SIViP
DOI 10.1007/s11760-015-0784-2

ORIGINAL PAPER

No-reference image quality assessment using Prewitt magnitude based on convolutional neural networks

Jie Li[1] · **Lian Zou**[1] · **Jia Yan**[1] · **Dexiang Deng**[1] · **Tao Qu**[1] · **Guihui Xie**[1]

Received: 21 December 2014 / Revised: 12 May 2015 / Accepted: 14 May 2015

Abstract No-reference image quality assessment is of great importance to numerous image processing applications, and various methods have been widely studied with promising results. These methods exploit handcrafted features in the transformation or space domain that are discriminated for image degradations. However, abundant a priori knowledge is required to extract these handcrafted features. The convolutional neural network (CNN) is recently introduced into the no-reference image quality assessment, which integrates feature learning and regression into one optimization process. Therefore, the network structure generates an effective model for estimating image quality. However, the image quality score obtained by the CNN is based on the mean of all of the image patch scores without considering the human visual system, such as edges and contour of images. In this paper, we combine the CNN and the Prewitt magnitude of segmented images and obtain the image quality score using the mean of all the products of the image patch scores and weights based on the result of segmented images. Experimental results on various image distortion types demonstrate that the proposed algorithm achieves good performance.

Keywords No-reference image quality assessment · Convolutional neural networks (CNNs) · Graph-based image segmentation · Prewitt magnitude

✉ Lian Zou
zoulian@whu.edu.cn

Jie Li
trackerdsp@163.com

[1] School of Electronic Information, Wuhan University, Wuhan 430072, Hubei, China

1 Introduction

With the rapid development and popularity of digital cameras, digital images have become the most important mean for us to acquire and exchange information. However, because of the degradation in image acquisition, storage and compression transmission process, the final image we get describes a degraded version of the original scene. Thus, image quality assessment (IQA) has become extremely important for most image processing applications. Subjective evaluation and objective evaluation are two existing image quality evaluation methods [1]. In subjective method, quality is evaluated by organized groups of human observers to mark the distorted images, which is time-consuming and expensive. In general, objective image quality measures can be classified into three categories: full reference (FR) IQA, reduced-reference (RR) IQA and no-reference (NR) IQA. FR-IQA directly compares the distorted images and their corresponding ideal images, thus obtaining the quality assessment of the distorted images. State-of-the-art FR measures, such as VIF [2] and FSIM [3], are well fitted with human judgment of quality. RR-IQA needs the partial information of reference image to compute images quality degradations [4]. The method presented in [5] accesses image quality by deploying the phase and magnitude in the frequency domain, which can serve as both a FR and an RR metric which is rarely mentioned in the previous literature. However, in many practical applications, the information of perfect versions of distorted images is not available, so NR-IQA is desirable. By exploiting features of discriminated ability, NR-IQA measures can directly acquire image quality without a reference image. The BIQI [6] uses several natural scene statistics (NSS) features to complete a simple NR-IQA. Then extracting features in image transformations, DIIVINE [7] is in the wavelet domain and BLIINDS-II [8] is in the DCT domain, which are usually

Published online: 29 May 2015

computationally expensive because of using image transformations. Through exacting features in the spatial domain, CORNIA [9] and BRISQUE [10] are efficient in computation time.

Recently, a new method for no-reference image quality assessment was proposed by Le Kang [11], and the structure of Le Kang's CNN has one convolutional layer with max and min pooling, two fully connected layers and an output node. In the literature of [12,13], systematic machine learning approaches are adopted in the domain of image quality assessment, which first acquire handcrafted features and then deploy machine learning for feature pooling process. However, CNN can automatically and directly learn features from data (images) and acquire quality scores from the output of it, which is different from [12,13]. As a result, CNN has the ability to accurately predict quality score on small image patches and integrate the feature learning and regression into one optimization process instead of using handcrafted features. Therefore, the CNN can accurately predict image quality without a reference image and achieve state-of-the-art performance on the LIVE dataset. In our study of the Le Kang's CNN, however, we found that it acquires the image quality only by averaging the patch scores without taking the property of human visual system (HVS) into account, such as edge and contour of images. According to many research results, human visual system pays close attention to the contour and edge information of an image; the contour and edge information are the sensitive information of an image's structure for human to understand the scene [14,15].

In this paper, an improved no-reference image quality assessment algorithm is put forward. An analogous network of CNN, the research priorities of [11], is applied to obtain quality scores of corresponding input image patches. Then, the Prewitt magnitude of segmented images is introduced to compute the weight of corresponding image patches. After weighting quality scores of image patches to get the mean of weighted quality scores, the final calculated quality score indicates the quality of images; a final quality score is calculated which means the quality of images.

The innovation of this article is mainly in two aspects. Firstly, we design the structure of CNN by ourselves and solve the traditional problem of quality evaluation through the perspective of deep learning. Secondly, our module combines the advantages of deep learning and the HVS which is sensitive to the edges and contours of the image. By using the segmentation algorithm to obtain the segmented image and using the Prewitt operator on the segmented image, the module can obtain the weighted values in line with the HVS, which improves the consistency of the subjective and objective image quality evaluation.

The remainder of this paper is organized as follows. In Sect. 2, the CNN is introduced for no-reference image quality assessment. Section 3 describes the improved algorithm in detail. Section 4 presents the experimental results and discussion. Finally, Sect. 5 draws the conclusion.

2 The CNN for no-reference image quality assessment

Recently, deep neural networks attract wide attention and have achieved good results in various computer vision tasks [16,17]. Specifically, the CNN in previous literature has acquired pretty good results on many applications. Before this, the CNN was primarily designed for object recognition and has not been applied in NR-IQA. A new philosophy of CNN for image quality measurement was proposed by Le Kang, which consists of five layers and is a $32 \times 32 - 26 \times 26 \times 50 - 2 \times 50 - 800 - 800 - 1$ structure and is similar to the architecture 1 of Fig. 2. Firstly, a local normalization is performed on a given grayscale images; Secondly, sampling non-overlapping patches from it, then using Le Kang's CNN to estimate the quality score for each patch; Finally, all the patch scores are averaged to obtain the final quality estimation for the image. A simple local contrast normalization is applied on the input of image patches, let $I(\mathrm{i},\mathrm{j})$ denotes the intensity value of a pixel at location (i, j), normalized value $\hat{I}(\mathrm{i},\mathrm{j})$ is computed as follows:

$$\begin{aligned}
\hat{I}(\mathrm{i},\mathrm{j}) &= \frac{I(\mathrm{i},\mathrm{j})-\mu(\mathrm{i},\mathrm{j})}{\sigma(\mathrm{i},\mathrm{j})+C} \\
\mu(\mathrm{i},\mathrm{j}) &= \frac{1}{(2\times M+1)\times(2\times N+1)} \sum_{m=-M}^{m=M} \sum_{n=-N}^{n=N} I(\mathrm{i}+m,\mathrm{j}+n) \\
\sigma(\mathrm{i},\mathrm{j}) &= \sqrt{\sum_{m=-M}^{m=M} \sum_{n=-N}^{n=N} (\mathrm{I}(\mathrm{i}+m,\mathrm{j}+n)-\mu(m,n))^2}
\end{aligned} \tag{1}$$

where C is a constant that prevents instabilities from dividing by zero. M and N are the normalization window sizes. Each pixel may have a different local mean and variance in the local normalized image patch. In the convolutional layer, there are 50 feature maps, each of which is generated through a filter that convolutes the local normalized image patches. Then, a min pooling and a max pooling are applied on each convolutional layer, reducing the filter responses to a lower dimension. Suppose that the response of the feature map obtained by the k-th filter at location (i,j) is $R_{i,j}^k$, and the max and min values of μ_k and υ_k are given by

$$\begin{aligned}
\mu_k &= \max_{i,j} R_{i,j}^k \\
\upsilon_k &= \min_{i,j} R_{i,j}^k
\end{aligned} \tag{2}$$

where $k = 1, 2, \ldots, K$ and K is the number of kernels. The pooling method decreases each feature map to a two-

dimensional feature vector. Therefore, a size $2 \times K$ vector will be the input of each node of the next fully connected layer. Le Kang trains his network on non-overlapping 32×32 patches taken from large images. In the test process, the average of all the patch scores of the predicted image is represented as the final quality score. Suppose the score of i-th image patch is z_i , Le Kang computes the image level quality score Z as follows:

$$Z = \frac{\sum_{i=1}^{K} z_i}{K} \tag{3}$$

where K represents the number of small image patches in original image. This novel framework allows learning and prediction of image quality on local regions which is rarely shown in previous literature.

3 Description of the improved algorithm in detail

Many research results show that human visual system pays close attention to the contour and edge information of an image [16,17]. Based on HVS, we propose an improved algorithm: the CNN framework using gradient based on segmented images, which obtains an image quality score through the average of all the product of the image patch scores and weights based on the result of segmented images, instead of the average of the image patch scores. The detail description about the proposed algorithm is as follows.

3.1 Graph-based image segmentation algorithm

In our framework, image segmentation is adopted to capture perceptual grouping, which is very important in human visual perception. And the segmentation method should have the properties of capturing perceptually important groupings or regions and high efficiency.

The image segmentation algorithm proposed by Pedro F. Felzenszwalb and D. Huttenlocher is based on graph theory [18]. And the result of segmented image is shown in Fig. 1. The formula of $G = (\mathrm{V}, \mathrm{E})$ represents an indirect graph of the original image, and $v_i \in V$ is a set of the elements that are divided corresponding to the pixels of image. Each edge $(v_i, v_j) \in \mathrm{E}$ corresponds to pairs of neighboring vertices and obtains a non-negative weight $w((v_i, v_j))$ which is a measure of the dissimilarity between neighboring elements v_i and v_j corresponding to the dissimilarity of the two pixels connected by this edge (e.g., the difference in color intensity, motion, location or some other local attribute). In the graph-based approach, the algorithm can get the segmentation result S and each component (region) $C \in S$ is a connected part of the graph $G' = (\mathrm{V}', \mathrm{E}')$, where $E' \in E$. In other words, a subset of the edges in E includes any component in segmentation result. The elements in the same component are intended to be similar and elements in the different component should be dissimilar. This means, the weights of edges between two vertices in the same component is relatively low and that of edges between two vertices in different should be higher [18].

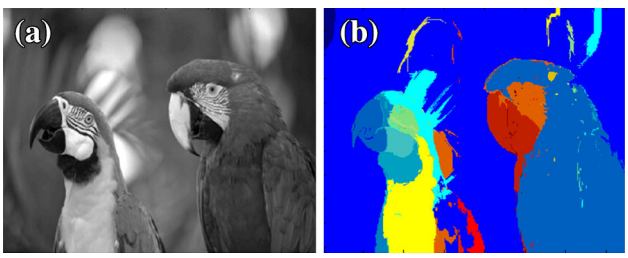

Fig. 1 An example of graph-based image segmentation. **a** Original image **b** segmented image

In [18], the algorithm proposes a predicate as D. Comparing the difference between elements along the boundary of two different regions (inter-difference) with difference among elements within each of the two regions (internal difference), the predicate evaluates whether or not there is evidence for a boundary between two regions of an image. The internal difference of a component C is defined to be the largest weight in the minimum spanning tree of the component, MST(C,E) ,where $C \subseteq V$.

That is

$$Int(\mathrm{C}) = \max_{e \in MST(\mathrm{C},\mathrm{E})} \omega(e) \tag{4}$$

The difference between two components $C_1, C_2 \subseteq V$ is defined to be the minimum weight edge connecting the two components. That is,

$$Dif\ (\mathrm{C}_1, \mathrm{C}_2) = \min_{v_i \in C_1, v_j \in C_2, (v_i, v_j) \in \mathrm{E}} \omega\left(v_i, v_j\right) \tag{5}$$

Based on the definition of (4) and (5), the pairwise comparison predicate is as follow, where $|C|$ represents the size of C, and k is same constant parameter.

$$\begin{aligned} &D(\mathrm{C}_1, \mathrm{C}_2) = \begin{cases} true & if\ Dif(\mathrm{C}_1, \mathrm{C}_2) > \mathrm{MInt}(\mathrm{C}_1, \mathrm{C}_2) \\ false & otherwise \end{cases} \\ &where, \\ &MInt(\mathrm{C}_1, \mathrm{C}_2) = \min(\mathrm{Int}(\mathrm{C}_1) + \tau(\mathrm{C}_1), \mathrm{Int}(\mathrm{C}_2) + \tau(\mathrm{C}_2)) \\ &\tau(\mathrm{C}) = k\,|C| \end{aligned} \tag{6}$$

3.2 Proposed method

Convolutional neural networks (CNNs) derived from the biological basis of the 1981 Nobel in medicine, according to the Hubel and Wiesel early work on the cat's visual above. In the visual cortex above, we know that there is a cell complex distribution of these cells that are very sensitive to some local

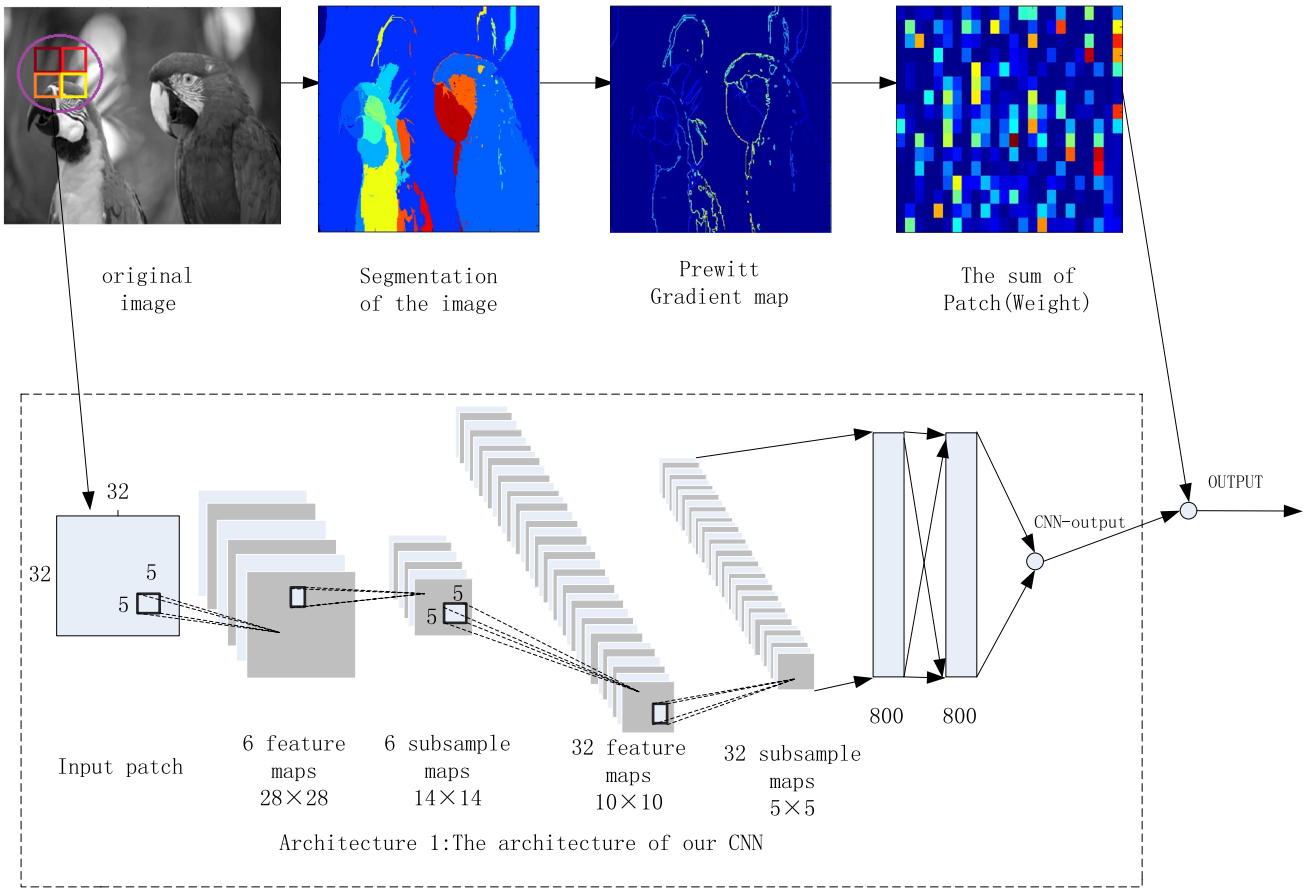

Fig. 2 Main components of our NR-IQA algorithm

input. They are called as perception, and adopt the combination of this special way to cover the entire field of version. These filters are sensitive to the input local space, so they can better detect the spatial correlation of different object in the natural images.

In this section, we present our no-reference NR-IQA algorithm in detail. Our algorithm is composed of three parts. The detailed description is shown in Fig. 2.

The first part is the proposed network, which consists of seven layers and is a $32\times32-28\times28\times6-14\times14\times6-10\times10\times32-5\times5\times32-800-800-1$ structure. The input is a normalized 32 × 32 pixel image patches. We employ [19] as a local contrast normalization. The normalized value $\hat{I}$ (i, j) is computed as follows:

$$\begin{aligned}\hat{I}(\mathrm{i},\mathrm{j}) &= \frac{I(\mathrm{i},\mathrm{j})-\mu(\mathrm{i},\mathrm{j})}{\sigma(\mathrm{i},\mathrm{j})+\mathrm{C}}\\ u(\mathrm{i},\mathrm{j}) &= \sum_{k=-K}^{K}\sum_{l=-L}^{L}\omega_{k,l}I_{k,l}(\mathrm{i},\mathrm{j})\\ \sigma(\mathrm{i},\mathrm{j}) &= \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L}\omega_{k,l}\left(\mathrm{I}_{k,l}(\mathrm{i},\mathrm{j})-\mu(\mathrm{i},\mathrm{j})\right)^2}\end{aligned} \tag{7}$$

where $\omega=\{\omega_{k,l}|k=-K,\ldots K,l=-L,\ldots L\}$ represents a two-dimensional circularly symmetrical Gaussian weighting function which samples out to three standard deviations and rescales to unit volume. In the implementation, we pick $K=L=3$. Compared with the (1), these normalized luminance values of natural images strongly tend toward a unit normal Gaussian characteristic. In early human vision, an operator of (7) can be adopted to model the contrast-grain masking processing.

The first layer is convolutional layer with six feature maps, and the map size is 28 × 28. Each element in every feature map is connected to a 5 × 5 neighborhood in the image patch. Followed with the convolutional layer is a max subsampling layer. In the every feature map, each unit is connected to a 2 × 2 neighborhood in the corresponding feature map in the first layer. The third layer is also a convolutional layer with 32 feature maps, and the map size is 10 × 10. Each element in every feature map is connected to all of the 5 × 5 neighborhoods at identical location in the second layer. Followed with the third layer is another max subsampling layer. In the every feature map, each unit is connected to a 2 × 2 neighborhood in the corresponding feature map in the fourth layer. The fifth and sixth layers are two fully connected lay-

ers, each of which has 800 nodes. The last layer gives the quality score of image patch, which is a linear regression with one-dimensional output. In order to obtain the network weights a, we employ an objective function as follow:

$$L = \frac{1}{N}\sum_{n=1}^{N} ||f(x_n; a) - y_n||_{l1}$$
$$a' = \min_a L \tag{8}$$

where x_n represents the input image patch and y_n represents its ground truth. And $f(x_n; a)$ is the predicted score of x_n using the network weights a. In the test process, we can use the network to predict the score z. In training process, we use the architecture 1 in Fig. 2 to train the parameters.

The second part is to compute the weight of the corresponding image patch. In image processing, gradient computation is a traditional topic to extract edge information of images which can be expressed by special masks and there are many gradient operators Canny, Sober, Roberts, Prewitt, etc. In this paper, the gradient map of an image is produced by using the Prewitt operator. Firstly, the Graph-based image segmentation is used to acquire the segmentation of the original image, and then the Prewitt operator is adopted to get the gradient map. Lastly, through computing the sum of corresponding patch, we can get the corresponding weight h. Suppose $\eta(i, j)$ to be the gradient-value at location (i, j) in the Prewitt gradient map, then we compute the corresponding weight of the patch h as follow:

$$h = \sum_{k=-K}^{K}\sum_{l=-K}^{K} \eta(i, j) \tag{9}$$

where $K = 16$.

Through the formula (10), the algorithm acquires the final predicted score Z,

$$Z = \frac{\sum_{i=1}^{m} h_i z_i}{\sum_{i=1}^{m} h_i} \tag{10}$$

where $i = 1, 2 \ldots m$ represent the sequence number of small image patches, m is the total number of patches in the original image. Taking the human visual system into account, we use the corresponding weight of patches in (10), which is different with (3).

4 Experimental results and discussion

In this Section, the performance of the proposed algorithm is evaluated based on the LIVE [20] and TID2008 [21] datasets. The LIVE dataset contains five different distortions: JP2K compression (JP2K), JPEG compression (JPEG), Gaussian blur (BLUR), fast fading (FF) and white noise compression (WN), which includes 29 original RGB color images and 779 distorted images. Difference mean opinion scores (DMOS)

Table 1 Experimental results of SROCC and LCC on LIVE. The result of best performance is marked in bold

	JP2K	JPEG	BLUR	FF	WN	All
SROCC						
PSNR	0.870	0.885	0.763	0.874	0.942	0.866
SSIM	0.939	0.946	0.907	0.941	0.964	0.913
FSIM	0.970	0.981	0.972	0.949	0.967	0.964
Q (c)	0.972	0.950	0.968	0.941	0.988	0.954
DIVINE	0.913	0.910	0.921	0.863	0.984	0.916
BLIINDS-11	0.929	0.942	0.923	0.889	0.969	0.931
BRISQUE	0.914	0.965	0.951	0.877	0.979	0.940
CORNIA	0.943	0.955	**0.969**	0.906	0.976	0.942
Le's CNN	0.952	0.977	0.962	0.908	0.978	0.956
Our method	**0.964**	0.935	0.941	**0.945**	**0.988**	**0.958**
LCC						
PSNR	0.873	0.876	0.779	0.870	0.926	0.856
SSIM	0.921	0.955	0.893	0.939	0.893	0.906
FSIM	0.910	0.985	0.978	0.912	0.978	0.960
Q (c)	0.965	0.966	0.960	0.942	0.986	0.949
DIVINE	0.922	0.921	0.923	0.888	0.988	0.917
BLIINDS-11	0.935	0.968	0.938	0.896	0.980	0.930
BRISQUE	0.922	0.973	0.951	0.903	0.985	0.942
CORNIA	0.951	0.965	**0.968**	0.917	0.987	0.935
Le's CNN	0.953	0.981	0.953	0.933	0.984	0.953
Our method	**0.978**	0.977	0.945	**0.960**	**0.993**	**0.966**

Table 2 The performance of 20, 40 and 100 iterations in JP2K with 60 % for train set and 40 % for test set

	20 Iterations	40 Iterations	100 Iterations
LCC	0.961	0.964	**0.978**
SROCC	0.947	0.952	**0.964**

Table 3 The result of 20 iterations of JP2K with 40 and 60 % for train set, corresponding as 60 and 40 % for test set

The percentage of train	40 % (Train)	60 % (Train)
SROCC	0.938	**0.947**
LCC	0.940	**0.961**

Table 4 LCC and SROCC obtained by training on LIVE and testing on TID2008 for each distortion

	JP2K	JPEG	WN	GB
LCC	0.8485	0.9367	0.8944	0.8792
SROCC	0.8390	0.9377	0.9000	0.8679

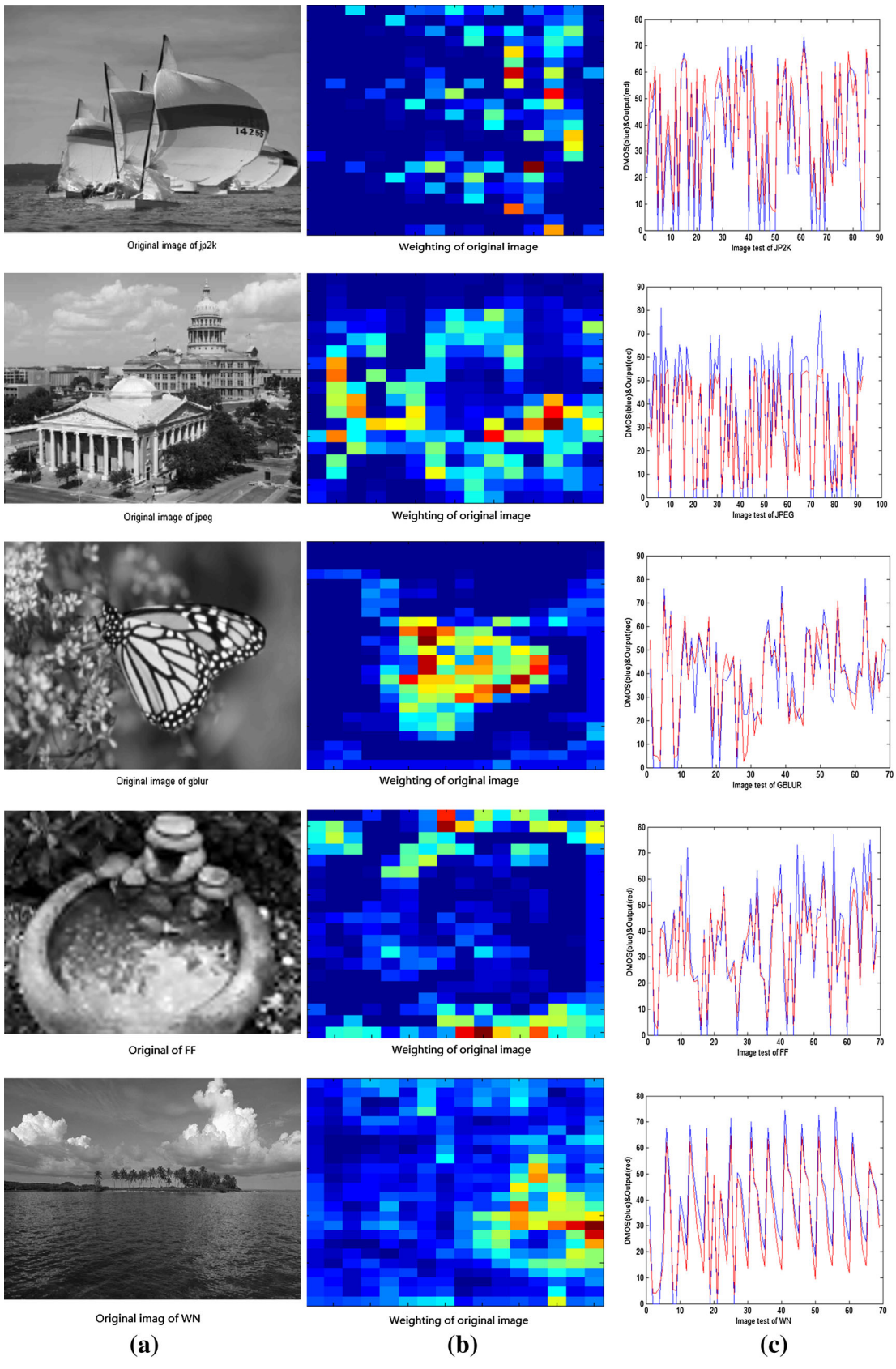

Fig. 3 **a** Original image of specified distortion type. **b** The weighting map of the patches of images at the corresponding location. **c** The fit between DMOS value and the prediction

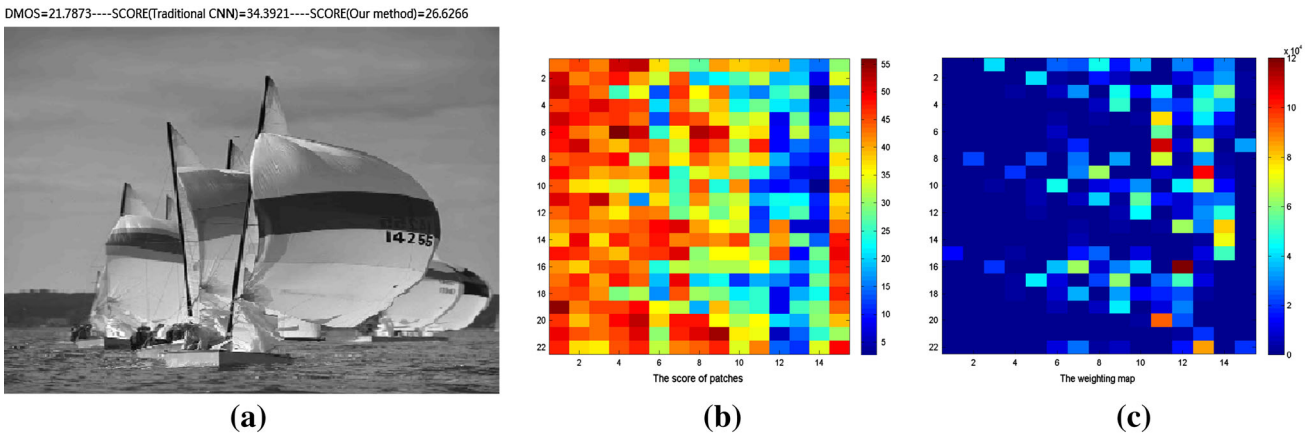

Fig. 4 An example of our algorithm's advantage

are associated with each image, and scores are in the range [0, 100]. Higher DMOS denotes lower quality. And the TID2008 dataset involves 1700 distorted images with 25 reference images which have been processed by 17 different types of distortions. Mean opinion scores (MOS) are associated with each image, and scores are in the range [0, 9]. Higher MOS shows higher quality, which is different with DMOS. The performance validation metrics such as Spearman rank-order correlation coefficient (SROCC) and the correlation coefficient (CC) are used as quantitative measures of evaluation.

4.1 Performance evaluation

The simulation results in Table 1 are obtained from 100 train iterations to learn the module, and we randomly select 60 % of reference images and their corresponding distorted images as the training set, and the test set is remaining 40 %. From Table 2, we find that the performance of JP2K improves by increasing the number of iterations. And the performance of other distortions also improves by increasing the number of iterations. For the reason that 100 train iterations cost a lot of time, we only use the instance of 20 iterations of JP2K to show the performance of different train percentage. The observation from Table 3 shows that the just little more than the training samples have better performance.

We train and test on the five distortions: JP2K, JEPG, BLUR, FF and WN on the LIVE dataset for distortion-specific experiment, and we train and test images of five distortions together.

Table 1 shows the performance of previous state-of-the-art IQA methods. And the result of best performance is marked in bold. As can be seen from it, the performance of our method is well across the five distortions. Furthermore, our method performs best on JP2K, WN and FF distortion types and all of five distortions together in both the metric of SROCC and LCC.

An original image of specified distortion type, the weighting map of corresponding original image and the fit between DMOS value and the prediction are shown in Fig. 3. It can be seen that our method performed quite well on the five distortion types, but it should also be noted that the difference of DMOS value and the prediction is lager when the DMOS value is bigger than 70. It means our method cannot predict the quality of images which are badly distorted. In Fig. 3b, the weights in most of the region are very low, and the weights with large values are relatively less. For example, in the picture of butterfly in the third line, the weights with large values locate on the edges of the butterfly which just rightly ignores the information of blur background that humans do not care. So we can see that our method is in line with HVS. The fit between DMOS value and the prediction is shown in the third column which demonstrates the good ability to predict.

Figure 4 shows an example of our algorithm's advantage. Fig. 4a is a distorted image from the JP2K in LIVE database whose DMOS value is 21.7873. Only using the architecture 1 in Fig. 2, we can obtain the score of patches of the image which is shown in Fig. 4b. And the score of image using the traditional CNN is 34.3921 which is the mean of all patches of image from the output of the architecture 1. Fig. 4c in Fig. 4 is the weighting map of Fig. 4a which is the sum of patch (weight) in Fig. 2. Combination with the Fig. 4c, the output is 26.6266, which is closer to the DMOS value. Our algorithm improves the performance of subjective and objective consistency which is important for NR-IQA.

4.2 Cross-database test

Since the proposed method involves training, the experiment in Table 4 is designed to test the generalization ability of our method. We train our CNN on LIVE and test on TID2008,

and only the shared four distortion types are examined. We can see that the proposed method performs well in TID2008 dataset.

5 Conclusion

In this paper, a novel image quality assessment (IQA) metric is proposed by combining the CNN and Prewitt magnitude based on segmented image. It takes full advantage of deep neural networks that integrates feature learning and regression into one optimization process and also takes HSV that is more intuitive to human into account. The HVS is sensitive to edge and contour information in an image. We incorporate such information into our metric, which is the primary reason for its better performance than only using the CNN. As research of deep learning and HVS continues, the method for NR-IQA metric will be improved in the future.

References

1. Thung, K.H., Paramesran, R.: A survey of image quality measures. In: Proceedings of international conference for technical postgraduates (TECHPOS). pp. 1–4 (2009)
2. Sheikh, H.R., Bovik, A.C., de Veciana, G.: An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Trans. Image Process. **14**(12), 2117–2128 (2005)
3. Zhang, L., Zhang, D., Mou, X., Zhang, D.: FSIM: a feature similarity index for image quality assessment. IEEE Trans. Image Process. **20**(8), 2378–2386 (2011)
4. Li, Qiang, Wang, Zhou: Reduced-reference image quality assessment using divisive normalization-based image representation. IEEE Signal Process. Soc. **2**(3), 202–211 (2009)
5. Narwaria, M., Lin, W., McLoughlin, I., Emmanuel, S., Chia, L.T.: Fourier transform based scalable image quality measure. IEEE Trans. Image Process. **21**(8), 3364–3377 (2012)
6. Moorthy, A.K., Bovik, A.C.: A two-step framework for constructing blind image quality indices. IEEE Signal Process. Lett. **17**(5), 513–516 (2010)
7. Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: from natural scene statistics to perceptual quality. IEEE Trans. Image Process. **20**(12), 3350–3364 (2011)
8. Saad, M., Bovik, A.C., Charrier, C.: Blind image quality assessment: a natural scene statistics approach in the DCT domain. IEEE Trans. Image Process. **21**(8), 3339–3352 (2012)
9. Ye, P., Kumar, J., Kang, L., Doermann, D.: Unsupervised feature learning framework for no-reference image quality assessment. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1098–1105 (2012)
10. Mittal, A., Moorthy, A., Bovik, A.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)
11. Kang, L., Ye, P.: Convolutional neutral networks for no-reference image quality assessment. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
12. Narwaria, M., Lin, W.: Objective image quality assessment based on support vector regression. IEEE Trans. Neural Netw. **21**(3), 515–519 (2010)
13. Narwaria, M., Lin, W.: SVD-based quality metric for image and video using machine learning. IEEE Trans. Syst. Man Cybern. Part B **42**(2), 347–364 (2012)
14. Narwaria, M., Lin, W.S., Enis Cetin, A.: Scalable image quality assessment with 2D mel-cepstrum and machine learning approach. Pattern Recognit. **45**, 299–313 (2012)
15. Chen, G.-H., Yang, C.-L., Xie, S.-L.: Gradient-based structural similarity for image quality assessment, conference: image processing. IEEE International Conference-ICIP, pp. 2929–2932 (2006)
16. Deng, L., Hinton, G.E., Kingsbury, B.: New types of deep neural network learning for speech recognition and related applications: an overview. In: IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP 2013) (2013)
17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. NIPS **1**, 4 (2012)
18. Felzenszwalb, PedroF, Huttenlocher, DanielP: Efficient graph-based image segmentation. Int. J. Comput. Vis.—IJCV **59**(2), 167–181 (2004)
19. Mittal, A., Moorthy, A., Bovik, A.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)
20. Sheikh, H.R., Wang, Z., Cormack, L., Bovik, A.C.: LIVE image quality assessment database release2. http://live.ece.utexas.edu/research/quality
21. Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., Battisti, F.: TID2008: a dataset for evaluation of full-reference visual quality assessment metrics. Adv. Mod. Radio Electron. **10**, 30–45 (2009)