularly-symmetric Gaussian weighting function sampled out to 3 standard deviations and rescaled to a unit volume.

For a stereoscopic 3D image which consists of two monocular 2D images, namely left view I_l and right view I_r, we use the difference image presented by Zhang *et al.* [4] as the representation of both two views and depth information. Specifically, we first conduct local contrast normalization on the left and right views and obtain normalized left and right views $\hat{I}_l$, $\hat{I}_r$, and then compute the difference image $\hat{I}_d$ as follows:

$$\hat{I}_d = \hat{I}_l - \hat{I}_r. \tag{4}$$

Fig. 2 shows an example of image preparation for a 3D image. After normalization, some pixels in the normalized left and right views and the difference image have positive values, and others have non-positive values. For illustration, we further normalize the pixel values in the normalized left and right views and the difference image to the range of [0, 1] and show the results in Fig. 2 (e)-(f). Take the normalized left image, $\hat{I}_l$ for example, we use the min-max normalization method to

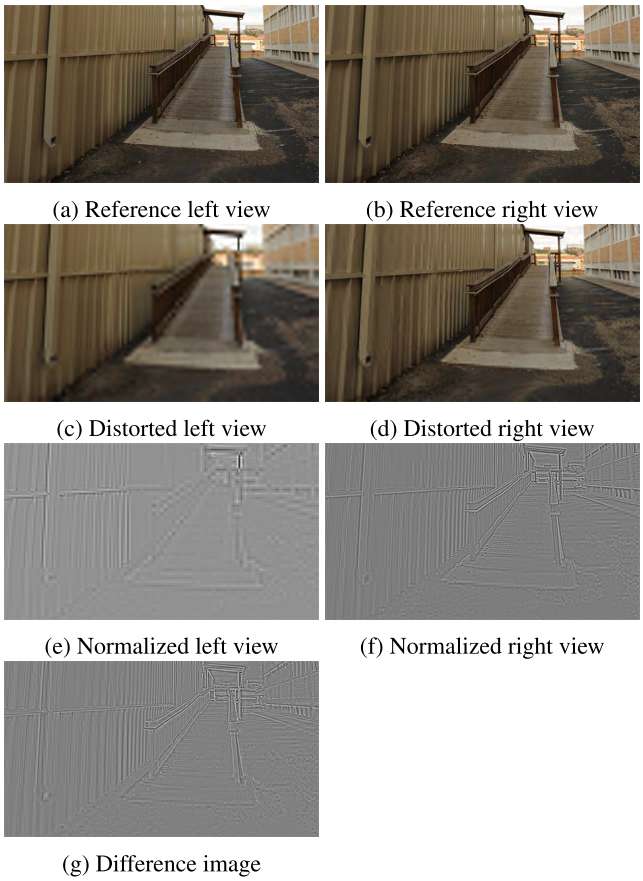


FIGURE 2. Example of image preparation for a 3D image. For illustration, figures (e)-(g) are normalized using the min-max normalization method.

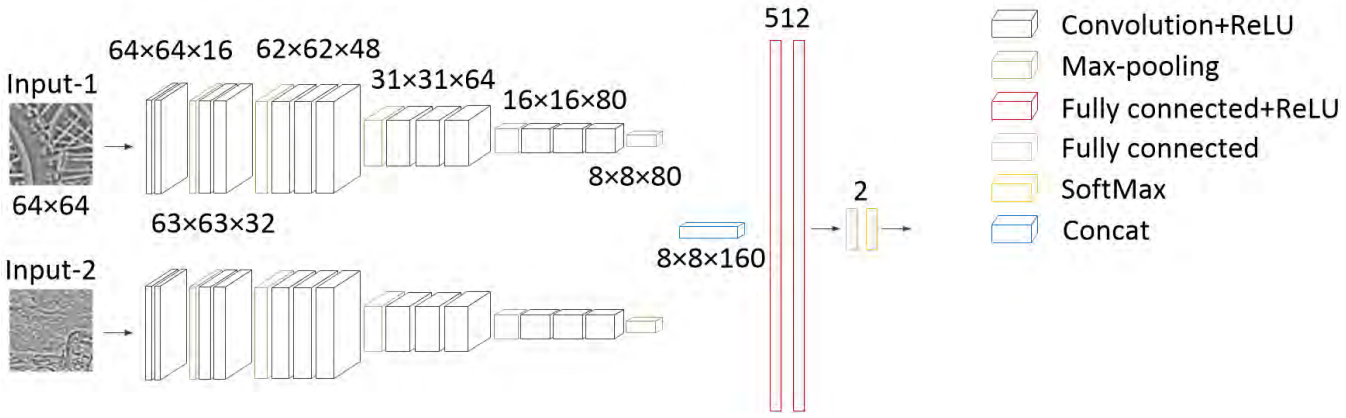


FIGURE 3. Architecture of the proposed LRSN model. The SCNN consists of two parallel branches, which have the same structure and share the weights.

normalize its pixel values to the range of [0,1] as follows,

$$\hat{I}'_l = \frac{\hat{I}_l - min}{max - min}, \tag{5}$$

where *min* and *max* represent the minimum and maximum values in the normalized left image, $\hat{I}_l$. It is worth mentioning that the min-max normalization (Equation (5)) is performed only for illustration, and not for the image preparation of the proposed LRSN model. Fig. 2 (e) and (f) are the normalized left and right views of the distorted 3D image. Fig. 2 (g) is the difference image between Fig. 2 (e) and (f).

2) IMAGE PATCH PAIRS

After image preparation, we divide each result image into equally sized image patches. Then we sort all image patches using the standard deviation of each image patch in ascending order and take a certain number (n) of image patches in the middle of the order as training data for the image.

There are mainly two reasons for taking image patches in the middle of the order. 1) It is difficult to use a small number of image patches to represent the entire image. However, if the number of image patches is large, it will inevitably lead to a large amount of training data. For efficiency, we select some image patches instead of using all image patches. 2) The quality scores of different image patches may be different. Therefore, we take the image patches in the middle of the order to reduce the impact of different image patches. In subsection IV-D, we show the effects of selecting image patches in different positions of the order, including image patches with the smallest, medium, and largest standard deviations.

Finally, the image patches selected from all the training images are combined into pairs to generate training data. In this way, the number of training data is increased to the square of the original number, which can overcome the shortage of training data for CNN-based IQA models.

We also compute a label for each pair of image patches. Because images in existing IQA databases have homogeneous distortions, we use the subjective quality score for each image as the quality scores for the patches obtained from it as in [1], [2]. For an IQA database, the quality score of an image is provided in the form of mean opinion score (MOS) or difference mean opinion score (DMOS) which is obtained through subjective evaluations. A higher MOS and a lower DMOS indicate a higher quality score. We compare the quality of two image patches by comparing their quality scores as follow,

$$L_{A,B} = \{MOS_A > MOS_B ? 1 : 0\}, \tag{6}$$

or

$$L_{A,B} = \{DMOS_A < DMOS_B ? 1 : 0\}, \tag{7}$$

where, A and B comprise an image patch pair $\langle A, B\rangle$, $L_{A,B}$ is the label of pair $\langle A, B\rangle$, MOS_A (MOS_B) and $DMOS_A$ ($DMOS_B$) are the quality scores of patch A (B) in the form of MOS and DMOS, respectively. If $MOS_A > MOS_B$ or $DMOS_A < DMOS_B$, the label of the image pair $\langle A, B\rangle$ is 1, otherwise it is 0.

To achieve high efficiency and effectiveness of the proposed LRSN, we compose image patch pairs using the following principles: 1) The image patch is not combined with the image patch from the same image; 2) If image patch A and image patch B composed an image patch pair, then B is no longer combined with A, thereby avoiding data redundancy; 3) Due to the small difference in quality between image patches with similar quality scores, composing image patch pairs with similar quality scores will increase the difficulty of preference learning. Therefore, if the difference in the quality scores between the two image patches smaller than a certain threshold, the image pair is not composed.

B. SIAMESE CONVOLUTIONAL NEURAL NETWORKS

In this paper, we propose a new SCNN whose structure is shown in Fig. 3. The SCNN consists of two parts: subnetworks I and II. Subnetwork I consists of two identical branches using five stacked convolutional structures for image feature extraction. Specifically, the structure of one of

the branches is composed of 13 convolution layers and 5 pooling layers (conv1-1, conv1-2, maxpool1, conv2-1, conv2-2, maxpool2, conv3-1, conv3-2, conv3-3, maxpool3, conv4-1, conv4-2, conv4-3, maxpool4, conv5-1, conv5-2, conv5-3, and maxpool5). The rectified linear units (ReLU) is used as an activation function for each convolution layer. Zero padding is used for all convolution layers. The convolution kernel size is 3×3, and the stride is one. All max-pool layers have 2×2 pixel-sized kernels, and strides in both directions are two.

Subnetwork II consists of three fully connected layers. The specific components are: FC6, FC7, and FC8. FC6 and FC7 both use ReLU as the activation function. In order to prevent overfitting, the dropout ratio is set to be 0.5. The features extracted by sub-network I are fused and used as input to the sub-network II. Sub-network II distinguishes the quality of the two input image patches according to the fused features.

The network uses cross entropy as loss function, and its formula is as follows:

$$L = -\frac{1}{N}\sum_{i=1}^{N}(y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})), \quad (8)$$

where N represents the number of image patch pairs; $y^{(i)} = [y_1^{(i)}, y_2^{(i)}]$ is a two-dimensional vector used to indicate the quality of two images. For the i^{th} image patch pair $\langle A, B\rangle$, if the quality of image patch A is greater than that of B, then $y^{(i)} = [1, 0]$, otherwise $y^{(i)} = [0, 1]$. $\hat{y}^{(i)} = [\hat{y}_1^{(i)}, \hat{y}_2^{(i)}]$ is also a two-dimensional vector, indicating the probability that the first image is better than the second. Conversely, the probability that the second image is better than the first image is $1 - \hat{y}^{(i)}$.

SCNN is iteratively trained over multiple epochs, and an epoch is defined as traversing the entire training set. In each epoch, the training set is divided into multiple mini-batches for batch optimization. We set the size of a mini-batch to 128. The initial learning rate is set to be 1×10^{-4}, and the batch-optimized learning rate of each parameter is adaptively controlled by the Adam method based on gradient variance. The parameters of Adam are referenced in the paper [24], $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$, and $\alpha = 10^{-4}$.

C. IMAGE QUALITY SCORE PREDICTION

Most of the traditional IQA models directly predict the quality score of the image by regression. In this paper, the relative quality score of the image is obtained by first comparing its image patches with many image patches of other images and counts the number of times that its image patches are ranked superior to other patches.

Similar to the training phase, in the prediction phase, we first perform image preparation. Then we divide the image into equal-sized image patches, sort all image patches by standard deviation, and select a number of image patches ranked in the middle. Finally, we generate image pairs for comparison. To ensure that the final score for each image is in the same range, we compare each image patch to all image patches in the test set, except for the image patch of its own image. The formula for predicting quality score for image I is as follows:

$$S_I = \sum_{A\in I}\sum_{B\notin I} P_{A,B}, \quad (9)$$

where image patch A belongs to image I, image patch B does not belong to image I, $P_{A,B}$ denotes the result of comparison between image patches A with B. $P_{A,B} = 1$ represents that the quality of image patch A is better than B. Otherwise the quality of image patch B is better than A. S_I represents the score of the image I.

In practical applications, there may be only one test image or the quality concentration of the test set. If there is only one image in the test set, there will be no comparable image. If the quality of the image in the test set is too centralized, the quality score of the image cannot be given accurately. Therefore, we also provide a test image set. When testing, the test image is compared to all the images in the test image set we provide.

IV. EXPERMENTS

In this section, we give the experimental protocol in subsection IV-A, show the performance comparison for 2D and 3D IQA in subsection IV-B, validate the generalization ability of the proposed model via cross-database tests in subsection IV-C, and discuss the effects of different strategies of image patch selection in subsection IV-D.

A. EXPERIMENTAL PROTOCOL

We use three 2D IQA databases to verify the performance of the proposed LRSN model for 2D IQA task.

1) LIVE [25]: The LIVE database consists of 29 reference images and 779 distorted images, each impaired by one of four or five levels of five types of synthetic distortions: JPEG compression (JPEG), JPEG2000 compression (JP2K), White noise (WN), BLUR and Fast-Fading (FF). Each distortion type contains 7 or 8 distortion levels. The subjective values for the distorted images are given by DMOS. A lower DMOS indicates higher visual quality.

2) CSIQ [26]: The CSIQ database consists of 30 reference images and 800 distorted images. Among them, 800 distortion images are generated from 30 reference images through six distortion types, each with 4 or 5 distortion levels. Its distortion types include JPEG, JP2K, WN, BLUR, FNOISE and CONTRAST. The subjective values for the distorted images are given by DMOS.

3) TID2013 [27]: The TID2013 database consists of 25 reference images and 3000 distortion images. The 3000 distortion images are transformed from 25 reference images through 24 distortion types and 5 distortion levels. The subjective values for the distorted images are given by MOS. A higher MOS indicates higher visual quality.

Besides, we use three 3D IQA databases to verify the performance of the proposed LRSN model for 3D IQA task.

1) LIVE 3D Phase-I [28]: The LIVE 3D Phase-I database consists of 20 reference stereoscopic images and 365 distorted stereoscopic images. This database includes five distortion types: JPEG, JP2K, WN, BLUR, and FF. Each image in the database is symmetrically distorted on its left and right views. Each distorted stereoscopic image is given a DMOS from subjective evaluation.

2) LIVE 3D Phase-II [29]: The LIVE 3D Phase-II database consists of 8 reference images and 360 symmetrically or asymmetrically distorted stereoscopic images. An asymmetrically distorted image has different distortion levels in its left and right views. This database includes five distortion types: JPEG, JP2K, WN, BLUR, and FF. For each distortion type, a reference image pair generates three symmetrically distorted images and six asymmetrically distorted images. The subjective evaluation values in the database are given in the form of DMOS.

3) NBU 3D IQA Database [30]: The NBU 3D IQA database consists of 12 reference images and 312 symmetrically distorted images. This database includes five distortion types: JPEG, JP2K, WN, BLUR, and H.264. The subjective values for the distorted images are given by DMOS.

We use Spearman Rank Order Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient (PLCC) to measure the performance of the IQA models. In our experiment, we randomly selected the distorted image corresponding to 80% of the reference images as the training set, and the remaining 20% of the data was used as the test set. All the results of the multi-distortion-type experiment were obtained by taking the median of 20 repeated experiments, and the results of the single-distortion-type experiment were obtained by taking the median of 10 repeated experiments.

The number of image patches selected for a single image is 8 ($n = 8$) in the four distortion-type experiments on all 2D IQA databases and all distortion-type experiments on all 3D IQA databases. Since the amount of data in every single distortion-type experiment is small, we set $n = 16$ for every single distortion-type experiment on 2D IQA database. The size of each image patch is 64×64.

Because composing image patch pairs with similar quality scores will increase the difficulty of preference learning, the image pair is not composed if the difference in the quality scores between the two image patches smaller than a certain threshold. In the experiments, we set the threshold based on the range of values of the image scores in each dataset. The MOS/DMOS ranges of values for three 2D IQA databases (LIVE, CSIQ, and TID2013) are $0-100$, $0-1$, and $0-10$, respectively, so we set their thresholds to be 4, 0.03, and 0.2, respectively. The DMOS ranges of values for three 3D IQA databases (LIVE 3D Phase-I, LIVE 3D Phase-II, and NBU) are all $0-100$, so we set their thresholds to be 4.

B. PERFORMANCE COMPARISON

1) PERFORMANCE FOR 2D IQA

In order to validate the performance of the proposed LRSN model on 2D IQA task, we compared it with eight no-reference IQA models, including QAC [13], NIQE [9], ILNIQE [10], BLIINDS−II [12], BRISQUE [11], BLISS [14], dipIQ [16], and MEON [17]. Following previous no-reference IQA metrics [1]–[3], we also compared the proposed LRSN model with four typical full-reference IQA metrics, including peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) [31], an index based on perceptual similarity measure (PSIM) [32], and an index based on distortion distribution-based gradient similarity (ADD-GSIM) [33]. Full-reference IQA uses the original image as a reference, and it usually achieves a higher consistency with human perception than no-reference IQA. We compare the proposed LRSN model with full-reference IQA to show the performance difference between no-reference IQA and full-reference IQA. A small difference indicates a good no-reference IQA index. Following works [14], [16], [17], in order to verify the generalization ability of the method, we experimented on four distortion types common to the three 2D IQA databases, including JPEG, JP2K, WN, and BLUR.

TABLE 1. Median SROCC and PLCC results on LIVE.

SROCC	JP2K	JPEG	WN	BLUR	ALL4
PSNR	0.908	0.894	0.984	0.814	0.883
SSIM [31]	0.961	0.974	0.970	**0.952**	0.947
PSIM [32]	**0.987**	**0.985**	**0.992**	**0.979**	**0.984**
ADD-GSIM [33]	**0.985**	**0.984**	**0.988**	**0.979**	**0.986**
QAC [13]	0.876	0.951	0.925	0.911	0.869
NIQE [9]	0.924	0.945	0.972	0.941	0.920
ILNIQE [10]	0.901	0.944	**0.979**	0.927	0.918
BLIINDS-II [12]	0.932	0.933	0.946	0.891	0.931
BRISQUE [11]	0.914	0.965	**0.979**	**0.951**	0.942
BLISS [14]	0.925	0.956	0.967	0.936	0.945
dipIQ [16]	**0.956**	**0.969**	0.975	0.940	**0.958**
MEON [17]	-	-	-	-	-
LRSN	**0.967**	**0.983**	**0.988**	**0.966**	**0.974**
PLCC	JP2K	JPEG	WN	BLUR	ALL4
PSNR	0.912	0.896	0.987	0.812	0.874
SSIM [31]	**0.968**	0.980	0.972	**0.951**	0.937
PSIM [32]	**0.984**	**0.990**	**0.992**	**0.971**	**0.975**
ADD-GSIM [33]	**0.984**	**0.987**	**0.990**	**0.971**	**0.981**
QAC [13]	0.876	0.960	0.895	0.912	0.855
NIQE [17]	0.932	0.956	0.979	**0.951**	0.912
ILNIQE [10]	0.912	0.966	0.976	0.936	0.913
BLIINDS-II [12]	0.939	0.943	0.964	0.899	0.929
BRISQUE [11]	0.923	0.973	**0.985**	**0.951**	0.949
BLISS [14]	0.933	0.972	0.978	0.948	0.945
dipIQ [16]	**0.964**	**0.980**	0.983	0.948	**0.957**
MEON [17]	-	-	-	-	-
LRSN	**0.966**	**0.992**	**0.988**	**0.966**	**0.978**

Experimental results on LIVE, CSIQ and TID2013 databases are shown in Table 1, Table 2, and Table 3, respectively. We show the top two performance values achieved by full-reference IQA or no-reference IQA models in bold and underlined the best performance values. A symbol "-"

TABLE 2. Median SROCC and PLCC results on CSIQ.

SROCC	JP2K	JPEG	WN	BLUR	ALL4
PSNR	0.941	0.901	0.943	0.936	0.928
SSIM [31]	0.962	**0.956**	0.912	0.965	0.935
PSIM [32]	**0.975**	**0.967**	**0.977**	**0.968**	**0.967**
ADD-GSIM [33]	**0.977**	**0.967**	**0.968**	**0.976**	**0.971**
QAC [13]	0.884	0.913	0.850	0.839	0.840
NIQE [9]	0.926	0.882	0.836	0.908	0.883
ILNIQE [10]	0.924	0.905	0.867	0.867	0.887
BLIINDS-II [12]	0.890	0.904	0.909	0.866	0.892
BRISQUE [11]	0.894	0.916	0.934	0.915	0.909
BLISS [14]	0.932	0.927	0.879	0.922	0.920
dipIQ [16]	**0.944**	0.936	0.904	**0.932**	0.930
MEON [17]	0.898	**0.948**	**0.951**	0.918	**0.932**
LRSN	**0.965**	**0.979**	**0.978**	**0.945**	**0.935**
PLCC	JP2K	JPEG	WN	BLUR	ALL4
PSNR	0.954	0.908	0.961	0.937	0.918
SSIM [31]	**0.973**	0.983	0.908	**0.956**	0.930
PSIM [32]	**0.982**	**0.985**	**0.977**	**0.974**	**0.974**
ADD-GSIM [33]	**0.982**	**0.984**	**0.966**	**0.974**	**0.975**
QAC [13]	0.898	0.942	0.865	0.855	0.847
NIQE [9]	0.944	0.946	0.824	0.935	0.900
ILNIQE [10]	0.942	0.956	0.880	0.903	0.914
BLIINDS-II [12]	0.905	0.927	0.931	0.893	0.922
BRISQUE [11]	0.937	0.960	0.947	0.936	0.937
BLISS [14]	0.954	0.970	0.895	0.947	0.939
dipIQ [16]	**0.959**	0.975	0.927	**0.958**	**0.949**
MEON [17]	0.925	**0.979**	**0.958**	0.906	0.891
LRSN	**0.975**	**0.990**	**0.985**	**0.950**	**0.954**

TABLE 3. Median SROCC and PLCC results on TID2013.

SROCC	JP2K	JPEG	WN	BLUR	ALL4
PSNR	0.898	0.929	**0.942**	**0.965**	0.924
SSIM [31]	0.950	0.935	0.896	**0.969**	0.924
PSIM [32]	**0.971**	**0.951**	**0.942**	0.923	**0.955**
ADD-GSIM [33]	**0.965**	**0.944**	**0.936**	0.935	**0.954**
QAC [13]	0.883	0.885	0.668	0.879	0.837
NIQE [9]	0.901	0.873	0.854	0.821	0.812
ILNIQE [10]	0.912	0.873	0.890	0.815	0.881
BLIINDS-II [12]	0.906	0.915	0.892	0.676	0.858
BRISQUE [11]	0.906	0.894	0.889	0.886	0.883
BLISS [14]	0.906	0.893	0.856	0.872	0.836
dipIQ [16]	**0.926**	**0.932**	0.905	**0.922**	0.877
MEON [17]	0.911	0.919	**0.908**	0.891	**0.912**
LRSN	**0.927**	**0.945**	**0.942**	**0.931**	**0.946**
PLCC	JP2K	JPEG	WN	BLUR	ALL4
PSNR	0.933	0.925	**0.963**	**0.958**	0.911
SSIM [31]	0.970	0.968	0.902	**0.958**	0.927
PSIM [32]	**0.983**	**0.980**	**0.953**	0.919	**0.960**
ADD-GSIM [33]	**0.978**	**0.974**	0.940	**0.923**	**0.957**
QAC [13]	0.892	0.929	0.719	0.877	0.829
NIQE [9]	0.912	0.928	0.859	0.848	0.819
ILNIQE [10]	0.929	0.944	0.899	0.816	0.890
BLIINDS-II [12]	0.921	0.931	0.908	0.695	0.896
BRISQUE [11]	0.919	0.950	0.886	0.884	0.900
BLISS [14]	0.930	0.963	0.863	0.872	0.862
dipIQ [16]	**0.948**	**0.973**	0.906	**0.928**	0.894
MEON [17]	0.924	**0.969**	**0.911**	0.899	**0.912**
LRSN	**0.951**	0.956	**0.944**	**0.931**	**0.934**

indicates that the corresponding experimental result was not provided in the original paper and the corresponding source code was not found for reproducing the result.

Tables 1, 2, and 3 show the results on LIVE, CSIQ, and TID2013, respectively. From the three tables, we can get the following conclusions: 1) The performance of the proposed LRSN model is superior to six state-of-the-art no-reference IQA models on the three databases, and even exceeds two classic full-reference IQA metrics. 2) For the single-distortion-type experiments on the LIVE Database, the performance of the proposed LRSN model surpasses six no-reference IQA models and two classic full-reference IQA models in the JP2K, JPEG, WN, and BLUR distortion types. Moreover, the performance on both JPEG and WN distortion types is close to the other two state-of-the-art full-reference IQA models. 3) For the single-distortion-type experiments on the CSIQ Database, the proposed LRSN model is superior to 6 no-reference IQA models in all four distortion types, and exceeds four full-reference IQA models in JPEG and WN distortion types. 4) For the single-distortion-type experiments on the TID2013 Database, the proposed LRSN model achieves improved performance compared to the 6 comparison no-reference IQA models, and the improvement in WN distortion type is significant. For WN distortion type, the performance of the proposed LRSN model is close to the best performance among the four full-reference IQA metrics.

2) PERFORMANCE FOR 3D IQA

In order to validate the performance of the proposed LRSN model on the 3D IQA task, we compared it with nine no-reference IQA models, including BRISQUE [11], BLIINDS-II [12], Shao *et al.* [19], Chen *et al.* [21], Zhou *et al.* [18], Jiang *et al.* [22], Shao *et al.* [20], Zhang *et al.* [4], and Shen *et al.* [23]. Among these nine no-reference IQA metrics, BRISQUE [11] and BLIINDS-II [12] were initially presented for 2D IQA. The quality scores of these 2D IQA metrics were obtained by calculating the average of the quality scores of the left view and right view. Zhou *et al.* [18] and Shao *et al.* [19], [20] are 3D no-reference IQA metrics based on binocular perception. Chen *et al.* [13] and Jiang *et al.* [22] are 3D no-reference IQA metrics based on depth perception. Shen *et al.* [23] and Zhang *et al.* [4] are 3D no-reference IQA metrics based on difference perception.

To compare the performance difference between the proposed LRSN model and the full-reference IQA, we also compared the proposed LRSN model with nine full-reference IQA models, including SSIM [31], FSIM [34], gradient magnitude similarity deviation (GMSD) [35], DCT subbands similarity (DSS) [36], Benoit *et al.* [37], Bensalma and Larabi [38], Chen *et al.* [29], Wang *et al.* [39], and Shao *et al.* [40]. Among these nine full-reference IQA metrics, SSIM [31], FSIM [34], GMSD [35], and DSS [36] were initially presented for 2D IQA. The quality scores of these 2D IQA metrics were computed in the same way as BRISQUE [11] and BLIINDS-II [12]. Benoit *et al.* [37] and Chen *et al.* [29] are 3D full-reference IQA metrics that combine depth/disparity and 2D IQA metrics. Bensalma and Larabi [38], Wang *et al.* [39], and Shao *et al.* [40] are binocular characteristics-based 3D full-reference IQA metrics.

TABLE 4. Experimental results on LIVE 3D IQA Phase-I, LIVE 3D IQA Phase-II, and NBU 3D IQA databases.

Type		LIVE 3D Phase-I		LIVE 3D Phase-II		NBU 3D IQA	
		SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
FR	SSIM [31]	0.877	0.873	0.792	0.803	0.909	0.914
	FSIM [34]	0.921	0.927	0.774	0.794	0.931	0.927
	GMSD [35]	**0.935**	**0.943**	0.763	0.793	**0.941**	0.939
	DSS [36]	**0.943**	**0.939**	0.745	0.776	**0.942**	**0.944**
	Benoit [37]	0.889	0.903	0.744	0.762	0.881	0.876
	Bensalma [38]	0.875	0.887	0.751	0.770	0.938	0.937
	Chen [29]	0.916	0.917	**0.889**	**0.900**	0.909	0.908
	Wang [39]	0.924	0.929	**0.918**	**0.915**	-	-
	Shao [40]	0.925	0.935	0.849	0.863	**0.941**	**0.941**
NR	BRISQUE [11]	0.901	0.910	0.770	0.770	0.920	0.917
	BLIINDS-II [12]	0.910	0.917	0.701	0.737	0.915	0.921
	Shao [19]	0.867	0.885	0.872	0.909	-	-
	Chen [21]	0.891	0.895	0.880	0.895	-	-
	Zhou [18]	0.921	0.941	**0.919**	**0.923**	-	-
	Jiang [22]	0.912	0.930	0.915	0.922	0.931	0.936
	Shao [20]	**0.944**	**0.953**	0.885	0.903	**0.938**	**0.949**
	Zhang [4]	**0.943**	0.947	-	-	-	-
	Shen [23]	-	-	**0.919**	0.919	-	-
	LRSN	0.939	**0.957**	**0.920**	**0.929**	**0.952**	**0.940**

Experimental results on LIVE 3D Phase-I, LIVE 3D Phase-II, and NBU 3D IQA databases are shown in Table 4. We show the top two performance values achieved by full-reference or no-reference IQA models in bold and underlined the best performance values. The symbol "-" indicates that the corresponding experimental result was not provided in the original paper and the corresponding source code was not found for reproducing the result.

From Table 4, we can get the following conclusions: 1) The proposed LRSN model outperforms the state-of-the-art full-reference and no-reference IQA metrics for 3D IQA. 2) For the experiments on the LIVE 3D Phase-I database, the proposed LRSN model achieves the best PLCC value and the third best SROCC value among all no-reference IQA models. 3) For the experiments on the LIVE 3D Phase-II database, the proposed LRSN model achieves the best PLCC and SROCC values among all no-reference IQA models. It worth mentioning that, the proposed LRSN model also outperforms all compared full-reference IQA models on the LIVE 3D Phase-II database which has both symmetrically and asymmetrically distorted 3D images. 4) For the experiments on the NBU 3D IQA database, the proposed LRSN model achieves the best SROCC value and the second best PLCC value among all no-reference IQA models.

In summary, the experimental results show that the proposed LRSN model can achieve good performance on both 2D and 3D IQA databases, which verify the effectiveness of the proposed LRSN model for both 2D and 3D IQA. Furthermore, the proposed LRSN model achieves significant improvements on the TID2013 database and the LIVE 3D Phase-II database.

C. CROSS-DATABASE TEST

To evaluate the generalization ability of the proposed model, we trained the proposed LRSN model on LIVE database and tested it on the CSIQ and TID2013 databases for 2D IQA, and trained the proposed LRSN model on LIVE 3D Phase-I and tested it on LIVE 3D Phase-II or NBU databases for 3D IQA.

For 2D IQA test, following the protocol of previous works [3], [16], we trained and tested the proposed LRSN model using four common distortion types in the LIVE, CSIQ, and TID2013 databases: JPEG, JP2K, WN, and BLUR. We used all the images contained in the four distortion types in the LIVE database for training, and used all the images contained in the four distortion types in the CSIQ and TID2013 databases for testing. We compared the proposed LRSN model with four no-reference IQA models: BRISQUE [11], DIIVINE [41], CORNIA [42], and dipIQ [16]. The experimental results are shown in Table 5.

TABLE 5. Cross-dataset evaluation (SROCC). Models are trained on LIVE and tested on CSIQ and TID2013.

SROCC	CSIQ	TID2013	Database Size Weighted Average
BRISQUE	0.909	0.883	0.897
DIIVINE	0.835	0.795	0.817
CORNIA	0.915	**0.893**	0.905
dipIQ	**0.930**	0.877	**0.906**
LRSN	**0.919**	**0.913**	**0.916**

For 3D IQA test, we used all the images contained in the LIVE 3D Phase-I database for training, and used all the images contained in the LIVE 3D Phase-II and NBU databases for testing. It should be noted that the H.264 distortion type in the NBU database does not appear in the LIVE 3D Phase-I database, so the results on the NBU database is not as good as those on the LIVE 3D Phase-II database. We compared the proposed LRSN model with three no-reference IQA models: BRISQUE [11], BLIINDS-II [12], and Jiang *et al.* [22]. Among them, metrics BRISQUE [11] and BLIINDS-II [12] were presented for 2D images, and Jiang *et al.* [22] was a depth perception-based

TABLE 6. Cross-dataset evaluation (SROCC). Models are trained on LIVE 3D Phase-I, and tested on LIVE 3D Phase-II or NBU.

SROCC	LIVE 3D Phase-II	NBU	Database Size Weighted Average
BRISQUE	**0.796**	0.396	0.612
BLIINDS-II	0.731	0.708	0.722
Jiang	0.758	**0.732**	**0.746**
LRSN	**0.809**	**0.730**	**0.772**

3D no-reference IQA metric. The experimental results are shown in Table 6.

As can be seen from Table 5, the proposed LRSN model achieves the best and second best performance on TID2013 and CSIQ databases, respectively. In the last column of Table 5, we show the database size weighted average over CSIQ and TID2013 databases. The weight for each database is the number of test images in the database. The weighted average values demonstrate the superior performance of the proposed LRSN model for 2D IQA. As can be seen from Table 6, the proposed LRSN model achieves the best and second best performance on LIVE 3D Phase-II and NBU databases, respectively. The weighted average values demonstrate the superior performance of the proposed LRSN model for 3D IQA.

In summary, the results of these cross-database experiments show that the proposed LRSN model does not depend on a specific database and has good generalization ability for both 2D and 3D IQA tasks.

D. DISCUSSION

In this subsection, we mainly discuss the effects of different strategies of image patch selection, effects of testing strategies, and effects of visual attention on the performance of the proposed LRSN model.

1) EFFECTS OF DIFFERENT STRATEGIES OF IMAGE PATCH SELECTION

We experimented with image patches obtained by using different strategies of image patch selection. Specifically, the same number of image patches with the smallest, medium, and largest standard deviations for each image were used as the selected image patches in the experiments. In Fig. 4, we show three image patches with the largest (a), medium (b), and smallest (c) standard deviations of an example image. Image patches (a) and (c) belong to the foreground and background, respectively. While image patch (b) partially belong to the foreground and partially belong to the background so that it can represent the image content better. We also carried out the experiment of selecting the number of image patches per image on LIVE database. The experimental results are shown in Table 7 and Table 8.

As can be seen from Table 7, when selecting the image patches with the standard deviation ranked in the middle for training, the experimental results are better than those when selecting image patches with the largest and smallest

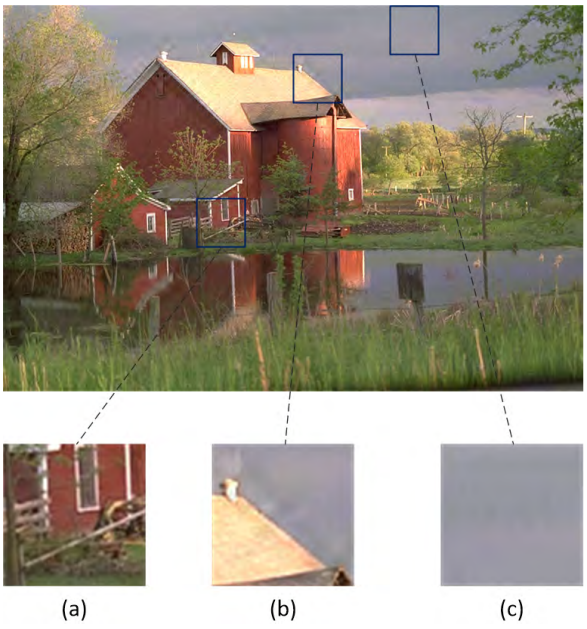


FIGURE 4. Three image patches with the largest (a), medium (b), and smallest (c) standard deviations of an example image.

TABLE 7. Experimental results of different values of the standard deviation of image patches on LIVE database.

Standard deviation	Small	Medium	Large
SROCC	0.948	0.974	0.959
PLCC	0.951	0.978	0.960

TABLE 8. Experimental results of selecting a different number of image patches for per image on the LIVE database.

Number of image patches	4	8	12	16
SROCC	0.932	0.974	0.976	0.974
PLCC	0.934	0.978	0.973	0.975

standard deviations. Therefore, selecting the image patches with the standard deviation ranked in the middle for training can reduce the impact of uneven image quality distribution and achieve a high consistency with subjective perception.

To further evaluate the effect of the number of image patches selected from each distorted image on the performance of the proposed LRSN model, we experimented with different numbers of image patches. Specifically, we experimented with four different numbers of image patches, namely 4, 8, 12, and 16. The experimental results are shown in Table 8.

From Table 8, we can see that a larger number of image patches selected for each image cannot guarantee a better performance of the proposed LRSN model. Specifically, the performance of selecting 4 image patches per image is inferior to the performance of selecting 8 image patches per image. However, when increasing the number from 8 to 12 and 16, the performance remains almost the same. Meanwhile, increasing the number of image patches will increase the training and testing time. Considering both performance and

efficiency, the number of image patches selected for each image is 8.

2) EFFECTS OF TESTING STRATEGIES

We experimented with different testing strategies on TID2013 database. During testing, we experimented with the different number of comparison for each test image to investigate the influence of the number of comparisons. Specifically, we compared each test image patch with a different number of image patches. It should be noted that there are 100 images in our test set and 8 image patches are selected from each image. Our original test strategy was to compare the patches of each image with the patches of all other 99 images. In the experiment, we compared each image with 99 images and the randomly selected 90, 80, 70, 60, 50, 40, 30, 20, and 10 images from the test set. The experiments were repeated three times. The experimental results are shown in Table 9.

TABLE 9. Experimental results of comparing each image with the different numbers of test images on TID2013 database. The best performance values for each of the three repeated experiments are formatted in bold.

No. of image (No. of image patches)	SROCC			PLCC		
	1	2	3	1	2	3
99 (792)	0.946	0.946	0.946	0.934	0.934	0.934
90 (720)	0.947	0.948	0.946	0.934	0.938	0.935
80 (640)	0.946	0.946	0.947	0.932	0.935	0.935
70 (560)	0.947	0.948	0.946	0.936	0.938	0.936
60 (480)	0.948	**0.950**	0.945	0.936	**0.941**	0.935
50 (400)	**0.949**	0.947	**0.949**	0.937	0.937	**0.939**
40 (320)	0.944	0.944	0.946	0.936	0.938	0.937
30 (240)	0.939	0.945	0.947	0.939	0.934	**0.939**
20 (160)	**0.949**	0.943	0.942	**0.941**	0.936	0.934
10 (80)	0.941	0.945	0.937	0.921	0.931	0.923

From Table 9, we can conclude that the number of compared images and image patches has little effect on the IQA results when using image patches no less than 160. Because comparing using randomly chosen images, the performance varies. In order to make the experimental results can be reproduced by others, we compared each test image with all other images in the test set in all experiments. In practice, good performance can be achieved when using image patches no less than 400.

Furthermore, we investigated the influence of different test sets of images. Specifically, we compared each test image in TID2013 test set with the same number of images in LIVE and CSIQ databases. The experimental results are shown in Table 10.

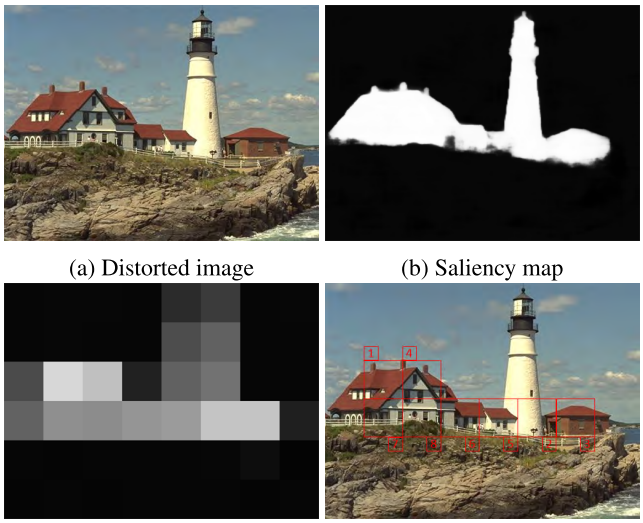
TABLE 10. Experimental results of TID2013 database when using the same number of images in TID2013, LIVE, or CSIQ databases as the test set.

	TID2013	CSIQ	LIVE
SROCC	0.946	0.942	0.945
PLCC	0.934	0.934	0.937

From Table 10, we can conclude that the change of the test set has little effect on the results. This also proves that the proposed LRSN model has good generalization ability.

3) EFFECTS OF VISUAL ATTENTION

Because visual attention has an important impact on IQA, we conducted experiments on TID2013 database to incorporate visual attention into IQA. Firstly, we use the method proposed by Hou *et al.* [43] to generate saliency maps of the distorted images, as shown in Fig. 5 (b). Then the average saliency values of all image patches (as shown in Fig. 5 (c)) of an image are computed and sorted in a non-ascending order. We experimented with three sets of image patches for training and testing, namely eight image patches with middle standard deviations, eight image patches with the largest average saliency values, four image patches with middle standard deviations and four image patches with the largest average saliency values. The experimental results are shown in Table 11.



(a) Distorted image (b) Saliency map

(c) Average saliency map (d) Selected image patches

FIGURE 5. Example of visual attention.

TABLE 11. Experimental results of incorporating visual attention into IQA on the TID2013 database.

Image patches	SROCC	PLCC
8 with middle standard deviations	0.946	0.934
8 with largest average saliency values	0.936	0.927
4 with middle standard deviations and 4 with largest average saliency values	0.946	0.926

From Table 11, we can conclude that using image patches with the largest visual attention for training and testing has limited influence on the overall performance of IQA. One reason is that images usually have salient objects of different sizes, therefore using the same number of image patches to represent the salient objects is not a good choice. Secondly, distortions usually have a negative influence on the

performance of salient object detection algorithms. Therefore, we will investigate more effective ways of incorporating visual attention into IQA in the future.

V. CONCLUSION

In this paper, we propose a 2D and 3D no-reference IQA model based on learning to rank, using SCNN for training. The SCNN uses two branches that share weights to extract features of the input two image patches. Then the relative quality of the two input image patches is compared. The relative quality score of the image is obtained by comparing image patches and counting times of preference. Extensive experiments on both 2D and 3D IQA databases show that the proposed LRSC model achieves superior performance to the state-of-the-art no-reference IQA models.

Existing researches evaluate the quality of a 3D image from an isolated perspective, such as distortion, aesthetics, or visual comfort. However, these perspectives should be all considered to get a comprehensive quality assessment for 3D images because they influence each other. To facilitate the research on objective comprehensive quality assessment for 3D images, we plan to carry out the subjective assessment using 3D images with different distortion types and levels, aesthetics values, and visual comfort values, and then construct a new database. We also plan to explore the SCNN-based preference learning for image restoration and enhancement [44].

REFERENCES

[1] 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. (CVPR)*, Jun. 2014, pp. 1733–1740.

[2] 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.

[3] 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.

[4] W. Zhang, L. Qu, L. Ma, J. Guan, and R. Huang, "Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network," *Pattern Recognit.*, vol. 59, pp. 176–187, Nov. 2016.

[5] Y. Niu, Y. Zhong, W. Guo, Y. Shi, and P. Chen, "2D and 3D image quality assessment: A survey of metrics and challenges," *IEEE Access*, vol. 7, pp. 782–801, 2019.

[6] Y. Fang, J. Yan, L. Li, J. Wu, and W. Lin, "No reference quality assessment for screen content images with both local and global feature representation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1600–1610, Apr. 2018.

[7] Y. Niu, H. Zhang, W. Guo, and R. Ji, "Image quality assessment for color correction based on color contrast similarity and color value difference," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 4, pp. 849–862, Apr. 2018.

[8] L. Xu, J. Li, W. Lin, Y. Zhang, Y. Zhang, and Y. Yan, "Pairwise comparison and rank learning for image quality assessment," *Displays*, vol. 44, pp. 21–26, Sep. 2016.

[9] 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. 2013.

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

[11] 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. 4595–4708, Dec. 2012.

[12] 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.

[13] 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.

[14] P. Ye, J. Kumar, and D. Doermann, "Beyond human opinion scores: Blind image quality assessment based on synthetic scores," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 4241–4248.

[15] H. Zeng, L. Zhang, and A. C. Bovik, "A probabilistic quality representation approach to deep blind image quality prediction," Aug. 2017, pp. 1–12, *arXiv:1708.08190*. [Online]. Available:https://arxiv.org/abs/1708.08190

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

[17] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018.

[18] W. Zhou, L. Yu, Y. Zhou, W. Qiu, M.-W. Wu, and T. Luo, "Blind quality estimator for 3D images based on binocular combination and extreme learning machine," *Pattern Recognit.*, vol. 71, pp. 207–217, Nov. 2017.

[19] F. Shao, K. Li, W. Lin, G. Jiang, and Q. Dai, "Learning blind quality evaluator for stereoscopic images using joint sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 10, pp. 2104–2114, Oct. 2016.

[20] F. Shao, Z. Zhang, Q. Jiang, W. Lin, and G. Jiang, "Toward domain transfer for no-reference quality prediction of asymmetrically distorted stereoscopic images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 3, pp. 573–585, Mar. 2018.

[21] M.-J. Chen, L. K. Cormack, and A. C. Bovik, "No-reference quality assessment of natural stereopairs," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3379–3391, Sep. 2013.

[22] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "Learning a referenceless stereopair quality engine with deep nonnegativity constrained sparse autoencoder," *Pattern Recognit.*, vol. 76, pp. 242–255, Apr. 2018.

[23] L. Shen, J. Lei, and C. Hou, "No-reference stereoscopic 3D image quality assessment via combined model," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 8195–8212, 2018.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

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

[26] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, pp. 1–21, Jan. 2010.

[27] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. Eur. Workshop Vis. Inf. Process.*, Jun. 2013, pp. 106–111.

[28] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Process., Image Commun.*, vol. 28, no. 8, pp. 870–883, Dec. 2013.

[29] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Process., Image Commun.*, vol. 28, no. 9, pp. 1143–1155, 2013.

[30] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1940–1953, May 2013.

[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] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro- and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, May 2017.

[33] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Analysis of distortion distribution for pooling in image quality prediction," *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 446–456, Jun. 2016.

[34] 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.

[35] 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.
[36] A. Balanov, A. Schwartz, Y. Moshe, and N. Peleg, "Image quality assessment based on DCT subband similarity," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 2105–2109.
[37] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, Dec. 2009, Art. no. 659024.
[38] R. Bensalma and M. C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Syst. Signal Process.*, vol. 24, no. 2, pp. 281–316, 2013.
[39] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3400–3414, Nov. 2015.
[40] F. Shao, K. Li, W. Lin, G. Jiang, M. Yu, and Q. Dai, "Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 2971–2983, Oct. 2015.
[41] 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.
[42] 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. (CVPR)*, Jun. 2012, pp. 1098–1105.
[43] Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. S. Torr, "Deeply supervised salient object detection with short connections," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5300–5309.
[44] Y. Niu, Y. Yang, W. Guo, and L. Lin, "Region-aware image denoising by exploring parameter preference," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2433–2438, Sep. 2018.

YUZHEN NIU received the Ph.D. degree in computer science from Shandong University, China, in 2010. She was a Postdoctoral Researcher with the Department of Computer Science, Portland State University, Portland, OR. She is currently a Professor with the College of Mathematics and Computer Science, Fuzhou University, China. Her current research interests include computer vision, artificial intelligence, and multimedia.

DONG HUANG is currently pursuing the M.S. degree with the College of Mathematics and Computer Science, Fuzhou University, Fujian, China. His current research interests include image processing, computer vision, and artificial intelligence.

YIQING SHI received the B.S. degree in electronic information engineering from Fuzhou University, Fuzhou, China, in 2016, where she is currently pursuing the Ph.D. degree in communication and information system. Her current research interests include pattern recognition, image processing, and computer vision.

XIAO KE received the Ph.D. degree in artificial intelligence from Xiamen University, China, in 2011. He is currently an Associate Professor with Fuzhou University. His research interests include multimedia, computer vision, pattern recognition, machine learning, and their relations with innovative technologies.