Signal Processing: Image Communication 81 (2020) 115703

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

An ensemble image quality assessment algorithm based on deep feature clustering☆

Tianliang Bian

Intelligent Risk Control Department, Qianbao Financial, Beijing, China

ARTICLE INFO

Keywords:
Feature extraction
Clustering
Image quality assessment
PSNR
SIMM
VIF

ABSTRACT

To effective handle image quality assessment (IQA) where the images might be with sophisticated characteristics, we proposed a deep clustering-based ensemble approach for image quality assessment toward diverse images. Our approach is based on a convolutional DAE-aware deep architecture. By leveraging a layer-by-layer pre-training, our proposed deep feature clustering architecture extracted a fixed number of high-level features at first. Then, it optimally splits image samples into different clusters by using the fuzzy C-means algorithm based on the engineered deep features. For each cluster, we simulated a particular fitting function of differential mean opinion scores with each assessed image's PSNR, SIMM, and VIF scores. Comprehensive experimental results on TID2008, TID2013 and LIVE databases have demonstrated that compared to the state-of-the-art counterparts, our proposed IQA method can reflect the subjective quality of images more accurately by seamlessly integrating the advantages of three existed IQA methods.

1. Introduction

At present, most image compression algorithms attempt to remove visual redundant information at the cost of the loss of visual information of the source image. Image quality assessment (IQA) after image compression algorithm or compression equipment directly reflects the performance of the algorithm/equipment, based on which image quality evaluation becomes a hot research topic nowadays. At the same time, the image quality assessment approaches can also be utilized to improve the fault tolerance of the image coding and decoding algorithm. Although image processing researchers have conducted plenty of investigations and made some achievements [1–4] in the related fields, the performance of image and video quality evaluation methods is still far from satisfaction.

The image quality assessment (IQA) can be split into two sub-topics: subjective estimation-based approaches and objective estimation-based approaches. The latter approaches estimate image quality by inspecting the quantitative indexes or parameters inherent in the quality model. The current research focuses on the latter and aims to enforce each image quality evaluation process accurately reflect the subjective quality of human visual perception.

Most image quality evaluation is based on the physiological characteristics of the human visual system (HVS), which is a highly complex and nonlinear system. Nevertheless, the current research on visual cognition of HVS is still limited. In this way, IQA can only be carried out based on certain assumptions and some restricted prior knowledge. Therefore, the existing IQA methods only partially reflect the impact of image content on human eyes and visual cognition, which might not completely represent the quality of images. Even worse, different image quality evaluation models exhibit their own advantages and disadvantages during the quality evaluation of images with different characteristics.

Peak signal-to-noise ratio (PSNR) [5] is a commonly used indicator to measure signal distortion, but PSNR does not involve the features of signal content. In this way, the quality evaluation of some images or video sequences exhibits a large deviation from the human subjective perception

The SSIM [2] is a simple algorithm to measure the similarity between the original signal and the processed signal based on the structural information. Empirically, the SSIM has a strong correlation with human subjective quality evaluation and thus can obtain better quality evaluation results than the PSNR. Notably, it is feasible for the SSIM method to have two signals with the same structure and similar value with respect to the same original signal. Thus, the subjective quality of the two signals might still be different. SSIM cannot completely solve the problem of PSNR.

In order to apply the aforementioned image evaluation algorithms more flexibly and accurately, in this paper, we designed a novel hybrid model for image quality assessment of diverse images. By extracting 10 high-abstract visual features from each image, the FCM clustering algorithm is conducted iteratively on these features to divide image

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.image.2019.115703.

E-mail address: tianliang.bian@gmail.com.

https://doi.org/10.1016/j.image.2019.115703
Received 25 June 2019; Received in revised form 3 November 2019; Accepted 11 November 2019
Available online 15 November 2019

dataset into multiple groups. Each group was rendered by utilizing a fitting function for accurate differential mean opinion scores (DMOS). In total, we summarize our contributions as follows:

(1) The proposed a framework tailored for image processing. It can effectively extract features from various images. The only fully connected layer existed in the middle of the architecture for ten extracted features. Simultaneously, we designed a layer-wise pre-training strategy on the convolutional DAE-based model.

(2) We utilized the FCM clustering algorithm to guarantee that the clustering procedure runs more efficient. Moreover, the loss of clustering can be propagated to the encoder part for feature weights fine-tuning.

(3) Based on the groups containing different image features, the values of fitting functions as predictions of DMOS are more close to the HVS scores.

2. Related work

For the related fields in this paper, the previously influential work is introduced with respect to three topics concerned: deep neural network, clustering methods, and image quality assessment methods.

2.1. Deep neural network

Deep learning models have a variety of architectures, such as deep belief networks, convolution neural networks, and automatic encoder, etc.

The stacked automatic encoder [6] was proposed based on the automatic encoder by adding the noise on each sample of the training set. According to the comparison between the reconstructed signal and original training samples as reconstruction error, the stacked automatic encoder has a strong ability in the tolerance of noises.

Since deep belief networks (DBN) ignored the two-dimensional structure of images, each position has to be learned separately to detect the weights for a particular visual feature, which remarkably increases the computational burden. The convolutional deep belief network (CDBN) was proposed [7,8] by combining the deep belief network with the convolutional neural network. By sharing weights in all positions of an image, in practice, the CDBN is widely applied in many domains like voice recognition and face recognition [9,10].

Ciresan et al. [11] applied multi-column deep convolution networks in six databases, including MNIST [12], NISTSD 19 [13], casia-hwdb1.1 [14], NORB [15], GTSRB [16] and cifar-10 [17]. With the assistance of the multi-thread GPU training, sample enhancement techniques and multi-model averaging, both the effectiveness and efficiency of this deep convolutional network are validated empirically.

Sermanet et al. [18] proposed to handle identification, positioning and detection tasks with deep convolution networks. The designed networks is comprised of multiple networks of different width-to-height ratios and sizes. The abstract features learned from the convolutional network were used as the input of the regression network to learn four coordinates of the object bounding box. This network have demonstrated the outstanding feature learning performance by leveraging the convolutional network. This architecture can derive not only the classification information but also the location and size information of the object.

2.2. Clustering methods

Clustering is a common data analysis tool, with the purpose of dividing a large number of samples into several categories. In this way, samples within each group are dense while ones in different groups are loose.

K-means-based approaches [19–23] are one of the most popular clustering methods. The k-means denotes the centroid of each category by the weighted average of all samples within, which is called the clustering centroid. Although the k-means-like clustering methods cannot be used to represent category attributes, it can optimally reflect the geometric/statistical significance of clustering for data with numerical attributes.

The FCM clustering algorithm [24] is a distance-based data partition algorithm. The FCM algorithm iteratively performs clustering partition until its target formula reaches the minimum value. Moreover, the FCM algorithm is a soft clustering algorithm. It utilizes the membership matrix to divide each sample according to its probability of belonging to each category. Therefore, each sample can belong to over one class, wherein the probability of each class is different.

DBSCAN [25] is a typical density-based clustering algorithm, which represents clusters as the points corresponding to density, and then performs clustering by enlarging the possible high-density areas.

The objective of self-organizing map (SOM) [26] is to utilize an artificial neural network for clustering. It is also a typical special case of vector quantization. This method exhibits two main features. First, it adopts an incremental strategy that the entire data points are processed one by one. Second, it can project the entire clustering center point onto a two-dimensional plane to facilitate visualization.

With the advancement of deep learning recently, deep clustering methods as variants of deep learning in unsupervised learning, are attracting increasing attention of machine learning researchers and is utilized popularly in domains such as dimension reduction, image recognition and etc.

The deep embedded clustering model (DEC) [27] and its upgraded version (IDEC) [28] are the representatives of the deep clustering methods. IDEC can enhance the performance of DEC by preserving the local information. In order to alleviate the distortion of embedded space during fine-tuning, the sum of relative entropy and reconstruction loss were employed as the loss function of deep architecture during pre-training layer-wise. In contrast to DEC, IDEC guaranteed the representativeness of embedded spatial features.

2.3. Image quality assessment methods

According to well-known SSIM [2] proposed by Wang et al. in 2004, the structural information extraction is the key function of human eyes which are highly adaptive to changes of signal structure. Based on this seminal work, many SSIM-based algorithms were proposed by encoding human visual system features. For example, the IQA method based on the ratio of visual signal to noise within the Wavelet domain was proposed by Chandler D M et al. [29] with visual perception threshold.

Since the quality evaluation of the compressed image is based on the jpeg and jpeg2000 format, Zhou Wang et al. [30] proposed a frequency-domain IQA algorithm to measure the distortion degree according to the peak value of distorted images at some specific frequencies and the transfer of energy from high frequency to low frequency.

From the perspective of information theory, Hamid et al. [31] proposed two image quality algorithms, information fidelity criterion (IFC) and visual information fidelity (VIF), by calculating the mutual information between the original image and the distorted image. To our best knowledge, these methods were the pioneers to represent the correspondence between the inherent content and visual quality of images.

According to the PQS algorithm [32] proposed by Makoto Miyahara et al. image distortions are divided into brightness differences, spatial frequency distortion, structural interference, and random error. First, the degree of different types of distortion is captured by designing features. Then, visual features are selected by the principal component analysis (PCA). Finally, the weights of different features in quality evaluation is determined by learning on the training image set.

The natural scene statistics (NSS) [33] algorithm performed wavelet decomposition on enormous natural images to calculate the statistical characteristics, such as the wavelet coefficients with different scales,

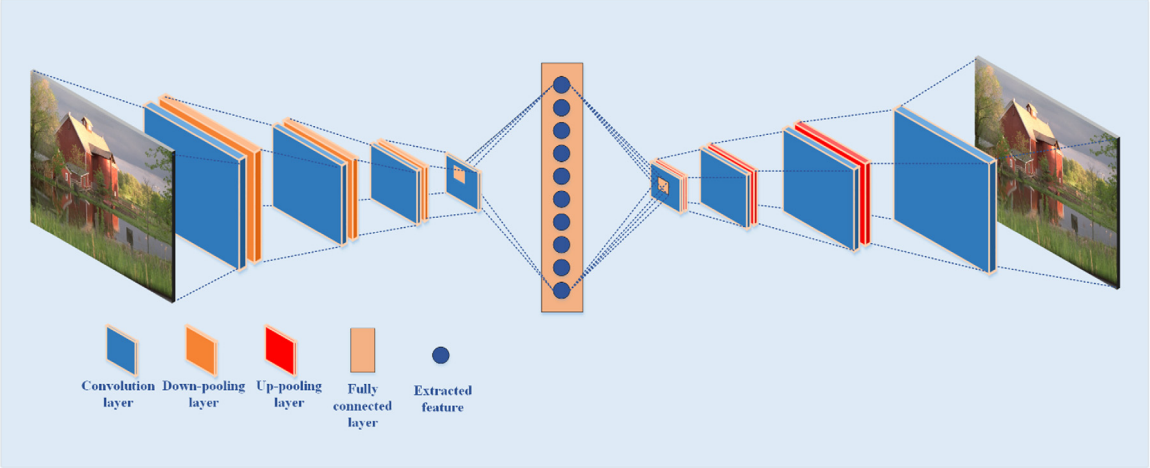

Fig. 1. The architecture of the DAE, which consists of eight convolutional layers, three down-pooling, three up-pooling layers, and one fully connected layer in the middle of architecture.

directions and positions. Subsequently, the Gaussian scale mixture (GSM) method was utilized to model these statistical features.

GAO and Weisi et al. [34,35] proposed a contour-let decomposition method based on the original visual feature to extract the visual sensitivity coefficient in each image. Through a statistical comparison between the distorted image and the visual sensitivity coefficient of the original image, the quality evaluation measure of the distorted image was calculated.

Ismail et al. [36] adopted the analysis of variance to investigate the sensitivity of various evaluation algorithms in multiple factors such as distortion types (JPEG compression, SPIHT compression, blur, additive noises) and image content (weaving, face, natural scene). The results have shown that one IQA algorithm is sensitive to a few types of visual distortion while less sensitive to the others.

In recent years, there were various novel algorithms were proposed. Hou et al. [37] presented a blind image quality assessment algorithm to directly learn qualitative measurement of images, where natural scene statistics features were utilized for image representation. Ahar et al. [38] proposed a novel method to verify the effectiveness of scene structure and perception quality. Sebastian et al. [39] presented a DNN-based IQA algorithm. The proposed DNN architecture consisted of 10 convolutional layers and 5 pooling layers, connected with 2 fully-connected layers. The proposed method was purely data-driven and did not require any hand-crafted features.

3. Proposed method

Our method for image quality assessment is formulated by three pipelines. The purpose of the first step is to extract the high-level representative features from assessed images by deep autoencoder architecture. Afterward, the FCM clustering method was employed on these abstract features for the clustering of assessed images while the loss of clustering result backpropagates to further fine-tune the encoder part of DAE delicately. For each clustered group, we finally simulate the fitting function for DMOS with PSNR, SSIM, and VIF scores of assessed images within.

3.1. Architecture of DAE

As shown in Fig. 1, in order to extract embedded high-level features from images efficiently, we used convolutional layers instead of fully connected (FC) layers to extract high-level features from input images except for the middle FC layer with fixed size. Besides a single FC layer and eight convolutional layers, the DAE architecture involves three down-pooling and as many up-pooling layers, ensuring each output in the same size as the input image.

It is noticeable that our entire DAE architecture is based on a CNN which uses a single image as the input. As a popular deep architecture for image modeling, CNN can be formulated as a learning function: $f: X \rightarrow Y$, where X denotes a collection of training images and Y denotes the predicted labels. During the CNN training, each image patch is sequentially fed into CNN. If we use the second last layer deep representation as the input, the CNN learning process can be formulated as:

$$f(\mathbf{H}) = \sum_{i=1}^{N} \sum_{t \in Y} \kappa(t = c) \cdot \log p(t = c | \boldsymbol{d}_n, \mathbf{H}_c), \tag{1}$$

where N denotes the number of training images, $\mathbf{H}$ is the inherent parameters in the deep CNN, $\kappa(\cdot)$ is and indicator function, that is, if $t = c$, then $\kappa = 1$, otherwise $\kappa = 0$.

The probability $p((t = c | \boldsymbol{d}_n, \mathbf{H}_c)$ is given as follows:

$$p\left(\left(t = c | \boldsymbol{d}_n, \mathbf{H}_c\right)\right) = \frac{\exp(H^T d_n)}{\sum_{t \in Y} \exp(H^T d_n)}, \tag{2}$$

All the convolutional layers adopt parameters with 3×3 kernel, one padding, and one stride to keep the size of output feature maps the same as the input. In the encoder module, each convolutional layer is followed by a max-pooling layer in stride two for down-sampling the input feature maps with the half size. The decoder uses three max-pooling layers in stride 1/2 to up-sample the corresponding feature maps with the double size.

We utilized a mean square error as the reconstruction loss function of the DAE which is formulated as:

$$L_{DAE} = L_r = \sum_{i=1}^{n} \left\| x_i - f_d(f_e(x_i)) \right\|_2^2 \tag{3}$$

For each sample x_i within training set X of size n, encoder function f_e maps x_i to the embedding features with symbol h_i in the low-dimensional eigenspace. The decoder function f_d is the inverse function of f_e, which reconstructs xi by performing inverse transformation $f_d(f_e(x_i))$. (see Fig. 2.)

The training process of our deep model is a forward and back propagated process. In the forward propagation process, we represent the output o_j of each neuron at the jth statistical layer is calculated as:

$$o_k = \sum_{i=1}^{M} \sum_{j=1}^{K} r_{ij \rightarrow k} \rightarrow o'_{jk} \tag{4}$$

Noticeably, $r_{ij \rightarrow k}$ can be treated as the "contribution" of the ij-th neuron to the kth neuron. In the back propagation step, we denote σ_t as the

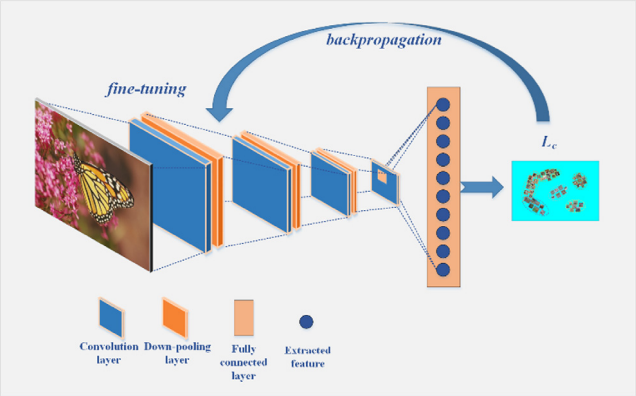

Fig. 2. The process of clustering, the result of which backpropagates to further fine-tune the encoder part of DAE.

error propagated to the jth neuron, then the error σ'_{ij} back-propagated to the ij-th neuron is calculated as:

$$\sigma'_{ij} = \sum_t r_{ij\to t}\sigma_t. \tag{5}$$

3.2. Clustering of the assessed images

One kind of soft clustering algorithms, fuzzy C-mean (FCM) algorithm, was adopted as a clustering model of the assessed images. The FCM employs the membership matrix to divide each sample based on its probability of belonging to each group. The FCM algorithm can be formulated as:

$$L_c = \sum_{j=1}^{k}\sum_{i=1}^{n} u_{ij}^b \left\|x_i - \mu_j\right\|_2^2 \tag{6}$$

$$\text{subject to} \quad \sum_{j=1}^{k} u_{ij}^b = 1$$

where u is a membership matrix and μ_j is the centroid of the jth cluster. Each element u_{ij} of u denotes the extent to which the sample x_i belongs to the cluster j. b is a weighted index, which is termed as a smoothing factor. It represents the degree to which the clustering pattern is shared between fuzzy classes.

The iterative formulas of μ_j and u_{ij} can be defined as:

$$\mu_j = \frac{\sum_{i=1}^{n} u_{ij}^b x_i}{\sum_{i=1}^{n} u_{ij}^b} \tag{7}$$

$$u_{ij} = \frac{\left\|x_i - \mu_j\right\|^{-2/(b-1)}}{\sum_{s=1}^{k}\left\|x_i - \mu_s\right\|^{-2/(b-1)}} \tag{8}$$

During the process of clustering, deep feature learning should be fine-tuned carefully. The feature extracted should not only capture the structure of the original data, but also reflect the underlying pattern within image data. For high-dimensional data, feature learning can make feature clustering more accurate , simple, and ensure samples in the same group densely intra-linked while loosely inter-linked to the rest by eliminating redundant information and facilitate the discovery of data category structure.

3.3. Fitting functions of DMOS for clustering groups

For each cluster, we simulate a particular linear fitting function of DMOS with three values of each assessed image including the PSNR, SSIM, and VIF score. Each fitting function lf_i can be defined as:

$$lf_i = w_{i1}s_{i1} + w_{i2}s_{i2} + w_{i3}s_{i3} + b_i \tag{9}$$

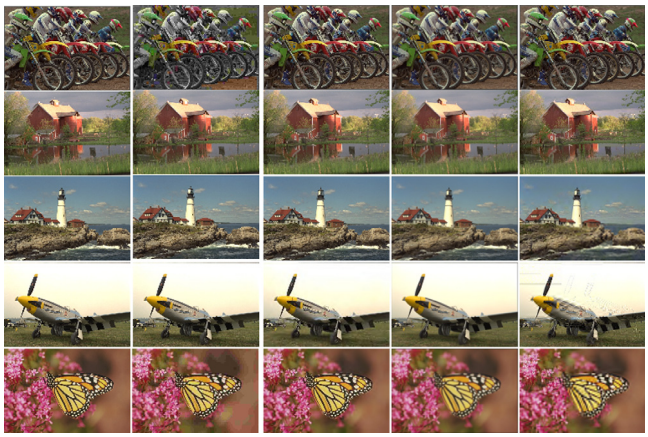

Fig. 3. Six reference images and their distorted counterparts in LIVE database. The first column shows the references while jpeg, jpeg2k, Gaussian blur, and faster Rayleigh attenuation are exhibited from the second to the fifth column.

where s_{i1}, s_{i2}, and s_{i3} represent the PSNR, SSIM, and VIF score of each image in the ith group respectively. w_{i1}, w_{i2}, w_{i3}, and b_i denote the weight coefficients and bias of f_i.

Toward a satisfactory fitting performance, we introduce polynomial fitting functions pf_i as a counterpart, defined as:

$$pf_i = w'_{i1}s_{i1}^2 + w'_{i2}s_{i2}^2 + w'_{i3}s_{i3}^2 + w'_{i4}s_{i1} + w'_{i5}s_{i2} + w'_{i6}s_{i3} + b'_i \tag{10}$$

Noticeably, in order to prevent overfitting, pf_i is only added three quadratic terms compared to lf_i.

4. Experiments and analysis

In this section, we conduct our proposed IQA algorithm and six comparative counterparts on LIVE [40], TID2008 [41] and TID2013 databases.

4.1. Image quality assessment datasets

We conduct experiments on two classic image quality database, TID2008 and LIVE, for verification.

TID2008 dataset originated from the National University of Aeronautics and Astronautics of Ukraine. It includes 25 reference images and 1700 images from 17 distorted types, including additive Gaussian noise, additive noise with color components than the lighting, the spatial location-related noise, mask noise, high frequency noise, impulse noise and quantization noise, JPEG compression, Gaussian blur, and image noise, JPEG2000 compression, JPEG transmission error, JPEG2000 transmission error, the eccentric type noise, upon local block distortion of different strength, strength the mean shift and contrast change. It is an extension version of TID2008.

The LIVE database, shown in Fig. 3, was jointly established by the Department of Electrical and Computer Engineering and the Department of Psychology at the University of Texas at Austin. The database is well-received and contains 29 reference images and 779 distorted images, including 175 JPEG2000, 169 JPEG, 145 white noise, 145 Gaussian blur, and 145 fast Rayleigh attenuation. The DMOS value of this database is obtained from approximately 25,000 data provided by 161 observers with the value range of [0, 100].

For each database, we randomly selected 80% images for training and the rest 20% are for testing.

4.2. Evaluation criterion

Two metrics were adopted to validate the effectivity of the proposed IQA method, the Pearson linear correlation coefficient (PLCC) and Kendall Rank Order Correlation Coefficient (KROCC). Herein, PLCC is defined as:

$$PLCC = \frac{n\sum_{i=1}^{n} y_i\hat{y}_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} \hat{y}_i}{\sqrt{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}\sqrt{n\sum_{i=1}^{n} \hat{y}_i^2 - (\sum_{i=1}^{n} \hat{y}_i)^2}} \tag{11}$$

where n represents the number of distorted images, y_i and $\hat{y}_i$ stand for the real value and test score of the ith image respectively.

KROCC calculates correlation coefficients to evaluate the statistical dependence between the prediction and the ground truth, which can be formulated as:

$$KROCC = \frac{2(C-D)}{n(n-1)} \tag{12}$$

where C and D in (12) represent the number of samples that are consistent and inconsistent between y and $\hat{y}$.

4.3. Implementation details

For the two data set, we first conducted pre-training on the proposed DAE structure in Fig. 1 in the way of greedy layer-by-layer pre-training as stack self-coding, since this strategy can effectively learn feature representation with respect to the original data distribution.

Iterations of each layer's pre-training repeated for 400 epochs. MBGD optimizer was utilized with an initial learning rate of 0.0001, which is multiplied by 0.1 every 100 epochs. Then, FCM clustering algorithm was adopted to choose the initially random clustering centroids after removing the decoder part of the network, leaving only the encoder and the fully connected layer. The cost function of FCM clustering constitutes the loss function to fine-tune the encoder. As the loss function is defined to minimize the distance between the target distribution and the fitting distribution, the end of the fine-tuning procedure denotes the sign of the clustering results.

The details of average clustering analysis and fitting modeling IQA algorithm are summarized as follows. First, we randomly selected 80% of each dataset for training and the other 20% for testing. Second, FCM clustering method was used to cluster the training samples into five groups, and the maximum iteration number of clustering was set to 1000. Third, each experiment was repeated five times with randomly selected training sat. The average value of the five experiments was treated as the final result.

4.4. Experimental results

In order to demonstrate the superiority of the proposed IQA method, we select six well-known image quality assessment models as comparative counterparts, including SSIM, MAD, VSNR, IFC, GSM, and VIF.

From experimental results of seven clustering algorithms on LIVE, TID2008, TID 2013 databases shown in Figs. 4, 5, and 6, the proposed IQA method achieves better results on the PLCC and KROCC measurements than the other six classic methods. (see Tables 1–3.)

The average experimental results of PLCC of seven IQA methods on the different distortion types of images in the LIVE database are shown in Tab 1,2 and 3. It can be seen from the table that our image clustering quality assessment method achieves better results than the others on four kinds of distorted images. In particular, Gaussian blur and fast Rayleigh attenuation are two difficult kinds for IQA methods to achieve satisfying performance. The performance of our proposed method surpassed all the other by a large margin, demonstrating our method is capable of choosing adequate IQA methods in different images.

It is worth emphasizing that, the proposed IQA method with the polynomial fitting function is inferior to that with the linear function,

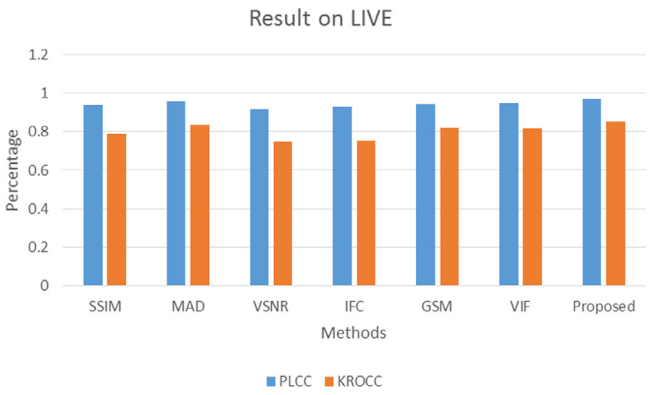

Fig. 4. The average results of PLCC, KROCC metric of seven methods in the LIVE database.

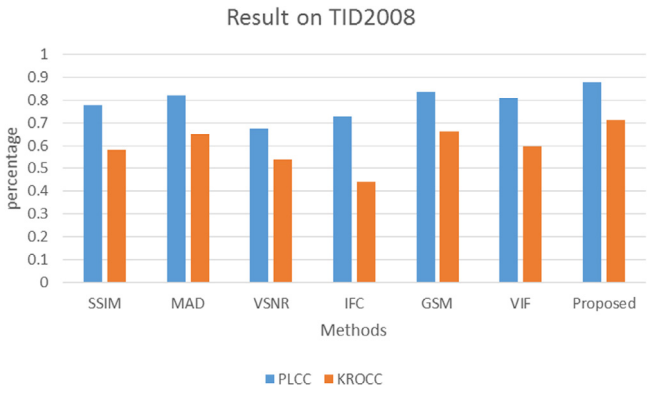

Fig. 5. The average results of PLCC, KROCC metric of seven methods in the TID2008 database.

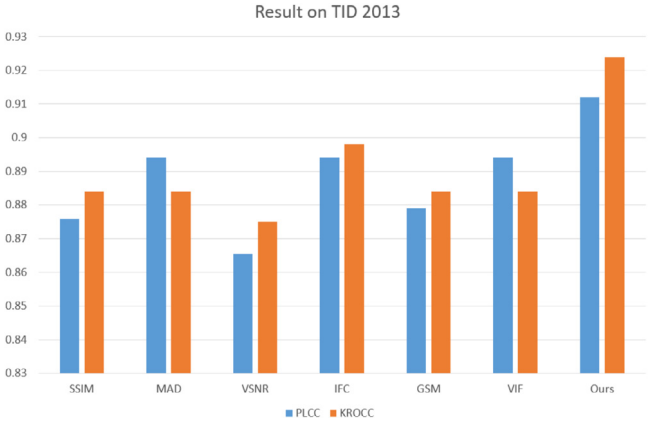

Fig. 6. The average results of PLCC, KROCC metric of seven methods in the TID2013 database.

Table 1
The average values of PLCC of seven IQA methods on the different distortion type of images in the LIVE database.

Methods	JPEG	JP2K	Gblur	FF	WN
SSIM	0.940	0.941	0.914	0.921	0.975
MAD	0.954	0.963	0.923	0.918	**0.980**
VSNR	0.894	0.907	0.823	0.864	0.928
IFC	0.930	0.927	0.865	0.875	0.949
GSM	0.957	0.948	0.927	0.883	0.937
VIF	0.947	0.939	0.926	0.878	0.977
Proposed	**0.965**	**0.970**	**0.953**	**0.942**	0.979

Table 2
The average values of PLCC of seven IQA methods on the different distortion type of images in the TID 2008 database.

Methods	JPEG	JP2K	Gblur	FF	WN
SSIM	0.912	0.911	0.906	0.911	0.908
MAD	0.912	0.906	0.915	0.921	0.922
VSNR	0.924	0.919	0.931	0.915	0.918
IFC	0.921	0.919	0.917	0.921	0.919
GSM	0.921	0.916	0.931	0.932	0.924
VIF	0.911	0.909	0.915	0.916	0.917
Proposed	**0.932**	**0.943**	**0.937**	**0.954**	**0.962**

Table 3
The average values of PLCC of seven IQA methods on the different distortion type of images in the TID 2013 database.

Methods	JPEG	JP2K	Gblur	FF	WN
SSIM	0.867	0.869	0.877	0.881	0.892
MAD	0.890	0.869	0.874	0.893	0.894
VSNR	0.899	0.903	0.896	0.897	0.903
IFC	0.895	0.897	0.904	0.903	0.908
GSM	0.901	0.902	0.897	0.899	0.904
VIF	0.906	0.903	0.907	0.896	0.898
Proposed	**0.921**	**0.917**	**0.923**	**0.931**	**0.914**

although the former had a better fitting performance during training. The reason is that, the lack of training samples results in the overfitting of polynomial fitting function.

In total, the impressive performance of our deep quality evaluation model is due to the following attributes: (1) a combination of deep features and a set of high-level quality-aware visual features is highly representative to reflect image quality. The DAE deep model is tailored for modeling image quality, which is fast to train and informative to quality-related visual elements; (2) our adopted clustering scheme treats images with different styles separately. This is the key to the effectiveness of our method. In practice, images with different styles have apparently different visual appearance and semantics. Toward a descriptive quality-aware feature, it is necessary to handle each type of images separately, i.e., each image type corresponds to a learned quality model; and (3) the fitting function can integrate the advantages of three different IQA algorithms with different assigned weights. Such multiple channel integration is very popular and effective to enhance the representativeness of visual features (such as the multi-view learning and multiple kernel learning), since the advantages of multiple models are encoded.

5. Conclusion

The topic of image quality evaluation is a useful and challenging in the field of image processing. In order to solve the problem that each existing image quality evaluation method is sensitive to images with specific characteristics, we proposed a deep clustering-based ensemble approach for quality assessment on a diverse set of images. The assessment procedure includes high-level feature extraction, clustering analysis and synthesis of multi-mode image quality evaluations. Comprehensive experimental results on LIVE, TID2008, and TID2013 databases have demonstrated that the proposed model has better image quality evaluation performance than the other six wide-spread IQA methods.

References

[1] I. Avcibaş, B. Sankur, K. Sayood, Statistical evaluation of image quality measures, J. Electron. Imaging 11 (2) (2002) 206–223.

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

[3] Z. Wang, A.C. Bovik, B.L. Evan, Blind measurement of blocking artifacts in images, in: Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101), vol. 3, IEEE, 2000, pp. 981–984.

[4] L. Guiling, W. Nannan, Z. Qiang, Study of moving image quality evaluation base on MPEG-2 system, J. Tianjin Univ. 34 (5) (2001) 573–576.

[5] F. Lukas, Z. Budrikis, Picture quality prediction based on a visual model, IEEE Trans. Commun. 30 (7) (1982) 1679–1692.

[6] P. Vincent, H. Larochelle, Y. Bengio, P.A. Manzagol, Extracting and composing robust features with denoising autoencoders, in: Proceedings of the 25th International Conference on Machine Learning, ACM, 2008, pp. 1096–1103.

[7] H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in: Proceedings of the 26th Annual International Conference on Machine Learning, ACM, 2009, pp. 609–616.

[8] H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, Unsupervised learning of hierarchical representations with convolutional deep belief networks, Commun. ACM 54 (10) (2011) 95–103.

[9] H. Lee, P. Pham, Y. Largman, A.Y. Ng, Unsupervised feature learning for audio classification using convolutional deep belief networks, in: Advances in Neural Information Processing Systems, 2009, pp. 1096–1104.

[10] G.B. Huang, H. Lee, E. Learned-Miller, Learning hierarchical representations for face verification with convolutional deep belief networks, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 2518–2525.

[11] D. Cireşan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, 2012, arXiv preprint arXiv:12022745.

[12] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.

[13] Grother P. J., NIST Special Database 19, Handprinted forms and characters database, National Institute of Standards and Technology, 1995.

[14] C.L. Liu, F. Yin, D.H. Wang, Q.F. Wang, Chinese handwriting recognition contest 2010, in: 2010 Chinese Conference on Pattern Recognition, CCPR, IEEE, 2010, pp. 1–5.

[15] Y. LeCun, F.J. Huang, L. Bottou, Learning methods for generic object recognition with invariance to pose and lighting, in: CVPR, vol. 2, 2004, pp. 97–104.

[16] J. Stallkamp, M. Schlipsing, J. Salmen, C. Igel, The German traffic sign recognition benchmark: A multi-class classification competition, in: IJCNN, vol. 6, 7, 2011.

[17] A. Krizhevsky, G. Hinton, Learning multiple layers of features from tiny images (vol. 1, no. 4), Technical report, University of Toronto, 2009, p. 7.

[18] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, Overfeat: Integrated recognition, localization and detection using convolutional networks, 2013, arXiv preprint arXiv:13126229.

[19] P.S. Bradley, U.M. Fayyad, Refining initial points for k-means clustering., in: ICML, vol. 98, 1998, pp. 91–99.

[20] P.S. Bradley, U.M. Fayyad, Refining initial points for k-means clustering, in: ICML, vol. 98, 1998, pp. 91–99.

[21] Zhang B., Generalized k-harmonic means–dynamic weighting of data in unsupervised learning, in: Proceedings of the 2001 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics, 2001, pp. 1–13.

[22] D. Pelleg, A.W. Moore, X-means: extending k-means with efficient estimation of the number of clusters, in: ICML, vol. 1, 2000, pp. 727–734.

[23] I. Sarafis, A.M.S. Zalzala, P.W. Trinder, A genetic rule-based data clustering toolkit, CEC'02 (Cat. No. 02TH8600), in: Proceedings of the 2002 Congress on Evolutionary Computation, vol. 2, IEEE, 2002, pp. 1238–1243.

[24] Dunn J. C., Well-separated clusters and optimal fuzzy partitions, J. Cybern. 4 (1) (1974) 95–104.

[25] M. Ester, H.P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Kdd, vol. 96, (34) 1996, pp. 226–231.

[26] J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map, IEEE Trans. Neural Netw. 11 (3) (2000) 586–600.

[27] X. Guo, X. Liu, E. Zhu, J. Yin, Deep clustering with convolutional autoencoders, in: International Conference on Neural Information Processing, Springer, Cham, 2017, pp. 373–382.

[28] X. Guo, L. Gao, X. Liu, J. Yin, Improved deep embedded clustering with local structure preservation, in: IJCAI, 2017, pp. 1753–1759.

[29] D.M. Chandler, S.S. Hemami, VSNR: A wavelet-based visual signal-to-noise ratio for natural images, IEEE Trans. Image Process. 16 (9) (2007) 2284–2298.

[30] Z. Wang, A.C. Bovik, B.L. Evan, Blind measurement of blocking artifacts in images, in: Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101), vol. 3, IEEE, 2000, pp. 981–984.

[31] H.R. Sheikh, A.C. Bovik, G. De Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, IEEE Trans. Image Process. 14 (12) (2005) 2117–2128.

[32] M. Miyahara, K. Kotani, V.R. Algazi, Objective picture quality scale (PQS) for image coding, IEEE Trans. Commun. 46 (9) (1998) 1215–1226.

[33] M.J. Wainwright, E.P. Simoncelli, Scale mixtures of gaussians and the statistics of natural images, in: Advances in Neural Information Processing Systems, 2000, pp. 855–861.

[34] Wang. Tisheng, Gao. Xinbo, Lu. Wen, A New Type of Reduced Reference Image Quality Assessment, vol. 35, (l) Xi'an: Xi'an university of electronic science and technology, 2008, pp. 20–36 (in Chinese).

[35] W. Lin, M. Narwaria, Perceptual image quality assessment: recent progress and trends, in: Visual Communications and Image Processing, vol. 7744, International Society for Optics and Photonics, 2010, 774403.
[36] I. Avcibaş, B. Sankur, K. Sayood, Statistical evaluation of image quality measures, J. Electron. Imaging 11 (2) (2002) 206–223.
[37] W. Hou, X. Gao, D. Tao, X. Li, Blind image quality assessment via deep learning, IEEE Trans. Neural Netw. Learn. Syst. 26 (6) (2017) 1275–1286.
[38] A. Ahar, A. Barri, P. Schelkens, From sparse coding significance to perceptual quality: a new approach for image quality assessment, IEEE Trans. Image Process. (2018) 1, PP(99).
[39] S. Bosse, D. Maniry, K.R. Muller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment, IEEE Trans. Image Process. (2017) 1, PP(99.
[40] K. Seshadrinathan, R. Soundararajan, A.C. Bovik, L.K. Cormack, Study of subjective and objective quality assessment of video, IEEE Trans. Image Process. 19 (6) (2010) 1427–1441.
[41] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, TID2008-a database for evaluation of full-reference visual quality assessment metrics, Adv. Mod. Radioelectron. 10 (4) (2009) 30–45.

Tianliang Bian obtained his B.Sc. and M.Sc. degrees in Communication Engineering from Jilin University (China) in 2013 and 2016. Now, he works in Intelligent Risk Control Department, Qianbao Financial, his research focuses on the application of artificial intelligence algorithms in the field of financial anti-fraud.