

No-Reference Image Quality Assessment via Degenerate Decision Logic

Haoyuan Deng*

The School of Electronic
Information
Wuhan University
Wuhan, China
denghaoyuan@whu.edu.cn

Zikai Wang

The School of Electronic
Information
Wuhan University
Wuhan, China
2019302120304@whu.edu.cn

Xiao An

The School of Electronic
Information
Wuhan University
Wuhan, China
2019302120305@whu.edu.cn

Guangyi Yang

The School of Electronic
Information
Wuhan University
Wuhan, China
ygy@whu.edu.cn

Abstract—In this study, a no-reference (NR) IQA method based on degenerate decision logic (DDL) is proposed. DDL uses wavelet transform to extract the wavelet sub-band features of images and then maps these features to higher-level features through a hybrid network, thereby achieving effective extraction of mixing characteristics and avoiding the redundant information of low level features. DDL also adds decision logic to the assessment, combining decision logic differentiation training and fine-tuning training to allow the assessment result to approach that of human subjective cognition as closely as possible. The performance of the proposed DDL method is evaluated on the LIVE database.

Keywords—no-reference image quality assessment, deep learning, deep neural decision forest, degenerate decision logic, feature hybrid network

I. INTRODUCTION

In general, humans tend to score images qualitatively rather than quantitatively. Classifying graphics as good, average, and bad is more consistent with subjective ratings, and these feelings do not depend on the distortion type of the image [1]. Thus, this assessment scale is suitable for different types of distortion in IQA. The main drawback of the existing NR IQA method based on machine learning (ML) is that simulation of HVS complexity comes at the price of ML model complexity. A targeted structure for restoring the perception and judgment of image degradation is lacking, resulting in the inability to interpret the model and the high computational complexity of the neural network. The number of training samples required increases dramatically and achieving an acceptable test performance is difficult. From the perspective of information theory, such a system does not effectively eliminate redundant information in natural images. Therefore, this study supposes that the IQA should satisfy the

following requirements simultaneously: measurement of the basic impression of human visual perception and measurement of the human visual perception of image details.

By simulating the human subjective assessment of image degradation, this study presents a feature extraction and image quality classification model.

The design concept of comparison of details on the basis of judging impression is determined, and an NR IQA model based on the degeneracy decision logic is established. The contributions of this study are as follows.

- An in-depth simulation of the degree of image degradation is performed. The use of graded scores instead of distortion type scores is more in line with the subjective feelings of humans; thus, the final IQA model yields results closer to those obtained from human observation and assessment of images.
- The feature extraction method is combined with the manual extraction of features and neural network. Manual extraction of features can effectively reduce the redundancy of image data. The manually extracted features are inputted into the neural network to perform deeper image feature mixing. The neural network node output as the final image feature not only ensures less data redundancy but also guarantees the depth of image representation.
- The decision-making forest is used to classify image quality. In this technique, only a few training parameters are used, the effect is stable, and the results do not easily fall into the local optimum. The contradiction between the depth of the classification function and the number of parameters processed is effectively solved.

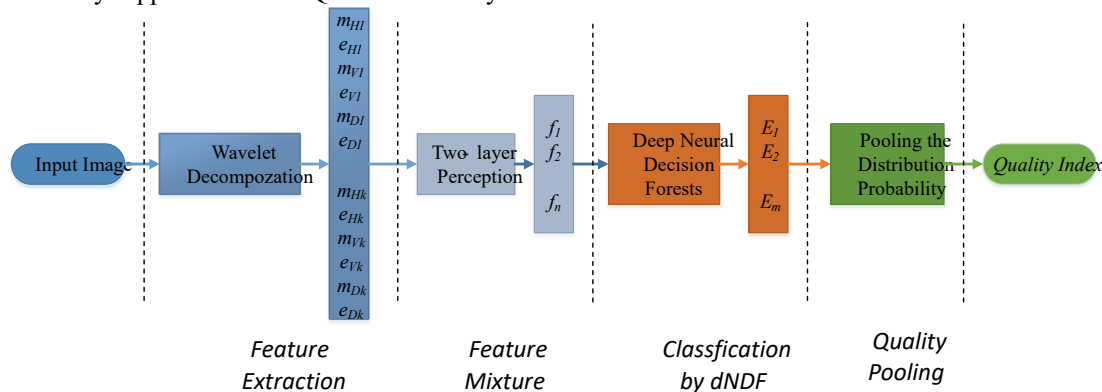


Fig. 1. Schematic of the no-reference IQA method based on degenerate decision logic.

II. NR IQA METHOD USING DEGENERATE DECISION LOGIC

This study presents a NR IQA method based on degenerate decision logic (DDL) which simulates the subjective assessment process of the HVS. The feature vector representing the degradation degree of image is obtained by initial extraction of the statistical characteristics of a natural scene and further selection and combination of the feature hybrid network. Then, the quality rating of the distorted image is obtained by classification decision logic, which indicates the assessment impression of the degradation degree of the distorted image. Finally, the objective score of the image is obtained by pooling.

Figure 1 shows the specific assessment process of the proposed DDL method.

A. Characteristics of the Wavelet Sub-Band

This study chooses a relatively concise wavelet sub-band feature [2]. The image is decomposed into K sub-bands by pyramid decomposition. Each scale contains components in the horizontal, vertical, and diagonal directions. The wavelet coefficients of these components are respectively denoted as H_k , V_k , and D_k ($k \in \{1, 2, \dots, K\}$). Given that H_k and V_k are only interchanged in the filtering order, only one of them is considered. The wavelet sub-band characteristics of natural images represent the degree of degradation well. We represent the energy information and structural information of the subband by amplitude m_{C_k} and entropy e_{C_k} , respectively [1].

$$m_{C_k} = \frac{1}{M_{C_k} N_{C_k}} \sum_{i=1}^{M_{C_k} N_{C_k}} \log_2 |C_k(i, j)| \quad (1)$$

$$e_{C_k} = - \sum_{i=1}^{M_{C_k} N_{C_k}} \sum_{j=1}^{M_{C_k} N_{C_k}} P[C_k(i, j)] \ln P[C_k(i, j)] \quad (2)$$

Where M_{C_k} and N_{C_k} are the length and width of the subband wavelet coefficient C_k , respectively, and $P(S)$ represents the occurrence probability of sample value S .

In this study, we take $K = 3$ as the number of wavelet sub-bands. Then, an image is represented as a 12-dimensional feature vector:

$$X = [m_{H_1}, m_{D_1}, m_{H_2}, m_{D_2}, m_{H_3}, m_{D_3}, e_{H_1}, e_{D_1}, e_{H_2}, e_{D_2}, e_{H_3}, e_{D_3}] \quad (3)$$

Figure 2 intuitively shows the correlation between the wavelet sub-band features and image quality. Figure 2(a) presents 29 undistorted reference images taken from the LIVE library, and Figure 2(b) shows the sub-band amplitude characteristics of the same reference images at different degradation degree intervals. Figure 2(a) indicates that, although the image contents are different, the decomposition characteristics of distortion-free images on different channels tend to be consistent. Figure 2(b) shows that the amplitude characteristics of the wavelet sub-band are obvious in differentiating the linear separability of difference mean opinion scores (DMOS). Although sub-band amplitude features may alias different image contents or degradation types, their separabilities are still preserved in large-span DMOS differences. This is the advantage of the quality level classification mechanism, i.e., it provides a theoretical basis for the use of wavelet sub-band characteristics for image feature statistics.

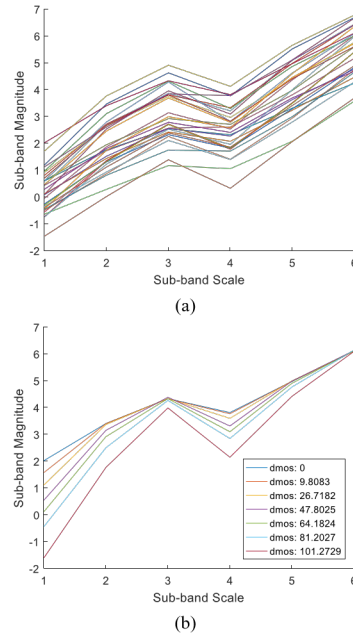


Fig. 2. Correlation between wavelet sub-band characteristics and image quality: (a) Twenty-nine non-distorted reference images from the LIVE database. (b) Sub-band amplitude characteristics of the same reference image in different degradation ranges.

B. Feature Hybrid Network

Based on the feature extraction of wavelet sub-bands, the feature mixture network is used to effectively map the wavelet feature vectors to high-level features [3] through linear combination and nonlinear mapping. As shown in Figure 3, the underlying feature vector X first passes through two layers of full-connection perceptions to implement the extraction of mixed features. The number of nodes of hidden layers H_1 and H_2 should be appropriately higher than that of input layer X , where we set $r = 50$. H_1 uses the ReLu activation function [4]. By contrast, H_2 does not use the activation function but waits for the activation of decision logic in the next stage. f is the mixed feature:

$$f = f(X; \Theta) \quad (4)$$

where Θ represents the network's pre-training parameters, including the linear mapping weights and offsets.

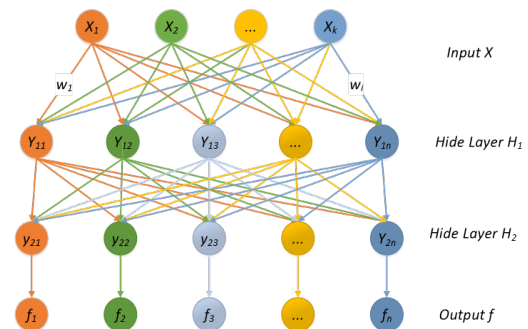


Fig. 3. Feature hybrid network.

C. Degenerate Decision Logic

In accordance with the BT.500-11 standard [5] published by the ITU-R, the absolute scale standard of manual test is divided into five levels, as shown in Table I. Among these levels, the quality and obstacle scales are the definitions of different thinking directions, which are provided to

professional and non-professional experimenters, respectively. By simulating human subjective assessment characteristics, this study adopts the same assessment level.

TABLE I. ABSOLUTE STANDARD OF SUBJECTIVE ASSESSMENT

Class	Quality Scale	Obstruct Scale
5	Great	Imperceptible
4	Good	Perceptible and unimpeded observation
3	Fair	Obstructive to observation
2	Bad	Interferes with observation
1	Poor	Seriously obstructive to observation

After the classification standard is determined, realizing image feature mapping to the classification level is the core of the entire IQA method. This study adopts a new approach to using the decision tree for image feature classification mapping. The use classification thresholds and decision paths is the fundamental difference between decision trees and neural networks. As the segmentation threshold used in the decision tree is non-derivable, the training of neural network and decision logic is cut off; thus, the classification accuracy is low and unstable. Therefore, the deep neural decision forest (dNDF) [6], the segmentation function of which is derivable, is adopted to replace the traditional decision tree [7]. The entire model can be trained in parameter fine-tuning through reverse propagation. The basic concept of dNDF is to strengthen the diversity among different decision trees by randomly arranging the input and nature of the decision tree, which helps exhaust all possible sample situations.

The dNDF has the features of the input sample vector f , and its decision tree T_i has H decision nodes. Then, according to certain mechanisms, random sampling, and arrangement of H elements in f , we form a subset of $\{f_1, f_2, \dots, f_h, \dots, f_H\}$ where f_h is an element in the subset. Obviously, f_h is "put" into the decision node h , thereby affecting the value of d_h , and d_h is the decision probability of dNDF.

$$d_h = \sigma(f_h) \quad (5)$$

Where $\sigma(x)$ is sigmoid function.

To enhance the distinction between various categories of decision logic and the consideration of parameter fine-tuning,

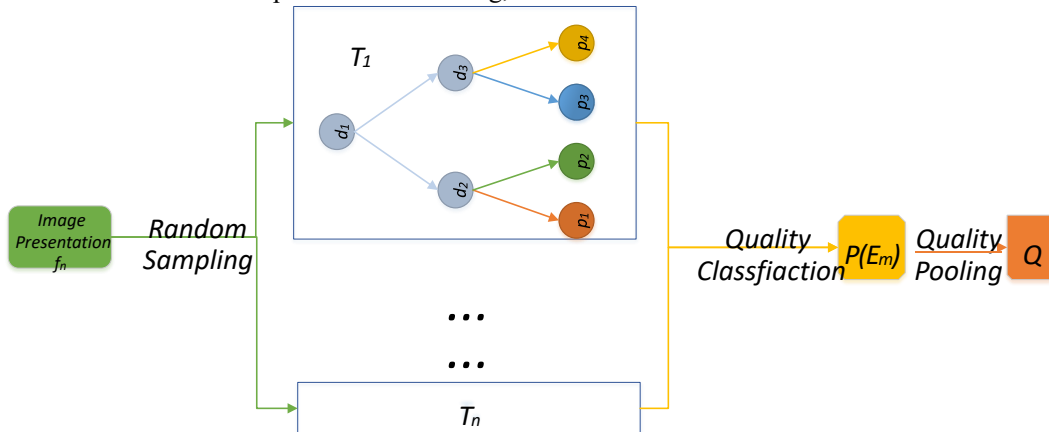


Fig. 4. Classification and pooling model based on dNDF.

Considering Sections II-C and II-D, the classification and pooling model based on dNDF is shown in Figure 4. Given that the continuous derivable sigmoid function is used to replace the non-derivable discrete function on decision

the classified probability of leaf nodes is the parameter that can be trained and updated. Update of π_n means the leaf nodes are differentiated into categories, which is a necessary implementation step for the decision logic model. Thus, the probability that sample X is mapped to the quality rating E_m after the decision tree T is obtained is as follows:

$$P_T(E_m; X) = \sum_{n=1}^N \pi_n E_m \mu_n \quad (6)$$

Where E_m corresponds to the mapping relationship established in Table I.

Set T_i ($n \in \{1, 2, \dots, F\}$) is the voting result of all decision trees forming the random forest F , and the final classification result of random forest is the voting result of all decision trees [8].

$$P_F(E_m; X) = \sum_{i=1}^F P_{T_i}(E_m; X) \quad (7)$$

D. Quality Rating Pooling

After mapping image features to classification levels through dNDF, the classification of image quality can be determined according to the size of various probabilities. However, many specific applications usually require an exact image quality score; thus, further pooling is performed in this study. From the perspective of information theory, the judging impressions established by the HVS are kept as statistically independent as possible to avoid duplication of information [9]. Therefore, we use a simple homogeneous linear pooling.

Supposing that the constants in the order of size $\{E_1, E_2, \dots, E_G\}$ divide the range of subjective scores, MOS/DMOS, evenly, then the objective IQA score is:

$$Q(x) = E[E_m] = \sum_{m=1}^G E_m P(E_m) \quad (8)$$

In this study, we set $G = 5$. We found through experiments that the values of E_m and G have no significant influence on the experimental results.

probability d_n , the decision tree of dNDF has a loss function gradient. Consequently, the DDL and the feature hybrid network establish a distinct relationship: the former and the latter respectively represent the judging impression and the

naked eye observation. When the judging impression is changed, the visual focus also changes. The observation focus and assessment impression are expected to eventually reach a balance, thus grasping the general extent of image degradation.

III. TRAINING OF THE DECISION MODEL

The present study divides the training of the decision model into two processes: decision logic differentiation training and parameter fine-tuning training. Decision logic differentiation training adopts a rough formation of judging impressions based on the data feed-forward approach. Parameter finetuning is based on the assessment of impressions to adjust the overall model. The feedback behavior of the data is strengthened in this step.

A. Differentiation Training of Decision Logic

The training set $(X, DMOS) \in T$ is defined. The sample size contained in sample T is denoted as $|T|$, the real assessment score tag of the database is DMOS, and $L(Q, DMOS; \theta, \pi)$ is the loss function of the ML model. The mean square error loss function is defined as:

$$L = \frac{1}{|T|} \sum_{(X, DMOS) \in T} (Q(X) - DMOS)^2 + \lambda \cdot L2 \quad (9)$$

Where λ is regularization rate.

To enhance the decision-making path differentiation further, we define the dNDF of each decision tree to update parameter π_n according to Eq. (10).

$$\pi_n^{(t+1)} = \sum_{(X, DMOS) \in T} R(DMOS) \cdot \frac{\pi_n^{(t)} \mu_n(f(X; \theta))}{Pr(E_m; X)} \quad (10)$$

Where $R(DMOS)$ is a measurement function representing the difference between the expected quality score and the quality rating threshold of the decision tree. $R(DMOS)$ is defined as:

$$R(DMOS) = e^{-\gamma(DMOS - E_m)^2} \quad (11)$$

Where $\gamma > 0$ and the optimum value can be determined by experiments.

To maintain the connection between decision logic and feature hybrid network, an updating function of the neural network must be introduced to the training process. Here, the random gradient descent method is used to complete the training for parameter Θ .

$$\theta^{(t+1)} = \theta^{(t)} - \eta \frac{\partial}{\partial \theta} L(Q, DMOS; \theta, \pi) \quad (12)$$

To avoid frequent decision logic updating, we use Eq. (10) to update π for one iteration after using Eq. (12) to update Θ for N iterations. In the present experiment, $N = 200$ is adopted to avoid the negative effect of frequent updating of decision logic and obtain a relatively obvious training effect of the differentiation decision logic.

B. Fine-Tuning Training

The quality rating classification of dNDF only simulates the basic impression that humans judge the degradation degree of distorted images. An insufficient fitting degree for the detailed comparison model that is dependent on the

assessment impression remains. Therefore, we further fine-tune the parameters (θ, π) involved in the decision logic through gradient descent optimization.

$$(\theta, \pi) = \operatorname{argmin}_{\frac{1}{|B|} \sum_{x, DMOS} L(x, DMOS; \theta, \pi)} \quad (13)$$

Where $B \subset T$ is a small batch subset of the training set.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

One universal IQA database, namely, LIVE [10], is used for the experiments. The experiments are conducted under the same training and testing conditions. Three assessment indicators are used to objectively assess the performance of the method, i.e., Spearman rank correlation coefficient (SROCC) [11], Pearson linear correlation coefficient (PLCC) [12], and root mean square error (RMSE) [13].

$$f(z) = \beta_1 \left[\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(z - \beta_3))} \right] + \beta_4 z + \beta_5 \quad (14)$$

Where z is the objective IQA score, $f(z)$ is the IQA regression fitting score, and β_i ($i = 1, 2, \dots, 5$) is the parameter of the regression function.

80% of the training samples and 20% of the test samples are randomly selected 100 times for training and testing. Table II lists the performance of proposed DDL method on the LIVE database. The IQA indicator of the DDL method achieves very good results on the LIVE database.

TABLE II. MEDIAN CROSS-VALIDATION METRICS FOR THE LIVE DATABASE

Metrics	SROCC	PLCC	RMSE
PSNR	0.866	0.862	13.89
SSIM	0.913	0.906	11.56
VIF	0.952	0.952	4.918
BIQI	0.824	0.833	15.05
DIIVINE	0.916	0.917	10.90
BLIINDS-II	0.933	0.924	9.047
BRISQUE	0.947	0.937	8.330
SSEQ	0.935	0.938	8.004
DLIQA	0.929	0.934	8.149
DDL	0.956	0.96	7.497

To further verify the performance of the proposed DDL method, Table III lists the SROCC values of different distortion types on the LIVE database of DDL and nine other state-of-the-art IQA methods.

TABLE III. CROSS-VALIDATION SROCC MEDIAN FOR LIVE DATABASE SAMPLES OF SPECIFIC DISTORTION TYPES

Metrics	JP2K	JPEG	Noise	Blur	FF	All
PSNR	0.868	0.885	0.943	0.761	0.875	0.866
SSIM	0.938	0.947	0.964	0.907	0.94	0.913
VIF	0.952	0.91	0.984	0.972	0.963	0.952
BIQI	0.802	0.874	0.958	0.821	0.73	0.824
DIIVINE	0.913	0.91	0.984	0.921	0.863	0.916
BLIINDS-II	0.946	0.935	0.963	0.934	0.899	0.933
BRISQUE	0.947	0.925	0.989	0.951	0.903	0.947
SSEQ	0.946	0.951	0.978	0.948	0.904	0.935
DLIQA	0.933	0.914	0.968	0.947	0.857	0.929
DDL	0.954	0.968	0.974	0.938	0.906	0.956

V. CONCLUSION AND FUTURE WORK

This study targets the process of restoring human perception and judging image degradation. Using neural

networks and NDFs that are relatively simple in computation process, a set of NR IQA methods based on DDL is implemented. First, the decision logic represented by dNDF is replaced by image quality rating, which effectively simulates human assessment impressions. Second, model training is divided into two parts, namely, decision logic differentiation and parameter fine-tuning, which correspond to the assessment process from assessment impression to detailed comparison. Finally, the connection between neural networks and decision forests is implemented in training, which reflects the influence of judging the impressions based on observation and ensures the validity of the model. The experimental results show that the proposed DDL method has good performance and generalizability.

REFERENCES

- [1] W. Hou, X. Gao, D. Tao, X. Li, Blind image quality assessment via deep learning, *IEEE transactions on neural networks and learning systems* 26 (2015) 1275-1286.
- [2] J. K. Romberg, H. Choi, R. G. Baraniuk, Bayesian tree-structured image modeling using wavelet-domain hidden markov models, *IEEE Transactions on image processing* 10 (2001) 1056-1068.
- [3] Y. Bengio, O. Delalleau, On the expressive power of deep architectures, in: *International Conference on Algorithmic Learning Theory*, Springer, 2011, pp. 18-36.
- [4] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in: *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 315-323.
- [5] C. Recommendation, Method for the subjective assessment of the quality of television pictures, CCIR Broadcasting Service (Television) (1992) 166-189, 1992.
- [6] P. Kotschieder, M. Fiterau, A. Criminisi, S. Rota Bulò, Deep neural decision forests, in: *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1467-1475.
- [7] I. Barandiaran, The random subspace method for constructing decision forests, *IEEE transactions on pattern analysis and machine intelligence* 20 (1998).
- [8] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1733-1740.
- [9] E. P. Simoncelli, B. A. Olshausen, Natural image statistics and neural representation, *Annual review of neuroscience* 24 (2001) 1193-1216.
- [10] H. R. Sheikh, Z. Wang, L. Cormack, A. C. Bovik, Live image quality assessment database release 2, 2005.
- [11] W. Pirie, Spearman rank correlation coefficient, *Encyclopedia of statistical sciences* 12 (2004).
- [12] J. Benesty, J. Chen, Y. Huang, I. Cohen, Pearson correlation coefficient, in: *Noise reduction in speech processing*, Springer, 2009, pp. 1-4.
- [13] T. Chai, R. R. Draxler, Root mean square error (rmse) or mean absolute error (mae)?—arguments against avoiding rmse in the literature, *Geoscientific model development* 7 (2014) 1247-1250.