

Blind Image Quality Assessment Bases On Natural Scene Statistics And Deep Learning

De Ge^{1,a}, Jianxin Song^{1,b}

¹Nanjing University. of Posts & Telecommunications, Nanjing 210003, China;

^awww.gedenet@163.com, ^bsongjx@njupt.edu.cn

Keywords: Blind/No-Reference, Natural Scene Statistics(NSS), Deep Belief Network (DBN), Image Quality Assessment(IQA)

Abstract. Measurement of the image and video quality is crucial for many aspects, such as transmission, compression, perception. The most of traditional methods learning-based image quality assessment(IQA) build the mapping function of the distortion and mass fraction. However, the mapping function is hard to built, and not accurate enough to show the relationship between the linguistic description and numerical number. In this paper, we proposed a new framework to blindly evaluate the quality of an image by learning the regular pattern from natural scene statistics (NSS). Our framework consists of two stages. Firstly, the distortion image is presented by NSS. The Deep Belief Network (DBNs) is used to classify the NSS features to several distortion types. Secondly, a newly qualitative quality pool is proposed according to the distortion types, which converts the distortion types of the image and the degree of the distortion into the numerical scores. In this paper, the proposed distortion classification method is not only more natural than the regression-based, but also more accurate. The experience is conducted on the LIVE image quality assessment database. Extensive studies confirm the effectiveness and robustness of our framework.

1. Introduction

Precise measurement the quality of the image or video is fundamental importance to many aspects (such as acquisition, compression, transmission, enhancement, and reproduction). The most accurate measurement is to get the opinion from human observers. However, this subjective methods are not suited for many application because its time-consuming and labor-consuming characters.

According to the accessibility of the original image, the existing IQA methods can be further classified into three categories [1]: full-reference (FR), reference-reference (RR) and no-reference (NR) or blind IQA. The FR get a high performance because of accessing to the original resource. Many researchers proposed lots of algorithms to predict the quality of the image, such as PSNR [2], SSIM [3], VSNR [4] and VIF. All these algorithms have some limitations, which strictly limits the extension. Therefore, the research of image quality evaluation is a hot spot in blind IQA. But it has a long way to go before the blind IQA reach a relatively high altitude.

The first type is easy to understand. We use a specific method to predict the quality of image according to specific distortion. Although this algorithm reach a satisfied effect, it has some limitations. The second approach is based on the learning, does not know the type of distortion, and establish the of the learning model. We need a large number of features to extract from image to learn, this method is not enough accuracy and stability, because different characteristics in response to different types of distortion and distortion effect is different. The third method is based on the first and second methods. The relationship between the image quality and the NSS is firstly studied. Then the mapping relationship between them is established. Many classical algorithms based on machine learning are based on third methods. BIQI [5] is proposed to firstly judge the type of image distortion, and employ the distortion-specific metric to predict the quality of image. With the continuous deepening of NSS research and understanding of the semantic understanding of the image features, people do not need to know the type of image distortion, but directly establish the mapping relationship between NSS features and subjective quality. In [6] and [7], Saad et al proposed the BLIND and BLIND-II. These two methods directly use the NSS feature of the image to predict the

quality of the image. Ye et al[8] used the K-means clustering to map the feature into the numeric scores. These methods all have some defects[9], and most of them use machine learning method to establish the mapping function between the mass fraction and NSS characteristics, but there is no way to guarantee the accuracy of the mapping function. Secondly, it is very difficult to guarantee the accuracy of the classification accuracy by using the low level features of the image. This paper is to do further research on the issue, using qualitative tags to develop the deep learning network. Experimental results show that the method has high accuracy, and the final prediction quality and the human visual system has been kept high consistency

The contribution of our approach are as followed: 1) our approach is non-distortion-specific. Many other classical methods are required to ensure that the distortion type is known. The extension of the method of this factor. 2) we employ the DBNs to judge the type of distortion of the image. Compared to machine learning method, we extract a deeper feature of NSS. Thus, it is more accurate to judge and predict the quality of the image. 3) we establish a quality pool according to the type of distortion of the image.

The rest of this paper is organized as follow. In Section 2, we describe the DBNs and the Wavelet domain NSS feature. In Section 3, we show how to use the DBNs to judge the type of distortion of the image based on the Wavelet domain-domain NSS features is known. Besides, we establish a quality pool depend on the type of distortion of image, which converts the type of distortion into numeric scores. In Section 4, the experience is conducted on the LIVE image quality assessment database, and we present the result. We conclude in Section 5.

2. Deep learning and natural scene statistics

2.1 Deep learning

In most of the classification, we use linear or logical regressions to do a simple classification. However, the limitation of this method is only applicable to the maximum input is simple, the more input features, the classification results is not accurate. Because of this case, it is not suited to classify the type of distortion of image using the linear or logical regressions methods. In order to solve this problem, many researchers have proposed various methods, but still can not avoid the instability problem in dealing with higher order features. At the same time, the neural network near the line of sight of people. Neural network can include several layers, each layer has the input and output, and can be a good simulation of the nonlinear process. The researchers have verified that the neural network has a good learning ability, and can complete the classification process well[10]. Several key issues need to be solved in the process of completing the classification. 1) it is hard to train the multi-layer neural network. 2) the multi-layer network are prone to gradient diffusion problems.

DBNs are composed of several layers of Restricted Boltzmann Machines. It consists of visible and hidden layers. Visible layer is used to show the low-order feature, and the hidden layer used to show the high-order feature. In this paper, our classification structure based on DBNs is as followed in Figure. 1.

Figure. 3 shows the histogram of normalized coefficient for natural image and its distorted versions in Figure. 2. We can find that it is obvious to distinguish different distortion types according to the wavelet coefficients.

2.2 Natural Scene Statistics

Our approach depends on the IQA algorithm learning how the image distortion affects the NSS. The approach is trained using features derived directly from natural image of wavelet domain coefficients against various perceptual level of image distortion. Our proposed model is based on different distortion types for NSS effects that are different. We judge the type of distortion of the image by the NSS wavelet domain of the image. The Fig. 2 show the natural and suffered different type of distortion and the histogram of normalized coefficient for natural image and its distorted versions.

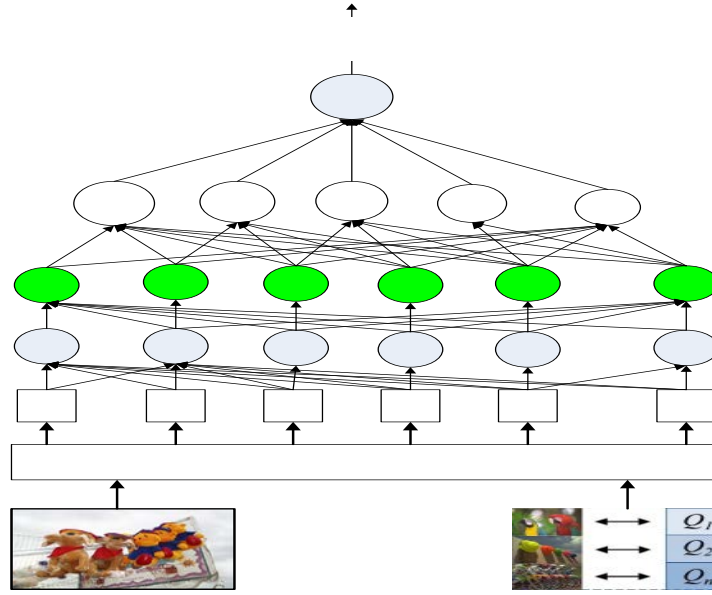


Fig. 1 The framework of our Classifier. Deep model is pre-trained by DBNs. The classification results represent Blur, WN, JPEG, JPEG200, Fastfading

3. BIQA Via Deep Learning

3.1 Framework Of Algorithm

The steps of the algorithm are as follows. Given an input image, we extract the NSS features of the image wavelet domain, and judge the type of the image by DBNs network. The distortion types considered in this paper include Blur, JPEG, JPEG2000, WN, Fast Fading, and their mutual combination. We denote the probability of each distortion as $p_i \{i=1, \dots, 5\}$. Under the premise of obtaining the image distortion type, quality score of each type of distortion is obtained by the quality pool donated as $q_i \{i=1, \dots, 5\}$. Finally, the final quality of the image is obtained by the weighted sum. Let $Q = \sum_{i=1}^5 p_i \cdot q_i$ as the quality of the distortion image. The end of this experiment has been saved through the Spearman rank order correlation coefficient (SROCC), linear correlation coefficient (LCC) and root mean squared error (RMSE) to verify. The Fig. 3 shows the framework of our algorithm in quality of the image.

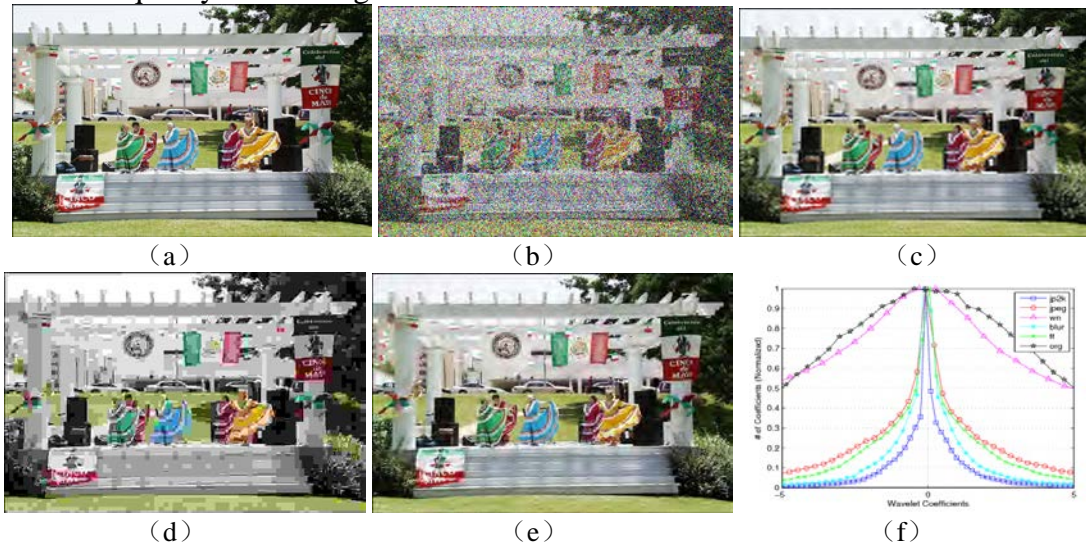


Fig. 2. LIVE image library dancers picture, (a) original image, (b) WN, (c) Fast fading, (d) JPEG, (e) JPEG 2000, (f) Histogram Of Normalized Coefficient

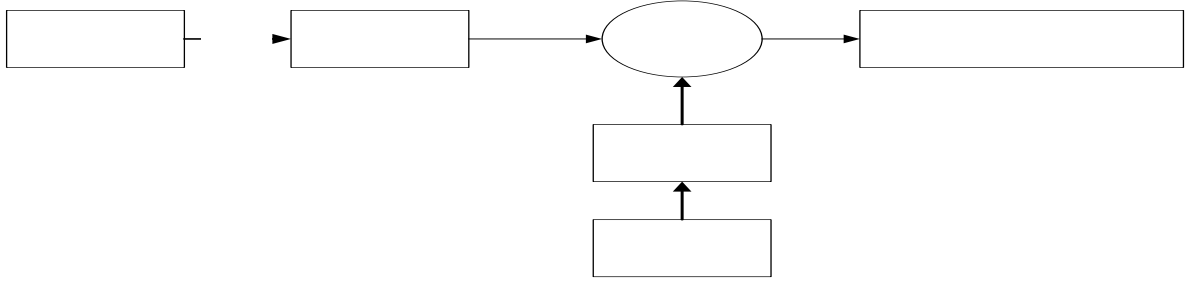


Fig. 3 The framework of our algorithm

After the introduction of the framework of the algorithm, we need to complete the following aspects.(1)How to extract the wavelet coefficient to judge the type of image distortion,How to extract the wavelet coefficient to judge the type of image distortion.(2)How to build quality pool.

3.2 Image Representation

The NSS characteristics can be divided into three levels:primary,secondary and tertiary.They have different applications in different directions, such as the primary attribute can give significant local characteristics of natural images, including resolution.The secondary attribute is mainly affected by the distortion of the image and the degree of distortion. But it is only through the two time that the property can not accurately get the image distortion type, because it does not have universal significance.The tertiary shows the self-similar property of scenes.It can reach a very high degree of accurate in the prediction of image distortion type combining the above two attributes.

Firstly, the input image is carried out in three directions by wavelet transform.Nine sub bands can be got by wavelet transform.Since low-high and high-low have similar attribute, we use only one of them.Therefore,we calculate the features in six sub-bands.

In each sub-bands,Since a and B have similar usage, we use only one of them.We calculate the magnitude m_k and entropy e_k of wavelet coefficients according to the following two equations[11]:

$$m_k = \frac{1}{N_k \times M_k} \sum_{j=1}^{N_k} \sum_{i=1}^{M_k} \log_2 |C_k(i, j)| \quad (1)$$

$$e_k = \sum_{j=1}^{N_k} \sum_{i=1}^{M_k} p[C_k(i, j)] \ln p[C_k(i, j)] \quad (2)$$

Where N_k and M_k are the length and width of the k -th sub-bands. $C_k(i, j)$ stands for the (i, j) coefficient of the k -th sub-band, and $p[.]$ is the probability density function of the sub-band.Therefore the distortion image is represented by the six sub-bands as a vector:

$$V = [m_1, m_2, \dots, m_6, e_1, e_2, \dots, e_6]^T \quad (3)$$

As shown in Fig. 1 ,the image represented by V ,and the L layer is filled with the wavelet NSS.The next two layers are simply and complexly respectively.The three two hidden layers is as followed:

$$P(V, h^1, h^2, L) = P(V|h^1)P(h^1|h^2)P(h^2, L) \quad (4)$$

Before use the DBNs,we should training the framework,there are two steps:

1)Phase 1: Pre-training parameters for the two adjacent layers using restricted Boltzmann machine (RBM).

2) Phase 2: Fine-tuning all the parameters by back propagation.

3.3 Quality Pooling

After we get the graphics distortion type,we need to establish the quality pooling according to the type of distortion.Firstly,we need to set up a training base for different distortion types and different degrees of distortion.The quality pooling consists of $30 \times 5 \times 80 = 12000$ images.Specific experimental parameters are as followed in Table 1:

Table 1.Demonstrating Minimum And Maximum Parameter Values Used For Inducing Distortions,Filter Parameters σ^2 .

Distortion types	Min	Max
WN	0.001	1
Fast Fading	0.001	1
JPEG (quality parameters)	10	75
JPEG 2000(bit rate)	0.05	1.75
Gblur	0.5	1.75

4. Experiments

In order to verify the accuracy of this algorithm in image distortion categories,the experiment is conducted in LIVE image quality assessment database.We randomly select 10 different types of image.The result is showed in Fig.4.

The classification accuracy of the proposed algorithm is more than 85% based on the deep learning algorithm.Most classification accuracy is more accurate than the BIQI algorithm. However,we find that there is still a lot of room for improvement in this paper.In the case of more than three types of distortion, the distortion of the type of a certain deviation.

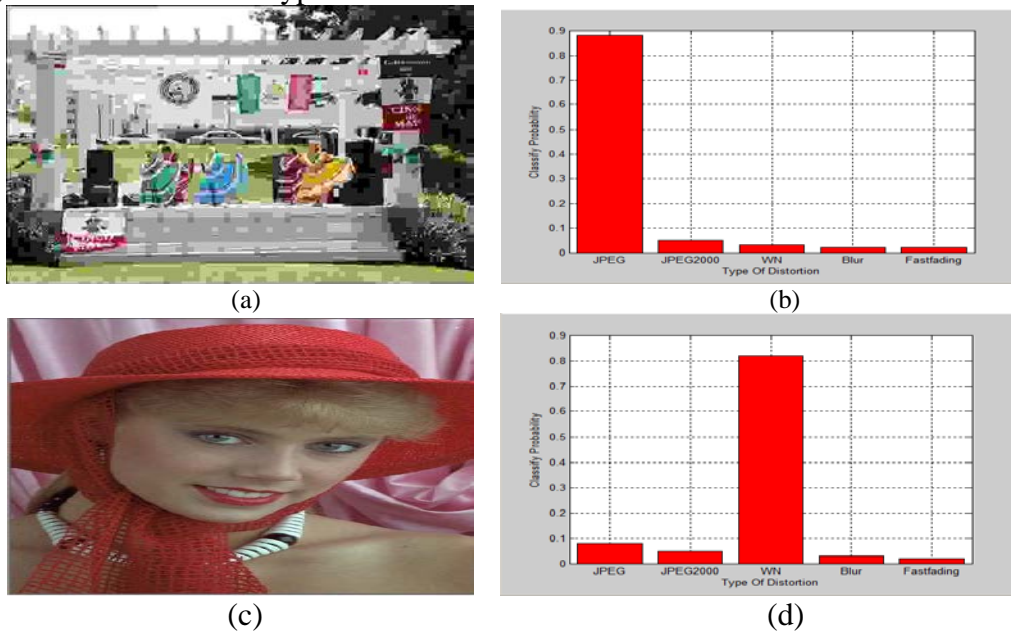


Fig. 4(a)JPEG distortion (b)accurate of prediction (c)WN distortion (d)accurate of prediction

Table 2 shows the median value of Spearman's rank ordered correlation coefficient (SROCC).In addition, the results of this paper are compared with the BIQI and the PSNR.We find that our algorithm has a higher accuracy than the BIQI algorithm.Our algorithm is consistent with Human Visual System(HVS).Compared to PSNR, the algorithm also has some competitiveness.

Table 2 SROCC between PSNR,BIQI and Our algorithm

Algorithm	WN	Fastfading	JPEG	JPEG2000	Gblur	Combine
PSNR	0.94	0.86	0.87	0.88	0.73	0.855
BIQI	0.951	0.73	0.88	0.80	0.84	0.82
Our Algorithm	0.96	0.81	0.83	0.85	0.86	0.85

We also calculate the LCC and RMSE between the PSNR,BIQI.Table 3 lists the LCC between PSNR and BIQI.The Table 4 lists the RMSE.Again,the performance of our algorithm is better than BIQI,and at most cases,better than PSNR.

Table 3 LCC between algorithm and DMOS

Algorithm	WN	Fastfading	JPEG	JPEG2000	Gblur	Combine
PSNR	0.92	0.86	0.88	0.86	0.75	0.85
BIQI	0.95	0.75	0.91	0.81	0.83	0.82
Our Algorithm	0.96	0.80	0.93	0.82	0.84	0.86

TABLE 4 .RMSE Between algorithm and DMOS

Algorithm	WN	Fastfading	JPEG	JPEG2000	Gblur	Combine
PSNR	0.92	0.86	0.88	0.86	0.75	0.85
BIQI	0.95	0.75	0.91	0.81	0.83	0.82
Our Algorithm	0.96	0.80	0.93	0.82	0.84	0.86

5. Conclusion

In this paper, We proposed a blind image quality evaluation algorithm based on the deep learning and NSS characteristics. A deep learning network is designed to classify the type of distortion of the image. Compared with other methods, it has higher accuracy. In order to make the quality of the image more accurate, we established the quality database based on the LIVE database. The effectiveness and robustness of the proposed algorithm are verified by experiments. Future research directions are how to predict the quality of mixed distortion.

References

- [1] Anish M, Anush Krishna M, Alan Conrad B. No-Reference Image Quality Assessment in the Spatial Domain.[J]. IEEE Transactions on Image Processing, 2012, 21(12):4695 - 4708.
- [2] Z. Wang, A.C.Bovik, H.R.Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600 - 612, Apr. 2004.
- [3] Z. Wang, E. P. Simoncelli, and A.C.Bovik, "Multiscale structural similarity image quality assessment," in Proc.37th Asilomar Conf.Signals Syst. Comput., Nov. 2003, pp. 1398-1402.
- [4] D.M. Chandler and S.S.Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," IEEE Trans. Image Process., vol.16, no.9, pp. 2284 - 2298, Sep. 2007.
- [5] Moorthy A K, Bovik A C. A Two-Step Framework for Constructing Blind Image Quality Indices[J]. Signal Processing Letters IEEE, 2010, 17(5):513 - 516.
- [6] Saad M A, Bovik A C, Charrier C. A DCT Statistics-Based Blind Image Quality Index[J]. IEEE Signal Processing Letters, 2010, 17(6):583-586.
- [7] Saad M A, Bovik A C, Christophe C. Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain.[J]. IEEE Transactions on Image Processing, 2012, 21(8):3339 - 3352.
- [8] Ye P, Kumar J, Kang L, et al. Unsupervised feature learning framework for no-reference image quality assessment[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2012:1098-1105.
- [9] H.Tang, N.Joshi, and A.Kapoor, "Learning a blind measure of perceptual image quality," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2011, pp. 305 - 312.
- [10] Hou W, Gao X, Tao D, et al. Blind Image Quality Assessment via Deep Learning[J]. IEEE Transactions on Neural Networks & Learning Systems, 2015, 26(6):1275-1286.

[11] Y. Xia, C. Sun, and W. X. Zheng, "Discrete-time neural network for fast solving large linear L1 estimation problems and its application to image restoration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 5, pp. 812 – 820, May 2012