Stereoscopic Image Quality Assessment by Deep Convolutional Neural Network

Yuming Fang, Jiebin Yan, Xuelin Liu, Jiheng Wang

PII: S1047-3203(18)30332-8
DOI: https://doi.org/10.1016/j.jvcir.2018.12.006
Reference: YJVCI 2371

To appear in: *J. Vis. Commun. Image R.*

Received Date: 23 May 2018
Revised Date: 11 October 2018
Accepted Date: 2 December 2018

Please cite this article as: Y. Fang, J. Yan, X. Liu, J. Wang, Stereoscopic Image Quality Assessment by Deep Convolutional Neural Network, *J. Vis. Commun. Image R.* (2018), doi: https://doi.org/10.1016/j.jvcir.2018.12.006

Stereoscopic Image Quality Assessment by Deep Convolutional Neural Network

Yuming Fang[a,*], Jiebin Yan[a], Xuelin Liu[a], Jiheng Wang[b]

[a] *School of Information Management, Jiangxi University of Finance and Economics, Nanchang, China*
[b] *Department of Electrical and Computer Engineering, University of Waterloo, ON, N2L3G1 Canada*

Abstract

In this paper, we propose a no-reference (NR) quality assessment method for stereoscopic images by deep convolutional neural network (DCNN). Inspired by the internal generative mechanism (IGM) in the human brain, which shows that the brain first analyzes the perceptual information and then extract effective visual information. Meanwhile, in order to simulate the inner interaction process in the human visual system (HVS) when perceiving the visual quality of stereoscopic images, we construct a two-channel DCNN to evaluate the visual quality of stereoscopic images. First, we design a Siamese Network to extract high-level semantic features of left- and right-view images for simulating the process of information extraction in the brain. Second, to imitate the information interaction process in the HVS, we combine the high-level features of left- and right-view images by convolutional operations. Finally, the information after interactive processing is used to estimate the visual quality of stereoscopic image. Experimental results show that the proposed method can estimate the visual quality of stereoscopic images accurately, which also demonstrate the effectiveness of the proposed two-channel convolutional neural network in simulating the perception mechanism in the HVS.

Keywords: Image quality assessment; stereoscopic images; no reference;

*Corresponding author
Email address: fa0001ng@e.ntu.edu.sg (Yuming Fang)

convolutional neural network

1. Introduction

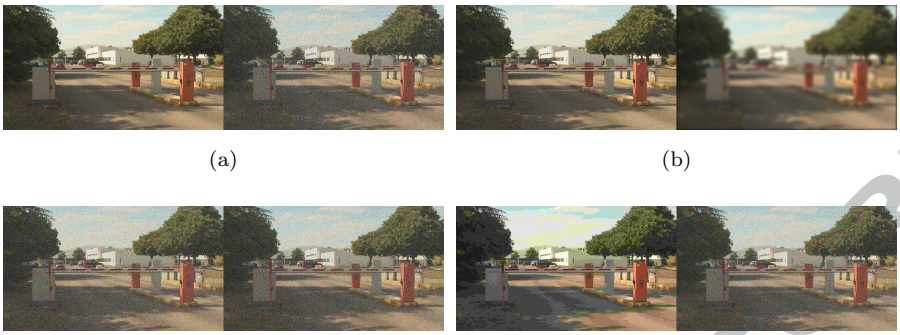
Recently, with the development of 3D imaging technology and multimedia applications, more and more stereoscopic images appear in our daily life, including 3D movies, 3D games and so on. The research on stereoscopic image processing has also attracted the attention of both academia research and industry applications. In the 3D image processing, transmission and reconstruction, stereoscopic images are inevitably affected by noise which may result in the loss of visual information, and image quality degradation. Image quality degradation not only affects the users' visual experiences but also its own value. Therefore, stereoscopic image quality assessment is crucial for various stereoscopic image processing applications. Stereoscopic image quality assessment includes subjective evaluation [1][2] and objective evaluation [3][4]. It is well known that subjective evaluation is time-consuming and power consumption, and subjective evaluation results are influenced by environment and other factors. Meanwhile, subjective evaluation cannot be embedded in multimedia applications. Thus, it is urgent to develop objective quality assessment methods which can predict the visual quality of stereoscopic images automatically.

Over the past few years, there have been many image quality assessment (IQA) methods proposed including full-reference (FR), reduced-reference (RR) and no-reference (NR) metrics. FR metrics require the complete reference information to estimate visual quality of images, such as peak signal-to-noise ratio (PSNR), Structure Similarity (SSIM) [5], SFUW [6]; RR metrics only need part of reference information to evaluate the visual quality of images [7][8]; NR metrics do not require any reference information for visual quality estimation of images, such as blind image spatial quality evaluator (BRISQUE) [9], Fang [10], NRLT [11], and Wu [12][13][14][15]. All these aforementioned IQA methods are designed for 2D images. Compared with 2D image, there are one left- and one right-view images in a stereopair. The difference between stereoscopic

IQA and traditional IQA is that stereoscopic image has depth information, and there are binocular vision characteristics in the perception of stereoscopic image visual quality, including binocular integration and binocular rivalry. For stereoscopic image quality assessment, one straightforward solution is utilizing 2D-IQA method to evaluate visual quality of left- and right-view images respectively, and then calculate the final quality score of stereoscopic image by combining the estimated visual quality scores of single-views based on a weighting strategy. As reported in [16][17], directly applying 2D-IQA method to estimate visual quality of stereoscopic images can obtain accurate evaluation results in the case of symmetrical distortion, but it obtains poor performance in the case of asymmetrical distortion, which cannot be consistent with the subjective perception. Some visual examples about symmetrical and asymmetrical distortions are illustrated in Fig. 1. For the symmetrically distorted stereoscopic image, there is the same amount of distortion in both left- and right-view images. For asymmetrically distorted stereoscopic image, there are different distortion levels or different kinds of distortions in left- and right-view images [2]. Designing universal visual quality assessment method for stereoscopic images is of great importance and practical in real-world applications.

2. Related Work

Compared with FR and RR methods, NR method is much desired in multimedia applications since it does not require information from reference images when evaluating image visual quality. The main idea of NR methods is exploiting discriminant features for image visual degradations, and the most successful IQA methods are based on Natural Scene Statistics (NSS) [18]. In [10], Fang *et al.* proposed a NR visual quality assessment method for contrast-distorted images based on moment and entropy features, in which the features' histograms are fitted by Gaussian distributions and an extreme value probability distribution respectively, and probability values are used as quality-aware features. Different from the method in [10], Mittal *et al.* proposed to use the general-



(a) (b)

(c) (d)

Figure 1: The visual examples of symmetric distortion and asymmetric distortion of stereoscopic images. (a) altered additive Gaussian noise with different distortion levels. (b) is altered by additive GN for left view and GB for right view; (c) is altered by additive GN with same distortion levels for both views; (d) is altered by additive JPEG compression for left view and additive GB for right view.

ized Gaussian distribution (GGD) to fit luminance value distribution of images, and the parameters of GGD are employed to perceive image distortion [9]. The common point of these two methods [9][10] is that both of them require fitting process. Instead of representing features by fitting function, Fang *et al.* adopted histogram to extract statistical luminance and texture features, and built a NR visual quality estimation model for screen content images. All of these three methods [9][10][11] follow two-step framework including feature extraction and regression by machine learning metric (support vector regression is widely adopted). And the same framework have been also adopted in visual quality assessment metrics for stereoscopic images. In [2], the authors developed a NR binocular IQA model by incorporating NSS features, where both 2D and 3D features are included. There have been also some visual quality assessment approaches proposed for stereoscopic images based on other machine learning metrics. In [19], Zhou *et al.* first computed binocular rivalry response (BRR) and binocular energy response (BER) of stereoscopic image, where BRR is a perceptual effect that occurs when both eyes see mismatched left- and right-

view images at the same retinal location, and BER is an important property of the HVS that can be represented as responses of a couple of monocular simple cells perceived by both eyes. The authors utilized the generalized local directional pattern and the local magnitude pattern to extract local patterns of BRR and BER for stereoscopic image quality prediction based on K-nearest neighbor method. In [20], the authors first constructed an over-complete dictionary matrix, by which they computed the coefficient vector of stereoscopic images. Then, KNN method is used to build stereoscopic quality assessment model based on the assumption that stereoscopic images with similar quality-aware features have similar visual quality.

Recently, convolutional neural network (CNN) has shown the superiority in computer vision, such as image classification [21], object detection [22][23] and action recognition [24]. CNN also has been applied in image visual quality assessment. In [25], the authors investigated the qualitative evaluation model for image visual quality assessment, where the NSS features extracted from wavelet domain are used to represent images and as the input of deep belief net. In [26], Kang *et al.* introduced CNN to design image visual quality estimation model, where the image patches in the spatial domain are taken as input instead of hand-crafted features [25] and the size of image patches are set to 32×32, while only one convolutional layer and two fullly connected layers (FCNs) are included. Inspired by deep residual network [21], the authors presented an end-to-end CNN model with two sum layers for image visual quality estimation [27], 32×32 image pathes are used as input and FSIM [28] values are employed as labels. Different from these two methods [26][27], Kim *et al.* designed a two-step CNN framework including local metric score regression and subjective score regression [29]. In the first step, the authors used four FR metrics including SSIM [5], GMSD [30], FSIM [28], VSI [31] to produce labels of image patches, and trained a CNN mode to estimate image local quality scores; in the second step, the patches of an image are used as the input of CNN, and pooled to the subjective quality score. The aforementioned three methods [26][27] contain an obvious drawback, all of them used objective quality estimation methods to

produce labels as substitute of subjective scores which makes the labels of image pathes contain much noise.

Owing to the fixed input of CNN, it greatly limits the application scope of the CNN models designed for visual quality assessment of 2D images, they cannot be applied to evaluate the visual quality of stereoscopic images directly. Thus, it is necessary to design CNN-based model for visual quality assessment of stereoscopic images. Recently, researchers have proposed using CNN to construct stereoscopic image quality assessment methods. In [32], Zhang *et al.* introduced 3-channel CNN for stereoscopic image quality assessment, where the input of CNN include left image patch, right image patch and difference image patch. In that model, the subjective quality score of the whole stereoscopic image is used as the output of the deep network for training. In [33], the authors proposed a NR IQA method for stereoscopic images based on binocular integration (BI) index and binocular self-similarity (BS). In that study, the BI index is used to evaluate monocular distortions and it is obtained by the trained deep neural network, while the BS index is computed by measuring the similarity of synthesized and original left-view images. Similar with the metric proposed in [29], Oh *et al.* proposed a no reference stereoscopic image quality assessment method based on CNN via local to global feature aggregation [34], where the FR stereoscopic image quality assessment method [16] is used to obtain local quality score and train the CNN model, then the patches extracted from stereoscopic image are used as the input to train the final stereoscopic image visual quality estimation model by local feature aggregation. The comparison of these three methods are summarized in Table 1. Although there are many CNN-based stereoscopic image visual quality estimation methods, it is still challenging to evaluate visual quality of stereoscopic images accurately. It is meaningful to put forward to an effective visual quality assessment model for stereoscopic images.

The contributions of this paper are summarized as follows: 1) inspired by the inter generative mechanism in the brain, we propose to use deep convolutional neural network to extract the high-level semantic information of stereoscopic images; 2) considering the binocular visual characteristics, we propose to mimic

Table 1: Descriptions of relevant stereoscopic image quality assessment methods.

Databases	Zhang [32]	Lv [33]	Oh [34]
Input	Hand-crafted	Patch	Patch
Size of input	96	32×32	32×32
Number of CNN	None	2	2
Number of FCN	5	3	3
Size of output	1	1	1
Activation function	Relu	Relu	Relu
Loss function	L2 Norm	L2 Norm	L2 Norm

the interaction process by fusing the high-level semantic features of left and right views.

The remaining of this paper is organized as follows. Section 3 introduces the proposed method in detail. In Section 4, we provide the experimental results from different aspects to demonstrate the advantages of the proposed method. Section 5 summarizes the paper.

3. Proposed method

The study of visual perception has found that there exists internal generative mechanism (IGM) [35] in the human brain when perceiving and understanding visual information. Moreover, the studies [36][37][38][39] on binocular perception revealed that there exists complex internal interaction when perceiving the visual information of stereoscopic images, binocular integration occurs in the HVS when both left- and right-view images are perceived by the HVS simultaneously, and binocular rivalry occurs in the HVS when only left- or right-view image is perceived by the HVS. These previous findings motivate us to design two-channel DCNN based model for visual quality assessment of stereoscopic images, which incorporates visual information extraction and interaction as well as visual quality regression.

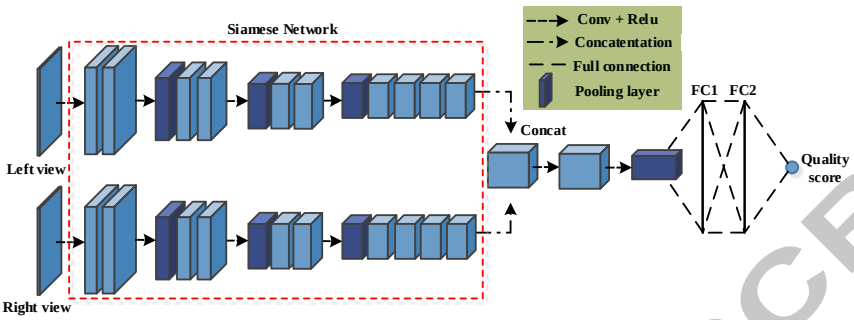


Figure 2: The proposed framework.

The framework of the proposed method is shown in Fig. 2. We feed the left- and right-view images into Siamese Network [40][41] containing four-group convolutional layers, and extract the high-level semantic features of the left- and right-view images, where the convolution kernel parameters of these two-channels in Siamese Network are shared. After that, we concatenate the extracted high-level feature maps of left- and right-view images and integrate these feature maps by convolutional operations to simulate the interactive process of the visual information in the HVS. Finally, the integrated feature maps are transformed into feature vector by using FCN, and the feature vector is presented in a non-linear way by the multi FCNs, the visual quality score of stereoscopic image is the output.

3.1. The Detailed Structure of The Proposed Framework

As shown in Fig. 2, the framework proposed in this paper includes two parts. The first part is used for high-level semantic information extraction of left- and right-view images, and the second part is used for visual quality estimation of stereoscopic image. In the first part, the input contain two image patches extracted from left- and right-view images in the same spatial position, and the size of image patch is set to 80×80. Previous work [42] shows that the high-level semantic information of the images can be effectively extracted by using 3×3 convolution kernel and stacking the convolutional layer repeatedly.

Thus following this work, we design a four-group CNN framework to extract the semantic information of left- and right-view images. The previous three groups contain only two convolution layers and the last group contains four convolution layers. There exists a pooling layer between each two groups, and we do not use down-sample operation in the fourth group for retaining the receptive field of left- and right-view images. Convolution operation is given as follows:

$$L_n^{(k)} = \sigma(\omega^k L_{n-1} + b^{(k)}) \tag{1}$$

where L_n^k denotes the k-th feature map in n-th layer; ω^k and b^k denote convolution kernel and bias which assigned by end-to-end learning; L_{n-1} denotes feature maps in (n-1)-th layer; σ is rectified linear unit (Relu) which can speed up the convergence of CNN and reduce the computational complexity. Relu is calculated as below:

$$L_n^{(k)} = max(0, L_n^{(k)}) \tag{2}$$

In order to reduce the complexity of the proposed DCNN model, each group convolution layer follows a pooling operation. The maximum pooling is adopted and the window size is set to 2×2, which means that we use the largest value to represent the receptive region.

The second part of the proposed DCNN framework is used for feature transformation and visual quality estimation of stereoscopic image. First, we concatenate the high-level semantic feature maps of left- and right-view images obtained by the proposed Siamese Network, and we can obtain the overall feature maps of stereoscopic image. The concatenation process is described as follows:

$$FM_o = Concat(FM_l, FM_r) \tag{3}$$

where FM_l and FM_r denote the high-level feature maps of left- and right-view images, respectively; and FM_o represents the fused feature maps of stereoscopic image.

Motivated by IGM [35] and binocular perception [36][37][38][39] which describe the mechanism of visual information extraction and interactive process

Table 2: The configurations of the proposed DCNN.

	Framework	Feature map	Parameters
Simese Network	Input	80×80×1	3×3×32
	First Group	80×80×32	3×3×64
	Second Group	40×40×64	3×3×128
	Third Group	20×20×128	3×3×256
	Fourth Group	10×10×256	
	Connected Layer	10×10×512	3×3×512
Part 2	Fusion Layer 1	5×5×512	3×3×512
	Fusion Layer 2	3×3×512	
	FCN 1	(3×3×512)	(3×3×512)×512
	FCN2	512	512×1
	Output	1	

of the visual information, we propose to fuse the overall feature maps of stereoscopic image by convolution operation. Then, we transform the fused feature maps into feature vector which is used to represent the visual quality of stereoscopic image by a non-linear way through the multi FCNs. The multi FCNs can be formulated as below:

$$Q = \sigma(\omega_2(\sigma(\omega_1 * f + b_1)) + b_2) \tag{4}$$

where Q represents the estimated visual quality score; f is feature vector; ω_1 and ω_2 denote the weighting matrixes of the first and the second FCN, respectively; b_1 and b_2 are two biases. These parameters including ω_1, ω_2, b_1 and b_2 are obtained by end-to-end learning. The detailed configurations of the proposed DCNN are summarized in Table 2.

3.2. Training Process

To train the proposed DCNN, the inputs are non-overlapping image patches extracted from stereoscopic images. Considering the balance between the size

of image receptive field and the requirement of large amount of samples when training the DCNN, we set the size of image pathes to 80×80. We augment the training data by rotating 90, 180 and 270 degrees. We also employ the subjective quality score of stereoscopic image to represent the labels of image pathes since these stereoscopic images have homogeneous distortions [43]. We use L_1 Norm as the loss function.

$$Loss = \frac{1}{M}\sum_{m=1}^{M} ||Q_m - P_m||^1 \tag{5}$$

where Q_m and P_m denote the estimated visual quality score and the subjective quality score, respectively; M represents bath-size, which is set as 20 in our experiments. We adopt the Adam optimization algorithm [44] for optimization. The learning rate α is set to 10^{-4} initially and subsequently descended by a factor of 10, we fixed α to 10^{-6} when loss is leveled off.

4. Experimental Results

4.1. Stereoscopic Image Databases

We use two 3D image quality databases to test the proposed NR 3D-IQA algorithm, which include LIVE 3D Image Quality Database Phase I and Phase II [1][2][16]. These two databases are shown in the following. It is worthy noting that LIVE Phase II contain symmetrically and asymmetrically distorted images. In the experiment, we use 80% images of each database to train the proposed DCNN model, and the rest of the same database are used for testing. For each database, we repeat this operation for 10 times and the median performance is reported.

LIVE 3D-IQA Database Phase I [1] This database is composed of 20 reference images and 365 symmetrically distorted ones by five common distortion types, including JPEG2000 (JP2K), JPEG, white noise (WN), Gaussian blur (GB) and Rayleigh fast-fading channel simulations (FF). GB is used to generate 45 distorted images, and other types of distortion are utilized to create 80 distorted images. The corresponding human scores are also provided in the

Table 3: Descriptions of 3D image quality databases.

Databases	LIVE Phase I	LIVE Phase II
Number of Subjects	32	32
Number of Images	365	360
Image Sizes	640×360	640×360
Distortions	JP2K, WN, GB, JPEG, FF	JP2K, WN, GB, JPEG, FF

form of DMOS. A lower DMOS denotes better visual quality of stereoscopic images.

LIVE 3D-IQA Database Phase II [2][16] Different from LIVE 3D-IQA Phase I database which only includes symmetrically distorted stereoscopic images, both asymmetrically and symmetrically distorted stereoscopic images are provided in LIVE 3D-IQA Phase II database. Five distortion types including JP2K, JPEG, WN, GB and FF are introduced to generate the distorted images. Each type of distortion is used to create 72 distorted images. Totally, there are 240 asymmetrically distorted and 120 symmetrically distorted stereoscopic images in this database.

The detailed descriptions of these databases are summarized in Table 3. As shown in Table 3, the resolutions of images in each database are the same.

4.2. Evaluation Methodology

We used two evaluation metrics to compute the correlation between the subjective and objective scores: Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank-order Correlation Coefficient (SRCC) [45]. Higher SRCC and PLCC values indicate better objective evaluation performance. Given the i-th image in the database (with N images in total), its objective and subjective scores are o_i and s_i. PLCC can be estimated as follows.

$$PLCC = \frac{\sum_{i=1}^{N}(o_i - \overline{o})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{N}(o_i - \overline{o}) * \sum_{i=1}^{N}(s_i - \overline{s})}}, \quad (6)$$

where $\overline{o}$ and $\overline{s}$ denote the mean values of o_i and s_i, respectively.

SRCC can be computed as follows.

$$SRCC = 1 - \frac{6\sum_{i=1}^{N} e_i^2}{N(N^2-1)}, \tag{7}$$

where e_i is the difference between the ith image's ranks in subjective and objective results.

Here, we use a five-parameter mapping function to nonlinearly regress the quality scores into a common space as follows:

$$f(x) = \Theta_1\left(\frac{1}{2} - \frac{1}{1+e^{(\Theta_2(x-\Theta_3))}}\right) + \Theta_4 x + \Theta_5 \tag{8}$$

where $(\Theta_1, ..., \Theta_5)$ are fitted by using subjective and objective quality scores, and they are determined in the testing stage.

4.3. Comparison Experiments and Discussions

The tested 2D-IQA or 3D-IQA models include FR methods PSNR, SSIM [5], MS-SSIM [16], IDW-SSIM [17], and NR methods Chen [2], Zhou [19], Shao [46], Lv [33], Zhang [32] and Oh [34]. Among the compared FR methods, PSNR and SSIM [5] are designed for quality assessment of 2D images, MS-SSIM [16] and IDW-SSIM [17] are designed for quality assessment of stereoscopic images. All of the compared NR methods are designed for quality assessment of stereoscopic images. The performance of IQA methods is evaluated by two criteria: PLCC and SRCC. It is worth noting that the results of the compared 3D-IQA methods are taken from their corresponding original papers.

The experimental results on LIVE 3D Image Quality Database Phase I and Phase II are shown Table 4, where it can be seen that on both Phase I and Phase II, the proposed method obtains the best quality prediction performance in terms of PLCC and SRCC. The consistent results from Phase I (symmetric distortion only) and Phase II (both symmetric and asymmetric distortion) suggest that the proposed method can automatically account for the relationship between IGM and binocular perception by the proposed two-channel DCNN.

Table 4: Experimental results on the LIVE 3D-IQA databases.

Databases	LIVE Phase I		LIVE Phase II	
Methods	PLCC	SRCC	PLCC	SRCC
PSNR	0.834	0.834	0.665	0.665
SSIM	0.873	0.877	0.802	0.793
MS-SSIM [16]	0.917	0.916	0.900	0.889
IWD-SSIM [17]	0.873	0.874	**0.916**	**0.919**
Chen [2]	0.895	0.891	0.880	0.880
Zhou [19]	0.928	0.887	0.861	0.823
Shao [46]	0.907	0.896	0.848	0.824
Lv [33]	0.901	0.898	0.870	0.862
Zhang [32]	**0.947**	**0.943**	-	-
Oh [34]	0.943	0.935	0.863	0.871
Proposed	**0.957**	**0.946**	**0.946**	**0.934**

As shown in Table 4, these 2D-IQA methods including PSNR and SSIM cannot accurately predict the visual quality of stereoscopic images, the main reason is that it does not consider the characteristics of stereoscopic images by averaging the estimated quality scores of left- and right-view images directly. With the consideration of binocular rivalry mechanism, IDW-SSIM can obtain accurate stereoscopic image quality estimation results on asymmetrical distortions, the weighting strategy in IDW-SSIM works not very well when applied to symmetrical distorted stereoscopic images, the results indicate the difficulty in designing the universal visual quality assessment method for stereoscopic images. Here, the outperformance of the proposed method indicates that the simple combination of the quality scores of left- and right-view images cannot sufficiently represent complex and non-intuitive interactions between multiple 3D visual cues including image quality, depth quality and visual comfort interactions during 3D visual perception. With the help of introduced IGM and

binocular perception inspired two-channel DCNN, the proposed method can achieve higher accuracy in predicting the visual quality of stereoscopic images.

Compared to the conventional approaches [2][19][46], these CNN based stereoscopic image visual quality evaluation models including Zhang [32], Oh [34] and the proposed method can obtain significant better performance in estimating visual quality of stereoscopic images, the experimental results demonstrate the validity of CNN when applied to stereoscopic image quality assessment. Among these methods including Zhang [32], Oh [34] and the proposed method, the proposed method obtains the best performance, especially for asymmetrical distorted stereoscopic images. The main reason is that we consider the internal generative mechanism and complex inner perception process in human brain when perceiving the stereoscopic images. In Zhang [32], the authors used subjective quality score of stereoscopic image to label the 32×32 image patch, which is too small to represent the whole stereoscopic image and it makes the labels of image patches contain much noise. At the same time, the CNN model proposed in [32] only include two convolution layers, it can not extract the high-level semantic features of the stereoscopic images effectively. Meanwhile, it does not take into account the interaction process in the brain when perceiving the stereoscopic images, just simply combines the feature vectors of the left- and right-view images and form the final feature representations of stereoscopic image. In Lv [33], the author employed hand-crafted features as the input of FCNs. Due to the limitation of the number of data, the fully connected layer is easily overfitting which affects the generalization of the network. At the same time, the author just used the trained FCNs to calculate the visual quality the left- and right-view images independently. In Oh [34], the author used FR stereoscopic image visual quality evaluation model [2] to generate labels of the image patches. The labels produced by this kind of data augmentation method might introduce a lot of noise. The noise data generated by this method is used to train this CNN model, its performance will be subject to the performance of FR visual quality evaluation metric.

Unlike these aforementioned methods, we take account of both the number

Table 5: PLCC performance for individual distortion types on the LIVE 3D-IQA databases Phase I.

Databases	LIVE Phase I				
Distortions	JP2K	JPEG	WN	GB	FF
PSNR	0.785	0.219	0.935	0.916	0.703
SSIM	0.865	0.487	0.939	0.919	0.721
Chen [2]	0.907	0.695	0.917	0.917	0.735
Shao [46]	0.901	0.458	0.916	**0.952**	-
Zhang [32]	**0.926**	**0.740**	**0.944**	0.930	**0.883**
Proposed	**0.975**	**0.753**	**0.973**	**0.953**	**0.868**

of training samples and the size of the receptive field of the image patches, and set the size of the input image patches to 80×80. And we extract the high-level semantic features of the left and right views through the weight sharing network, which accelerate the convergence rate of the network. The convolution operation is used to fuse the high-level semantic features of the left- and right-view images to simulate the inner interaction in human brain. Experimental results show that the proposed stereoscopic image visual quality evaluation method in this paper can get more consistent results with subjective perception that other relevant methods.

Moreover, in Tables 5 and 6, we report PLCC scores on individual distortion types to validate the generalization ability of the proposed method on LIVE 3D image quality databases. As shown in Tables 5 and 6, directly applying 2D-IQA method to stereoscopic image quality prediction works well in the case of white noise distortion. From Tables 5 and 6, it can be observed that the proposed method, which does not attempt to recognize the distortion types or give any specific treatment for any specific distortion type, pronounces the best prediction performance for all distortion types.

Table 6: PLCC performance for individual distortion types on the LIVE 3D-IQA databases Phase II.

Databases	LIVE Phase II				
Distortions	JP2K	JPEG	WN	GB	FF
PSNR	0.597	0.491	0.919	0.690	0.730
SSIM	0.704	0.678	0.922	0.838	0.834
Chen [2]	**0.867**	**0.867**	**0.950**	0.900	**0.933**
Shao [46]	0.826	0.828	0.928	**0.984**	-
Zhang [32]	-	-	-	-	-
Proposed	**0.975**	**0.952**	**0.972**	**0.983**	**0.929**

4.4. Cross-database Validation

We have conducted a cross-validation experiment to further validate the proposed method. We use one of LIVE Phase I and II databases as the training set and the other as the test set, the training strategies are same as those adopted in the proposed method. The experimental results are shown in Table 7. From Table 7, we can observe that the proposed method still can obtain accurate estimation results, which demonstrates the robustness of the proposed method. The performance of the proposed method when training on LIVE Phase II is better than it when training on LIVE Phase I, the main reason is that LIVE Phase II contains both symmetrical and asymmetrical distorted stereoscopic images, the proposed DCNN framework trained on LIVE Phase II can capture both symmetrical and asymmetrical distortions.

5. Conclusion and Future Work

Inspired by the IGM of the brain, in this paper, we propose a novel NR 3D-IQA metric for stereoscopic images based on two-channel DCNN. First, in order to simulate the visual information process in the brain, we design a DCNN-based framework to extract the high-level semantic information of the

Table 7: Experimental results of cross-database validation. 1) Use Phase I as training set and Phase II as testing set; 2) use Phase II as training set and then Phase I as testing set.

	Training Database	
Performance	LIVE Phase I	LIVE Phase II
PLCC	0.811	0.899
SRCC	0.797	0.898

left- and right-view images. Then, considering that binocular fusion and binocular rivalry occur in the HVS when stereoscopic image is perceived by the HVS, and there is a complex inner perception generative process in the brain before making a decision about stereoscopic image visual quality. Thus, we design a feature fusion network to integrate the high-level semantic information of the left- and right-view images, and simulate the inner inference mechanism in the brain by the multi-layer convolution layer. Finally, the visual characteristics of the stereoscopic image are presented in a non-linear way through the multi-layer FCN, and the visual quality score of the stereoscopic image is the output. Experimental results show that the proposed method can obtain accurate evaluation results when estimating the visual quality of stereoscopic images, and the proposed model works well in asymmetrical distortions and can obtain higher performance than other relevant methods, which also prove the validity of the proposed visual feature extraction and fusion approach.

Acknowledgement

This work was supported in part by the Natural Science Foundation of China under Grant 61822109 and 61571212, Fok Ying-Tung Education Foundation under Grant 161061 and by the Natural Science Foundation of Jiangxi under Grant 20181BBH80002.

References

[1] A. K. Moorthy, C.-C. Su, A. Mittal, A. C. Bovik, Subjective evaluation of stereoscopic image quality, Signal Processing: Image Communication 28 (8) (2013) 870–883.

[2] M.-J. Chen, L. K. Cormack, A. C. Bovik, No-reference quality assessment of natural stereopairs, IEEE Transactions on Image Processing 22 (9) (2013) 3379–3391.

[3] Y. Fang, J. Yan, J. Wang, No-reference quality assessment for stereoscopic images by statistical features, in: IEEE International Conference on Quality of Multimedia Experience, 2017, pp. 1–6.

[4] J. Wang, K. Zeng, Z. Wang, Quality prediction of asymmetrically distorted stereoscopic images from single views, in: IEEE International Conference on Multimedia and Expo, 2014, pp. 1–6.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612.

[6] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li, Z. Guo, Objective quality assessment of screen content images by uncertainty weighting, IEEE Transactions on Image Processing 26 (4) (2017) 2016–2027.

[7] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, IEEE Journal of Selected Topics in Signal Processing 3 (2) (2009) 202–211.

[8] J. Wu, W. Lin, Y. Fang, L. Li, G. Shi, I. N. S, Visual structural degradation based reduced-reference image quality assessment, Signal Processing: Image Communication 47 (2016) 16–27.

[9] A. Mittal, A. K. Moorthy, A. C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Transactions on Image Processing 21 (12) (2012) 4695–708.

[10] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, G. Zhai, No-reference quality assessment for contrast-distorted images based on natural scene statistics, IEEE Signal Processing Letters 22 (7) (2015) 838–842.

[11] Y. Fang, J. Yan, L. Li, J. Wu, W. Lin, No reference quality assessment for screen content images with both local and global feature representation, IEEE Transactions on Image Processing 27 (4) (2018) 1600–1610.

[12] Q. Wu, H. Li, F. Meng, K. N. Ngan, B. Luo, C. Huang, B. Zeng, Blind image quality assessment based on multichannel feature fusion and label transfer, IEEE Transactions on Circuits and Systems for Video Technology 26 (3) (2016) 425–440.

[13] Q. Wu, H. Li, Z. Wang, F. Meng, B. Luo, K. N. Ngan, Blind image quality assessment based on rank-order regularized regression, IEEE Transactions on Multimedia 19 (11) (2017) 2490–2504.

[14] Q. Wu, H. Li, K. N. Ngan, K. Ma, Blind image quality assessment using local consistency aware retriever and uncertainty aware evaluator, IEEE Transactions on Circuits and Systems for Video Technology 28 (9) (2018) 2078–2089.

[15] Q. Wu, H. Li, F. Meng, K. N. Ngan, A perceptually weighted rank correlation indicator for objective image quality assessment, IEEE Transactions on Image Processing 27 (5) (2018) 2499–2513.

[16] M.-J. Chen, C.-C. Sun, D.-K. Kwon, L. K. Cormack, A. C. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, Signal Processing: Image Communication 28 (9) (2013) 1143–1155.

[17] J. Wang, A. Rehman, K. Zeng, S. Wang, Z. Wang, Quality prediction of asymmetrically distorted stereoscopic 3D images, IEEE Transactions on Image Processing 24 (11) (2015) 3400–3414.

[18] H. R. Sheikh, A. C. Bovik, L. K. Cormack, No-reference quality assessment using natural scene statistics: Jpeg2000, IEEE Transactions on Image Processing 14 (11) (2005) 1918–1927.

[19] W. Zhou, L. Yu, Binocular responses for no-reference 3D image quality assessment, IEEE Transactions on Multimedia 18 (6) (2016) 1077–1084.

[20] W. Zhou, W. Qiu, M. Wu, Utilizing dictionary learning and machine learning for blind quality assessment of 3-D images, IEEE Transactions on Broadcasting 63 (2) (2017) 404–415.

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

[22] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (9) (2015) 1904–1916.

[23] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (6) (2017) 1137–1149.

[24] S. Ji, W. Xu, M. Yang, K. Yu, 3D convolutional neural networks for human action recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (1) (2012) 221–231.

[25] W. Hou, X. Gao, D. Tao, X. Li, Blind image quality assessment via deep learning, IEEE Transactions on Neural Networks and Learning Systems 26 (6) (2015) 1275–1286.

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

[27] B. Bare, K. Li, B. Yan, An accurate deep convolutional neural networks model for no-reference image quality assessment, in: IEEE International Conference on Multimedia and Expo, 2017, pp. 1356–1361.

[28] L. Zhang, X. M. L. Zhang, D. Zhang, FSIM: A feature similarity index for image quality assessment, IEEE Transactions on Image Processing 20 (8) (2011) 2378–2386.

[29] J. Kim, S. Lee, Fully deep blind image quality predictor, IEEE Journal of Selected Topics in Signal Processing 11 (1) (2017) 206–220.

[30] W. Xue, L. Zhang, X. Mou, A. C. Bovik, Gradient magnitude similarity deviation: A highly effecient perceptual image quality index, IEEE Transactions on Image Processing 23 (2) (2014) 684–695.

[31] L. Zhang, Y. Shen, H. Li, VSI: A visual saliency-induced index for perceptual image quality assessment, IEEE Transactions on Image Processing 23 (10) (2014) 4270–4281.

[32] W. Zhang, C. Qu, L. Ma, J. Guan, R. Huang, Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network, Pattern Recognition 59 (C) (2016) 176–187.

[33] Y. Lv, M. Yu, G. Jiang, F. Shao, Z. Peng, F. Chen, No-reference stereoscopic image quality assessment using binocular self-similarity and deep neural network, Signal Processing: Image Communication 47 (9) (2016) 346–357.

[34] H. Oh, S. Ahn, J. Kim, S. Lee, Blind deep S3D image quality evaluation via local to global feature aggregation, IEEE Transactions on Image Processing 26 (10) (2017) 4923–4936.

[35] J. Wu, W. Lin, G. Shi, A. Liu, Perceptual quality metric with internal generative mechanism, IEEE Transactions on Image Processing 22 (1) (2014) 43–54.

[36] L. Kaufman, Sight and mind: An introduction to visual perception, American Journal of Psychology 87 (4).

[37] M. J. M. Levelt, The alternation process in binocular rivalry, British Journal of Psychology 57 (3) (1966) 225–238.

[38] D. V. Meegan, L. B. Stelmach, W. J. Tam, Unequal weighting of monocluar inputs in binocular combination: Implications for the compression of stereoscopic imagery, Journal of Experimental psychology Applied 7 (2) (2001) 143–153.

[39] R. Blake, D. H. Westendorf, R. Overton, What is suppressed during binocular rivalry?, Perception 9 (2) (1980) 223–231.

[40] S. Chopra, R. Hadsell, Y. LeCUN, Learning a similarity metric discriminatively, with application to face verifiction, in: IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 1–1.

[41] S. Zagoruyko, N. Komodakis, Learning to compare image patches via convolutional neural networks, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–1.

[42] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, Computer Science.

[43] Y. Peng, D. Doerman, No-reference image quality assessment using visual codebooks, IEEE Transactions on Image Processing 21 (7) (2012) 3129–3138.

[44] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, CoRR abs/1412.6980.

[45] VQEG, Final report from the video quality experts group on the validation of objective models of video quality assessment, March 2000, [Online]. Available:http://www.vqeg.org.

[46] F. Shao, W. Lin, S. Wang, G. Jiang, M. Yu, Q. Dai, Learning receptive fields and quality lookups for blind quality assessment of stereoscopic images, IEEE Transactions on Cybernetics 46 (3) (2016) 730–743.

Highlights

- Inspired by the inter generative mechanism in the brain, we propose to use deep convolutional neural network to extract the high-level semantic information of stereoscopic images;
- Considering the binocular visual characteristics, we propose to mimic the interaction process by fusing the high-level semantic features of left and right views.