Overview and research progress of no reference image quality evaluation methods

HaoYue Wang, XiangNing Chen*, ShuHan Du, BiJie Xu, Yan Liu
Space Engineering University, Beijing Huairou 101416
Email: xn_chen_edu@163.com

Abstract. Image quality will greatly affect the acquisition of human visual information, but the image will inevitably cause a certain degree of distortion in the process of transmission, compression, coding and so on. At this time, the accuracy of image quality evaluation (IQA) is particularly important, which also makes IQA an important research direction in the field of image processing. IQA methods are divided into subjective image quality evaluation method and objective image quality evaluation method. There are many subjective evaluation methods, and the international telecommunications union (ITU) has also proposed many subjective evaluation methods and standards. There are three most commonly used methods, namely, the double stimulation continuous quality classification method, the single stimulation continuous quality classification method and the double stimulation damage classification method. Correspondingly, there are three kinds of objective evaluation methods, full reference, half reference and no reference. At present, this paper selects the research focus of IQA field—no reference image quality evaluation method as the main research object, summarizes the advantages and disadvantages of each non-reference evaluation method, and prospects the future development of the non-reference image quality evaluation method.

Keywords: image quality evaluation criteria; no reference; representative algorithm

1. Introduction
As an important source of information for human information acquisition and machine recognition, the accuracy and importance of the information obtained by sharpness and SNR are important. However, distortion is unavoidable in the process of image acquisition, processing, compression, transmission and display. IQA is a technology that analyzes image features, then evaluates image quality, and finally realizes image optimization. IQA occupies an extremely important position in the field of image processing, and an effective IQA system is indispensable to evaluate whether the image meets the requirements of a specific application. Nowadays, the non-reference image quality evaluation method (NR) has become a hot topic in the field of IQA because it does not need to use reference image, has high practicability and wide application range. The following will focus on several popular and superior performance non-reference image quality evaluation methods.

2. Tradeoff Technical Index of Image Quality Evaluation Algorithm
With the increase of IQA algorithms year by year, each algorithm has put forward its own innovation and superiority, but the advantages and disadvantages of these algorithms also need certain measurement standards. This section will focus on some commonly used technical indicators to evaluate the advantages and disadvantages of IQA algorithms, as well as common public IQA databases[1].

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
2.1. No reference
Blind image quality (BIQ) evaluation method is also called blind image quality evaluation method. It is divided into specific distortion type and non-specific distortion type. Compared with the other two methods, the non-reference method is more practical and widely used. However, reference image is not used in the evaluation process, and the image content and distortion types are complicated, which makes the evaluation algorithm of non-reference image quality most difficult. The following will introduce several commonly used referenceless image quality evaluation criteria.

2.1.1. A Support Vector Machine (SVM)-based Non-reference Image Quality Evaluation Method.

1) blind image quality indices (BIQI) and image authenticity and integrity evaluation based on distortion recognition (DIIIVNE)

Moorthy and other researchers believe that the distortion of the image will affect the natural statistical characteristics of its wavelet domain. By studying the laws of this kind of influence, a non-reference quality evaluation method based on wavelet domain BIQI [3], proposed. The proposed model consists of two layers, which are described in detail below:

The first layer, for the input training image with 5 distortion types (JPEG/JPEG 2000 compression distortion, Gao Si blur, white noise, fast fading), first do 3 level 3 scale wavelet transform, extract 2 feature parameters from each of the 9 wavelet subbands obtained, then use multi-classification support vector machine (Multiclass-SVM) to train the extracted feature parameters, predict the possibility that each input image contains these 5 distortions.

At the second layer, for the above five distortion types, the feature vector is used to train the support vector regression (SVR) to obtain the mapping relationship between the feature vector and the human eye subjective evaluation of the image, and finally calculate the quality evaluation score of the image.

\[
BIQI = \sum_{i=1}^{5} p_i \cdot q_i, \quad i = 1,2,...,5
\]  

(1)

The p\(_i\) is the possibility of each distortion in the input image, q\(_i\) is the mapping relationship between the feature vector and the subjective scoring of the human eye.

After that, Moorthy et al. improved the image authenticity and integrity evaluation (DIIIVNE) algorithm based on distortion recognition [4]. And the innovation of this algorithm is to discard BIQI previous algorithm to extract only some simple edge description features when extracting eigenvalues, higher statistical features with HVS consistency were selected. The operation steps of the algorithm are as follows: the input image is decomposed by 2-level 6-direction wavelet; It was then separated and normalized, extracting its normalized wavelet coefficients statistical features (subband distribution features, inter-scale correlation, spatial correlation, directional correlation and spatial autocorrelation), this step enables the subbands of natural images to be more similar to the Gaussian distribution; Finally, The quality score of the image is obtained by using a similar calculation method to the BIQI [5].

2) Evaluation of image quality without reference in spatial domain (BRISQUE)

This method is a non-reference image quality evaluation algorithm in spatial domain. The algorithm represents the image to be evaluated as an artificially designed feature vector, Support vector machines (SVM) are then used for classification [6]. The length of the eigenvector is 36. Each image needs to be extracted twice, Extract 18 feature elements at a time, the second extraction requires a 0.5-fold scaling based on the original plot. How to extract the feature vector is as follows: extracting brightness mean contrast normalization coefficient (MSCN), Synthesizing MSCN coefficients into asymmetric generalized Gaussian distributions (AGGD), Enter into the SVM for regression, Finally, the evaluation results of image quality are obtained.
Calculation of MSCN coefficient \( \hat{I}(i, j) \) :

\[
\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C} \tag{2}
\]

\[
\mu(i, j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k, l} I_{k, l}(i, j) \tag{3}
\]

\[
\sigma(i, j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k, l} (I_{k, l}(i, j) - \mu(i, j))^2} \tag{4}
\]

where \( \mu(i, j) \) is the result after gaussian filtering and \( \sigma(i,j) \) is the standard deviation.

The calculation of generalized Gaussian distribution (GGD) is as follows:

\[
f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta \Gamma\left(\frac{1}{\alpha}\right)} \exp\left(-\left(\frac{|x|}{\beta}\right)^\alpha\right) \tag{5}
\]

\[
\beta = \sigma \sqrt{\frac{\Gamma\left(\frac{3}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)}} \tag{6}
\]

\[
\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt \quad \alpha > 0 \tag{7}
\]

For adding the correlation information between the connected pixels, the BRISQUE selects four directions to calculate the product of the adjacent elements, and the current pixel is calculated in four directions:

\[
H(i, j) = \hat{I}(i, j)\hat{I}(i, j + 1) \tag{8}
\]

\[
V(i, j) = \hat{I}(i, j)\hat{I}(i + 1, j) \tag{9}
\]

\[
D_1(i, j) = \hat{I}(i, j)\hat{I}(i + 1, j + 1) \tag{10}
\]

\[
D_2(i, j) = \hat{I}(i, j)\hat{I}(i + 1, j - 1) \tag{11}
\]

where \( H(i,j) \) is horizontal, \( V(i,j) \) is vertical, \( D_1(i,j) \) is left diagonal and \( D_2(i,j) \) is right diagonal.
Calculation of asymmetric generalized Gaussian distribution AGGD:

\[
f(x; \nu, \sigma_l^2, \sigma_r^2) = \begin{cases} 
\frac{\nu}{(\beta_l + \beta_r) \Gamma \left( \frac{1}{\nu} \right)} \exp \left( -\left( \frac{x}{\beta_l} \right)^\nu \right) & x < 0 \\
\frac{\nu}{(\beta_l + \beta_r) \Gamma \left( \frac{1}{\nu} \right)} \exp \left( -\left( \frac{x}{\beta_r} \right)^\nu \right) & x \geq 0 
\end{cases}
\]

(12)

\[
\beta_l = \sigma_l \sqrt{\frac{\Gamma \left( \frac{1}{\nu} \right)}{\Gamma \left( \frac{3}{\nu} \right)}}
\]

(13)

\[
\beta_r = \sigma_r \sqrt{\frac{\Gamma \left( \frac{1}{\nu} \right)}{\Gamma \left( \frac{3}{\nu} \right)}}
\]

(14)

The \( \nu \) is the shape of the image, \( \nu \) is the mean value of the image, and the \( \beta_l \) and \( \beta_r \) are the left variance and the right variance of the image.

Finally, using the method of machine learning, the image is first converted into feature vector, then, the feature vectors and outputs (quality scores) of all images in the training dataset are put into the SVM for training. The trained model is used to predict image quality. After pre-training, the evaluation scores of JPEG2K compression, recompression, Gaussian noise and median ambiguity can be obtained.

2.1.2. Non-reference Image Quality Evaluation Method Based on Probability Model.

1) Methods for evaluating image integrity (BLIINDS) and BLINDS-II based on statistical properties of offline cosine transform (DCT) domain.

Saad et al. proposed an image integrity evaluation method (BLIINDS) [7] based on DCT domain statistical characteristics based on the feature that HVS is more sensitive to image structure information and contrast. The specific operation steps of the evaluation method are as follows: first, the statistical features (structural features, contrast features, and anisotropy) of the DCT coefficients of the training image are extracted; then, the extracted feature parameters are fitted using the multivariate gaussian model (MGD) or the multivariate laplacian distribution model (MLD) and the probability model is obtained by training. Finally, the above feature parameters are extracted for the image to be evaluated, and the quality score is calculated by substitution model.

Subsequently, the improvement of Saad et al. was BLIINDS-II [8], on the basis of BLIINDS. The innovation of this method lies in changing the selection and fitting of characteristic parameters and making the final evaluation results closer to the subjective score of human eyes.

Finally, the GGD shape parameter characteristics, energy quantum band ratio, direction characteristics and frequency variation coefficient are selected. The BLIINDS-II is similar to the BLIINDS in the calculation method. His contribution is mainly to re-select the characteristic
parameters, which makes the evaluation results more in line with the HVS characteristics and greatly improves the effect of the algorithm.

2) Natural image quality evaluation methods (NIQE)

From the simple natural scene, the author extracts similar features from the BRISQUE and constructs a systematic image quality evaluation index. The innovation of this algorithm mainly lies in the fact that the distorted image with subjective score is not used as the training data, but only the informative image block is selected as the training data in the original image library. The quality of the measured image can be evaluated by fitting these indexes into a multivariate Gao Si model[9]. The steps are as follows:

Firstly, Formula (45) is used to measure the local sharpness of the image, and the local average variance of the image block is calculated to filter out the richer image block:

$$\delta(b) = \sum \sum_{(i,j) \in patch} \sigma(i,j)$$  \hspace{1cm} (15)

$\delta$ is a local deformation coefficient defined by the author.

Select the image blocks that meet $\delta > T$ conditions ($T$ is usually 0.6–0.9 times the maximum value of $\delta$) are screened out, and then the normalized coefficients of these image blocks are fitted by GGD model. After estimating the normalized GGD model parameters, the product of four adjacent coefficients is fitted with AGGD, and 16 adjacent coefficient parameters are obtained. By fitting the extracted characteristic parameters with MVG, the model parameters are calculated:

$$f_\chi(x_1,x_2,...,x_k) = \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^\frac{1}{2}} \exp\left(-\frac{1}{2}(x - \nu)^T \Sigma^{-1}(x - \nu)\right)$$  \hspace{1cm} (16)

The $x_i$ is the extracted feature parameter, the $\nu$ is the mean value of the MVG model, and the $\Sigma$ is the variance matrix of the MVG model.

Finally, the statistical feature parameters of the image to be evaluated are extracted, and the mean value and variance matrix are fitted by the MVG model. The quality of the image is evaluated by calculating the "distance" between the fitting parameters of the image to be measured and the training image.

$$D(\nu_1,\nu_2,\Sigma_1,\Sigma_2) = \sqrt{\left(\nu_1 - \nu_2\right)^T \left(\Sigma_1 + \Sigma_2\right)^{-1} \left(\nu_1 - \nu_2\right)}$$  \hspace{1cm} (17)

Where $\nu_1,\nu_2,\Sigma_1,\Sigma_2$ are the mean value and variance matrix of the MGD model of the training image and the image to be tested respectively.

3. Image quality evaluation method based on deep learning

In machine learning research, deep learning is a new field with very strong modeling and analysis capabilities, which can solve many practical problems and show superior performance in handling a series of tasks. In recent years, image quality evaluation methods based on deep learning have also received extensive attention. The following are several non-reference image quality evaluation methods based on machine learning [2]:

3.1. Quality evaluation of sorted images (RankIQA)

In the process of image quality evaluation, it is very difficult to obtain the absolute quality evaluation of the picture, but it is relatively easy to distinguish the one with better quality from the two pictures. Based on this, the author proposes a non-reference image quality evaluation method for depth learning.
from the relative comparison of image quality, which solves the problem of insufficient training data when using depth learning to evaluate image quality[10].

Firstly, the original image is distorted by different distortion methods combined with different distortion intensities, so a large number of images with different degrees of distortion can be obtained. Then the twin network is used for training. As a result, the model learns the difference between different pictures, without accurate labels. After the model is trained, one of the twin models is taken as the initialization model, and the IQA data set is passed into the model. Regress according to the true score of the picture and fine-tune the model parameters.

Secondly, the author also proposes a more effective backpropagation method. A disadvantage of twin networks is the large amount of computation. Under traditional methods, if you have an original image plus n distorted picture, you need to pass \( n^2 - 1 \) pictures, and this paper proposes to add a layer at the end of the network that can produce picture pairs. In essence, the quality score of all pictures in this operation is calculated and the gradient is calculated.

3.2. Depth Image Quality Appraiser (DIQA)

Since the data set of manual scoring is too small, the method, like rankIQA, has experienced two stages of training [11].

The first stage only needs to determine that the relevant information of a pair of pictures is different degrees of distortion. And the trained data need to be normalized, that is, the extraction of high frequency information. The high frequency information loss part of distorted image is extracted by depth learning.

The reason of this thought is that: first, the distortion of the image has little effect on its low frequency information; second, the human visual perception system is not very sensitive to the change of the low frequency information of the image. So the data used in the whole training process of depth learning is not the original image, but the high frequency information extracted from the image. The high frequency information of the image can be extracted by using the edge detection operator, or the indirect method of subtracting the Gao Si fuzzy graph of the original graph can be used. Figure 1 shows the flow chart of the DIQA algorithm.

Figure 1. Overall flowchart of DIQA.

The graph is a flow chart of the whole training process. The goal of the first stage has two inputs, one is an error graph, the other is a trusted graph. where the error graph is a feature graph formed by deep learning network.

The expression of the loss function is:

\[
L_1(\hat{I}_d; \theta_f, \theta_g) = \| g(f(\hat{I}_d; \theta_f) - e_{fg}; \theta_g) \|_2^2 \quad (18)
\]
\[ e_{gr} = |\hat{I}_r - \hat{I}_d|^p \]  
(19)

\[ \hat{r} = \frac{1}{H_r \cdot W_r} \sum_{(i,j)} r(i, j) \]  
(20)

\[ r = \frac{2}{1 + \exp(-\alpha \|\hat{I}_d\|)} - 1 \]  
(21)

The \(\theta\) is the CNN parameter, the \(\hat{I}_r\) is the high frequency information image of the reference image, the \(\hat{I}_d\) is the high-tech information map of the distorted image, the \(P\) is the index parameter, and the \(\hat{r}\) is the reliability map. \(r\) is a weight distribution for texture and flat parts (the weight of the high frequency part is greatly increased). Because the model inputs high frequency information of the distorted image, combined with the loss function, the coefficient can eliminate the adverse effect of the flat part on the prediction error map.

3.3. No-Reference Image Quality Assessment via Adversarial Learning (Hallucinated-IQA)

An illusion-guided mass regression network is presented. the method first generates an hallucination reference image to compensate for the lack of a true reference image [12]. then, the hallucination reference information is paired with the distorted image information and, under the guidance of implicit sorting relationships within the generator, it is submitted to the regression analyzer to learn the differences and thus produce accurate quality predictions.

The model consists of three parts: quality perception generation network, quality discrimination network and hallucination guided quality regression network. The generation network is used to generate hallucination reference images, so the smaller the gap between the generated hallucination images and the real images, the more accurate the function of the quality regression network is. To this end, we adopt a stacked hourglass as the benchmark for generating the network. Using the loss of feature space and the loss of quality perception as constraints, we propose a constraint function consisting of three parts: quality difference, semantic difference and pixel difference. Then, in order to ensure that the generator produces high-perceived outputs with real high-frequency details, especially for those samples that seriously lack structural and texture information, we introduce adversarial learning mechanisms. The training discriminator distinguishes between images that improve the quality of regression networks and images that reduce the quality of regression networks. Then, the difference graph (the absolute value of the difference between the hallucination image and the distorted image) and the distortion graph are input into the quality regression model for the training of the regression network. Finally, the quality score of the measured image is obtained by doing high-order semantic fusion.

3.4 NIMA of neuroimaging evaluation

NIMA based on the latest deep object recognition (object detection) neural network, it can predict the distribution of human evaluation opinions on images from the technical and aesthetic perspectives [13]. NIMA algorithm is to generate a score histogram for any image, that is, to score the image 1-10 points, and directly compare the image of the same subject. This design coincides with the visual characteristics of human beings in form. And the evaluation effect is closer to the result of human evaluation.

Based on the probability distribution of image quality evaluation, the EMD (earth over's distance)-based loss is calculated and backpropagated.
\[
EMD(p, \hat{p}) = \left( \frac{1}{N} \sum_{k=1}^{N} \left[ CDF_{\hat{p}}(k) - CDF_p(k) \right]^r \right)^{1/r} \tag{22}
\]

Where \( p \) is the fraction probability distribution of human evaluation picture quality, \( \hat{p} \) is the predicted fraction probability distribution; the \( N \) is the number of fractions, the \( r \) is the type of norm, and the \( CDF \) is the cumulative function of probability distribution.

The comprehensive quality evaluation of the picture to be evaluated is determined by the mean value and standard deviation of the fractional probability distribution. The mean value represents the quality score of the picture, and the standard deviation represents the unconventional degree of the picture.

4. Future Outlook

The accuracy of image quality evaluation algorithm is of great significance to the development of the whole image processing field, and the research in IQA field has been paid more and more attention by more scholars. However, there are still the following research directions:

1) Now the research based on HVS characteristics is still focused on the physiological characteristics of human eye vision, and its psychological characteristics are less involved. In the future, we need to study this aspect more deeply, combine the advanced knowledge of image analysis, and create a HVS model closer to human visual perception;

2) As for the evaluation methods, we make a clear classification of subjective evaluation and objective evaluation, and in the future research direction, we should transform the IQA algorithm from simple subjective or objective method to the combination of main and objective;

3) At present, the IQA algorithm for color image and video image is not comprehensive enough, there is still a lot of room for improvement;

4) With the development of technology, stereo image may become a mainstream image form in the future, then the IQA algorithm for stereo image should also keep pace with the times and become another important field of IQA research.

5. References

[1] ShuFang Z, A Review of general reference - free Image Quality evaluation algorithms. Computer Engineering and applications, 2015(19): p. 13-23,151.

[2] Peng L, GenQiao I, GuoLiang Z, Research on the evaluation method of image quality without reference based on information entropy and deep learning. Computer application research, 2018. 35(11): p. 3508-3512.

[3] Moorthy A K ,Bovik A C.A two-step framework for constructing blind image quality indices[J].Signal Processing Letters ,2010 ,17(5): p. 513-516.

[4] Moorthy A K ,Bovik A C.Blind image quality assessment : from natural scene statistics to perceptual quality[J]. IEEE Transactions on Image Processing ,2011 ,20(12): p. 3350-3364.

[5] Simoncelli E P ,Freeman W T ,Adelson E H ,et al.Shift able multiscale transforms[J].IEEE Transactions on Information Theory ,1992 ,38(2) :587-607.

[6] Mittal, A., A.K. Moorthy and A.C. Bovik, No -Reference Image Quality Assessment in the Spatial Domain. IEEE Transactions on Image Processing, 2012. 21(12): p. 4695-4708.

[7] Saad M A ,Bovik A C ,Charrier C.A DCT statistics -based blind image quality index[J].Signal Processing Letters , 2010 ,17(6) :583-586.

[8] Saad M A ,Bovik A C ,Charrier C.DCT statistics model based blind image quality assessment[C]/Proceedings of the 18th IEEE International Conference on Image Processing(ICIP) ,2011 : p. 3093-3096.
[9] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," IEEE Trans. Circ. Syst. Video Technol., vol. 5, no. 1, pp. 52-56, 1995.

[10] X. Liu, J. Van De Weijer and A. D. Bagdanov, "RankIQA: Learning from Rankings for No-Reference Image Quality Assessment," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 1040-1049.

[11] Kim, J., A. Nguyen and S. Lee, Deep CNN-Based Blind Image Quality Predictor. IEEE Transactions on Neural Networks and Learning Systems.

[12] 2018CVPR: Hallucinated-IQA: No-reference image quality assessment via adversarial learning

[13] H. Talebi and P. Milanfar, "NIMA: Neural Image Assessment," in IEEE Transactions on Image Processing, vol. 27, no. 8, pp. 3998-4011, Aug. 2018.