Blind Image Quality Assessment by Visual Neuron Matrix

Hua-Wen Chang, Xiao-Dong Bi, and Chen Kai

***Abstract*—Modeling of the human visual system (HVS) is considered to be the most appropriate method to evaluate the image quality. The natural image statistics (NIS) can not only reflect certain response characteristics of the visual system, but also provide accurate quantitative descriptions of the HVS. Inspired by this fact, a visual model, which is named as visual neuron matrix (VNM), is built by a statistical learning process on natural image samples. On the basis of the VNM, a blind image quality assessment (BIQA) model is proposed. First, the VNM is obtained by independent component analysis, which can simulate the visual neurons in the cerebral cortex as well as extract the visual features from test images. Then, a regression model (neural network or support vector machine) is used to associate the visual features with the quality score of a test image. The experimental results show that the proposed method is more competitive and efficient than the state-of-the-art BIQA models.**

***Index Terms*—Human visual system, natural image statistics, visual neuron matrix, image quality assessment.**

I. Introduction

SINCE the human visual system (HVS) is the ultimate judge of perceptual image quality, the most accurate way to measure image quality is subjective evaluation. However, subjective image quality assessment (IQA) is a time-consuming and expensive solution [1]. Therefore, we need objective IQA methods that can automatically predict the quality of images.

According to the availability of reference image information, objective IQA models can be classified into full-reference IQA (FR-IQA)[2], reduced-reference IQA (RR-IQA)[3] and no- reference IQA (NR-IQA) or blind IQA (BIQA)[4]. Because there is little or no reference information for the image in real world applications, BIQA models had attracted considerable research interests [5].

Researchers have studied the BIQA models extensively, in which image feature extraction is an important part of the BIQA models [6]. Existing general BIQA models are divided into two categories according to the type of features used [7]. The first category relies on manually extracted features of images[8]. The second category tries to get the image features through learning techniques [8], [9].

The first category is based on manually extracted features that are sensitive to the presence of different distortions [9]–[12]. Manual extraction of image features is usually based on some specific physical properties or biological characteristics [13]. These low-level features are usually represented by the parameters of specific probabilistic models [7]. Because the extracted features cannot adapt to new types of distortions [14], [15], their performance is relatively poor [16]. The other type of methods can provide better performance, which are based on features that are extracted by learning techniques, such as CORNIA [17], MEON [18] and DB-CNN [19]. These methods rely on large code-book or complex network structure, the evaluation results depend heavily on huge amount of computation and annotated data. All these factors are time-consuming and not applicable in practice [7].

The problem of IQA is an interactive process between images and the HVS. So we need to consider both the HVS and image information when designing IQA models [20]. Studies have shown that there are important statistical rules in natural images [21], such as sparseness [20], independence [21], the correlation of energies [10], etc. The natural image statistics (NIS) can display the statistical adaptive representation of visual input. Moreover, it also reflects some important response characteristics of the visual system. Therefore, the NIS models can be a helpful tool to deal with IQA problems [20], [22].

As an important model of NIS, independent component analysis (ICA) can provide accurate qualitative predictions that usually turn out to be in line with the measurements from the visual cortex [23]. Inspired by the above facts, we tried to build a visual model though a training process based on NIS. This visual model is learned from a large amount of natural image samples by means of ICA training. The obtained visual model, which is named as the visual neuron matrix (VNM), can simulate the cerebral cortical neurons. The spatial organization of each unit in the VNM is very similar to the structure of the simple-cell receptive fields in the visual cortex. The VNM is trained only once, then it is used to extract visual features that can be applied to the designing of IQA models. In fact, the training of the VNM is independent of the quality prediction task. Once the VNM is obtained, it can be directly used for different quality prediction task without additional learning steps.

In this paper, a visual model (i.e., VNM) for the BIQA is proposed on the basis of NIS theory. Then a BIQA method is presented by incorporating this visual model with neural network (NN) or support vector machine (SVM). The VNM,

Manuscript received June 6, 2021; revised August 8, 2021; accepted August 14, 2021. Date of publication August 20, 2021; date of current version September 17, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 61401404, 61502435, and 61602423; in part by the Key Research Project of Colleges in Henan Province under Grant 15A520107; in part by the Basic Research Project of Henan under Grant 142300410374; and in part by the Science and Technology Project of Henan Province under Grants 192102210136 and 212400410154. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Demetrio Labate. *(Corresponding author: Xiao-Dong Bi)*

The authors are with the Zhengzhou University of Light Industry, Zhengzhou 450001, China (e-mail: changhuawen@gmail.com; bixiaodong@zzuli.edu.cn; chenkai@zzuli.edu.cn).

Digital Object Identifier 10.1109/LSP.2021.3106579

which is considered as a visual model, can simulate the processing of the visual neurons and extract visual features from test images. Finally, the visual features of the test image are mapped to the quality score by using the NN. In addition, the quality mapping can also be performed by SVM, which can form another BIQA method. The two BIQA methods are named as VNM-NN and VNM-SVM for short, respectively.

II. The Training Process of VNM

The acquisition of the VNM requires a training process. At the beginning of the training process, ten natural images with no distortion are randomly selected as the sample source from CIFAR-10 database [24]. Then 10000 color image blocks of size $16\times16\times3$ are randomly taken from these sample images. Each image block is scanned row by row and channel by channel, then each image block is vectorized into a column vector. Since a color image block has three channels, the length of the column vector is $16\times16\times3 = 768$. Then each vector subtracts its mean pixel value. Finally, all the column vectors of the sample images will form a matrix, $\mathbf{X}$.

In order to remove the redundant information and retain the basic information of the sample images to the maximum extent, principal component analysis (PCA) is used to reduce the dimension of the sample data, which can also be used for whitening of the sample data. PCA can be accomplished by eigenvalue decomposition of the covariance matrix of the data, $\mathbf{X}$. The covariance matrix can be obtained by:

$$\mathbf{U} = \frac{1}{N}\left(\mathbf{X}\times\mathbf{X}^{\mathbf{T}}\right) \tag{1}$$

where N is the number of the sample image blocks.

Then the covariance matrix, $\mathbf{U}$, is decomposed into the maximum eigenvalues and the corresponding eigenvectors. The eigenvectors are represented as $\mathbf{V} = (\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_L)$, the eigenvalues are $\mathbf{D} = \mathrm{diag}(d_1,d_2,\ldots,d_L)$, where L is the number of retained principal components. It means the first L principal components are selected for the subsequent process. According to our experiments, top performance is achieved when $L = 128$.

After extracting the principal components, the sample data $\mathbf{X}$ is whitened into $\mathbf{Z}$ by:

$$\mathbf{Z} = \frac{1}{\sqrt{\mathbf{D}}}\times\mathbf{V}^{\mathrm{T}}\times\mathbf{X} \tag{2}$$

Finally, the whitened data, $\mathbf{Z}$, will be used for the training of the VNM. We need the ICA algorithm to separate the data into statistically independent non-Gaussian linear combinations. The basic form of iteration for finding a maximum non-Gaussian direction by ICA algorithm is as follows:

$$\mathrm{VNM}^{\omega}_{(\mathrm{n})} = \mathbf{V}\left\{\mathbf{Z}\times g\left(\left(\mathrm{VNM}^{\omega}_{(\mathrm{n}-1)}\right)^{\mathrm{T}}\mathbf{Z}\right)\right\} - \mathbf{V}\left\{g'\left(\left(\mathrm{VNM}^{\omega}_{(\mathrm{n}-1)}\right)^{\mathrm{T}}\mathbf{Z}\right)\right\}\mathrm{VNM}^{\omega}_{(\mathrm{n}-1)} \tag{3}$$

where $g(\mathrm{x}) = tanh(\mathrm{x})$ is a hyperbolic tangent function, and $g'(\cdot)$ is the derivative of $g(\cdot)$.

Each new iteration of the model starts with a random orthogonal matrix, and each iteration requires updating VNM^{ω} and then orthogonalizing the matrix.

$$\mathrm{VNM}^{\omega} = \mathrm{VNM}^{\omega}\times\frac{1}{\sqrt{\left(\mathrm{VNM}^{\omega}\times\mathrm{VNM}^{\omega\mathrm{T}}\right)}} \tag{4}$$

This process is repeated until the final result reaches convergence. The convergence criterion is as follows:

$$\left\|\left|\mathrm{VNM}^{\omega}_{(\mathrm{n})}\times{\mathrm{VNM}^{\omega}_{(\mathrm{n}-1)}}^{\mathrm{T}}\right| - \mathbf{I}\right\| \le 10^{-9}\times L \tag{5}$$

where $\|\bullet\|$ denotes the Frobenius norm, $\mathbf{I}$ stands for an identity matrix.

After the training process is completed, the obtained matrix should be converted from the whitened space to the original image space by:

$$\mathrm{VNM} = \mathrm{VNM}^{\omega}\times\frac{1}{\sqrt{\mathbf{D}}}\times\mathbf{V}^{\mathrm{T}} \tag{6}$$

Since only the first L ($L = 128$) principal components of the sample data are retained during the whitening step, the size of the VNM is 128×768. In other words, the image column vector with the length of 768 will be transformed into a visual feature vector with the length of 128.

Each row element of the VNM has three parts corresponding to three color channels, and each part of the row can be transformed into a block unit which is similar to the receptive field of simple cell in the visual cortex. Then the VNM can be represented by three groups of blocks units which can be considered as the model of visual neurons, as shown in Fig. 1. In Fig. 1, each group has 128 visual neurons corresponding to a color channel, and the size of each neuron is 16×16. Since these blocks are similar to the receptive fields of the cerebral cortex cells, the VNM can be used as visual neurons in the visual cortex, which can convert images into visual responses.

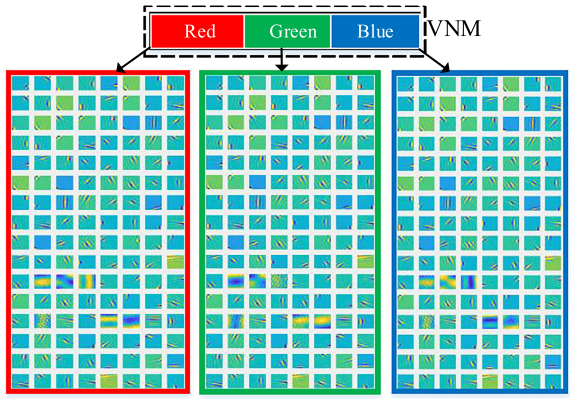

Fig. 1. A graphical representation of the VNM. VNM corresponds to three color channels, and each color channel has 128 receptive fields of neurons.

III. Image Quality Assessment by VNM

A. Image Preprocessing

First, the test image is divided into color image blocks with a size of $16\times16\times3$. Each image block, $\mathbf{B}$, subtracts its mean pixel value as follows:

$$\mathbf{C} = \mathbf{B} - \mu\left(\mathbf{B}\right) \tag{7}$$

where $\mu(\mathbf{B})$ represents the mean value of the image block $\mathbf{B}$.

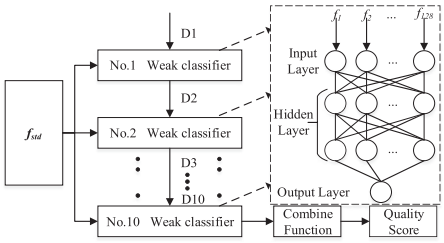

Fig. 2. The NN structure diagram.

Then the values of the image block, **C**, needs to be scanned row by row, by which a block will be converted into a column vector with the size of 768×1. After that, all the vectors from a test image are formed into a new matrix, **M**. Each column vector of **M** represents each color image block:

$$\mathbf{M} = (\boldsymbol{m}_1 \cdots \boldsymbol{m}_t) \tag{8}$$

where *t* is the total number of image blocks in a test image.

B. Feature Extraction Using VNM

After the preprocessing step, the extraction of visual features from a test image can be simply achieved by the computation of matrix multiplication, as shown in equation (9).

$$\mathbf{F} = \mathrm{VNM} \times \mathbf{M} \tag{9}$$

where **F** represents the set of obtained visual features.

The obtained visual feature matrix, **F**, has a size of *L*×*t*, and *t* represents the total number of image blocks. Each column represents the output signal of an image block from the visual model, VNM. Then, we calculate the standard deviation of each row in **F**, as shown in equation (10).

$$f_i = \sqrt{\frac{1}{t}\sum\nolimits_{j=1}^{t}\left(\boldsymbol{F}_{ij} - \mu\left(\boldsymbol{F}_i\right)\right)^2} \tag{10}$$

where $i \in [1, 2, \ldots, L]$, *L* is the number of rows in **F**, which can be considered as the number of neurons, f_i is the standard deviation of the *i*-th row in **F**. Finally, all the standard deviations of the row elements form a vector $\boldsymbol{f}_{std}$:

$$\boldsymbol{f}_{std} = \left(f_1 \cdots f_i \cdots f_L\right)^{\mathrm{T}} \tag{11}$$

C. Quality Score Mapping

Through the above step, a test image is transformed into a vector with the length of *L*. Finally, Ada-boosting NN is used to achieve the mapping from visual features to quality scores, as shown in Fig. 2.

This NN makes a combination of multiple weak classifiers to become a strong classifier. There are 10 weak classifiers in the network, which are represented by D1, D2, . . ., D10 in Fig. 2. For each classifier, there are three components, i.e., input layer, hidden layer and output layer. The number of nodes in the input layer is *L* which is the size of the vector, $\boldsymbol{f}_{std}$, while the output layer is composed of only one node, and its output is the quality prediction of the test image. The structure of the weak classifier was designed as two hidden layers, the first hidden layer used tangent sigmoid function as the activation function:

$$\mathrm{S}\left(x_{dif}\right) = \frac{1}{1 + e^{-a\left(x_{dif} - c\right)}} \tag{12}$$

where x_{dif} represents the difference between the predicted and subjective quality scores, which is calculated by the weak classifier. Let *a* define the width of the transition area, and *c* define the center of the transition area. In this study, *a* is set to 1 and *c* is set to 0.5.

The activation function of the second hidden layer is a radial basis function. As shown in equation (13), a Gaussian kernel function is selected as the radial basis function:

$$\mathrm{K}\left(\|x_{test} - x_c\|\right) = e^{\left(-\frac{\|x_{test} - x_c\|^2}{2\times\sigma^2}\right)} \tag{13}$$

where x_c represents the center of the kernel function, and σ is taken as the width parameter for the radial range, σ defaults to 16. $\|x_{test} - x_c\|$ represents the Euclidean distance between any point x_{test} in the space and the center of the kernel function x_c.

TABLE I
SRCC RESULTS ON LIVE, CSIQ TID2013, LIVEMD AND KADID-10K

Methods	LIVE	CSIQ	TID2013	LIVEMD	KADID-10K
BRISQUE	0.9391	0.8608	0.5675	0.8864	0.4734
CORNIA	0.9471	0.6789	0.6786	0.8991	0.7804
IL-NIQE	0.9021	0.8683	0.5126	0.8911	0.5015
OG-IQA	0.9500	0.7986	0.6875	0.8765	0.6647
dip-IQA	0.9324	0.8754	0.8233	0.4801	0.3048
BIECON	0.9611	0.8154	0.7179	0.9099	0.4576
MEON	0.9625	0.9322	**0.8262**	0.6879	0.2345
DB-CNN	**0.9682**	**0.9461**	0.8161	0.9275	0.5019
MsKLT	0.9413	0.9255	0.7988	0.9157	0.9088
UNIQUE	**0.9692**	0.9024	0.7719	0.9214	**0.8781**
AIGOA	0.9601	0.9277	0.8211	**0.9336**	0.8635
VNM-NN	0.9545	**0.9537**	**0.8382**	**0.9574**	**0.8885**

IV. EXPERIMENT AND RESULTS

A. Experimental Protocols and Performance Criteria

The experimental process follows the same protocol in [19], [25]. The experimental comparison was conducted on five databases, including the LIVE [26], CSIQ [27], TID2013 [28], LIVEMD [29] and KADID-10K [30] database. Spearman rank order correlation coefficient (SRCC) and Pearson linear correlation coefficient (PLCC) are adopted to measure the performance of IQA models. Higher values of SRCC and PLCC indicate better performance. In order to maintain the same experiment protocols as for other BIQA models [17]–[19], the distorted images were divided into two parts, 80% of which were used for training and the remaining 20% for testing. To avoid randomness in the results, the training and testing procedures were repeated 10 times. The median of the 10 results was selected as the final result.

B. Performance on Individual Database

We compared the proposed model against several classic and the latest BIQA methods, including BRISQUE [31], CORNIA [17], IL-NIQE [32], OG-IQA [33], dip-IQA [34], BIECON [25], MEON [18], DB-CNN [19], UNIQUE [35], MsKLT [36] and AIGOA [37]. Table I lists the SRCC results which show that VNM-NN performs significantly better than other classical and CNN based BIQA methods on CSIQ, TID2013, LIVEMD and KADID-10K.

C. Individual Distortion Types

In order to further look at the behaviors of VNM-NN on individual distortion types along with several competing BIQA models, we test them on images of various distortion types.

TABLE II
SRCC RESULTS OF DIFFERENT DISTORTION DATA ON CSIQ

Methods	JPEG	JP2K	WN	GB	PN	CC
BRISQUE	0.8062	0.8401	0.7232	0.8203	0.3788	0.8042
CORNIA	0.5132	0.8311	0.6641	0.8362	0.4931	0.4622
IL-NIQE	0.9052	0.9244	0.8673	0.8672	0.7894	0.8013
OG-IQA	0.8124	0.8645	0.7678	0.8547	0.7894	0.7549
dip-IQA	0.9361	0.9442	0.9044	**0.9321**	0.8945	0.8614
BIECON	0.9014	0.9214	0.9001	0.9215	0.8661	0.8319
MEON	**0.9481**	0.8893	**0.9511**	0.9184	0.8342	0.8501
DB-CNN	0.9400	**0.9531**	0.9482	**0.9473**	**0.9405**	**0.8704**
MsKLT	0.9475	0.8928	0.9466	0.9005	0.9212	0.8457
UNIQUE	0.9279	0.9014	0.9135	0.8997	0.8912	0.8212
AIGOA	0.9356	0.9517	0.9214	0.8816	0.9018	0.8489
VNM-NN	**0.9557**	**0.9651**	**0.9679**	0.9238	**0.9531**	**0.9481**

TABLE III
SRCC RESULTS OF DIFFERENT DISTORTION DATA ON TID2013

Types	CORNIA	IL-NIQE	MEON	DB-CNN	AIGOA	VNM-NN
AGN	0.7561	**0.8760**	0.8132	0.7902	0.8621	**0.8930**
ANC	**0.7498**	**0.8159**	0.7225	0.7009	0.7161	0.6437
SCN	0.7265	**0.9233**	**0.9264**	0.8263	0.8442	0.8374
MN	0.7262	0.5120	**0.7281**	0.6469	0.6629	**0.8166**
HFN	0.7964	0.8685	**0.9112**	0.8791	**0.9534**	0.8736
IN	0.7667	0.7551	**0.9012**	0.7089	0.8913	**0.9038**
QN	0.0156	**0.8730**	0.8884	0.8251	0.8589	**0.9308**
GB	**0.9209**	0.8142	0.8872	0.8592	**0.9175**	0.8083
DEN	0.8315	0.7500	0.7964	**0.8652**	**0.9148**	0.7857
JPEG	0.8743	0.8349	0.8509	**0.8945**	0.8454	**0.9200**
JP2K	0.9103	0.8578	0.8912	**0.9163**	**0.9120**	0.9038
JGTE	0.6856	0.2827	0.7465	0.7728	**0.8582**	**0.7869**
J2TE	0.6784	0.5248	0.7168	**0.7735**	0.6983	**0.8083**
NEPN	**0.2857**	0.0805	0.1169	0.2704	0.1306	**0.8291**
BLOCK	0.2188	0.1357	**0.5005**	0.4445	0.2238	**0.7942**
MS	0.0645	0.1845	0.1778	0.0098	**0.5544**	**0.7233**
CC	0.1823	0.0141	0.2524	**0.5487**	0.4306	**0.8599**
CCS	0.0807	0.1628	**0.6849**	0.6313	0.6814	**0.8451**
MGN	0.6438	0.6932	**0.8491**	0.7115	0.8486	**0.8506**
CN	0.5341	0.3599	0.4068	**0.7525**	0.6867	**0.8734**
LCNI	**0.8623**	0.8287	0.7724	**0.8601**	0.8400	0.7618
ICQD	0.2717	0.7487	**0.8579**	0.8331	0.8284	**0.8351**
CHA	0.7922	0.6793	0.7794	0.7321	**0.8892**	**0.9022**
SSR	0.8624	0.8650	0.8559	**0.9021**	**0.9087**	0.8735
Count	4	4	8	8	8	16

Table II and Table III show the experimental results of various distortion types in the CSIQ and TID2013 image database, the best two SRCC results are highlighted in bold.

The CSIQ contains six types of distortions: white Gaussian noise (WN), JPEG compression (JPEG), JPEG-2000 compression (JP2K), pink Gaussian noise (PN), Gaussian blur (GB), and global contrast decrements (CC). Table II shows that VNM-NN has good performance on JPEG, JP2K, WN, PN and CC. The overall performance proves that our model is universal for each distortion type. This conclusion is also strongly confirmed in the TID2013 database.

Table III shows the results of the 6 best BIQA methods on various distortions in the TID2013 database. VNM-NN achieves the top two performance on 16 out of 24 distortion types. Moreover, the performance of VNM-NN is relatively consistent among all distortion types.

Most BIQA models perform well on some distortions but fail to provide satisfactory performance on non-eccentricity pattern noise (NEPN), local block-wise distortions (BLOCK) and mean shift (MS) [19]. Whereas, the VNM-NN has a much higher prediction accuracy than all the other models on these three distortion types. Therefore, VNM-NN can provide not only relatively more accurate prediction results for all the distortions, but also stable performance on different distortion types.

In addition, the most BIQA methods perform poorly on color distortions, e.g., changes of color saturation (CCS) and image color quantization with dither (ICQD) in TID2013 database. Since VNM-NN uses the all RGB color components in a perceptual way, it is very sensitive to color distortions.

TABLE IV
SRCC AND PLCC RESULTS OF DIFFERENT REGRESSION MODELS ON THE CSIQ, LIVE AND TID2013

Methods	LIVE		CSIQ		TID2013	
	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
VNM-NN	0.9465	0.9445	0.9603	0.9537	0.8679	0.8382
VNM-SVM	0.9478	0.9425	0.9613	0.9514	0.8489	0.8217

TABLE V
TIME COST OF EACH BIQA METHOD

Methods	BRISQUE	CORNIA	IL-NIQE	OG-IQA
Time Cost(s)	0.333	2.897	4.092	0.199
Methods	MEON	DB-CNN	AIGOA	VNM-NN
Time Cost(s)	0.689	1.756	2.156	0.121

D. Performance of Different Regression Models

In order to validate the effect of the regression model we used, we also chose SVM as the regression function for testing. And the results are listed in Table IV. There is little difference between the performance of the VNM-NN and VNM-SVM, indicating that the VNM can work with different regression models and provide stable performance.

E. Computational Cost

To compare the computational efficiency of BIQA methods, we measured the time cost of each method in evaluating the quality of a 512×512 color image (taken from CSIQ). The results are in Table V which shows that VNM-NN is faster than other methods. Because deep learning-based BIQA methods (e.g., MEON, DBCNN and AIGOA) contain multi-layer networks and a large number of parameters, the process of extracting image features consumes a lot of time [38]. On contrary, the feature extraction stage of IQA can be quickly finished by matrix multiplication with the VNM which has been obtained beforehand.

V. CONCLUSION

In this paper, we proposed a BIQA method on the basis of VNM. The VNM, which is considered as a visual model, can be used to extract visual features from test images in the form of multiplication. Then a regression model (SVM or NN) is used to map the visual features of a test image to a quality score. The experimental results on LIVE, CSIQ, TID2013, LIVEMD and KADID-10K show that both the VNM-NN and VNM-SVM can provide accurate results and better performance than the state-of-the-art BIQA models. However, according to further experiments, VNM-NN is not sufficient to solve authentic distortion types, given its extremely simple structure and learning methods. An interesting line of research will then be how to construct a more complicated or deeper (more stages) structure for this model.

REFERENCES

[1] C. Strauss, F. Pasteau, F. Autrusseau, M. Babel, L. Bedat, and O. Deforges, "Subjective and objective quality evaluation of lar coded art images," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2009, pp. 674–677.

[2] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[3] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 824–829, Apr. 2011.

[4] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.

[5] S. Gabarda and G. Cristobal, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Amer.*, vol. 24, no. 12, pp. B42–B51, 2007.

[6] W.-H. Zhu, W. Sun, X.-K. Min, G.-T. Zhai, and X.-K. Yang, "Structured computational modeling of human visual system for no-reference image quality assessment," *Int. J. Autom. Comput.*, vol. 18, no. 2, pp. 204–218, 2021.

[7] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4444–4457, Sep. 2016.

[8] A. B. Jeripothula, S. K. Velamala, S. K. Banoth, and S. Mukherjee, "Blind image quality assessment using a combination of statistical features and CNN," *Multimed. Tools Appl.*, vol. 79, no. 31–32, pp. 23243–23260, Aug. 2020.

[9] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[10] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2457–2469, Dec. 2016.

[11] C. Feichtenhofer, H. Fassold, and P. Schallauer, "A perceptual image sharpness metric based on local edge gradient analysis," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 379–382, Apr. 2013.

[12] L. He, D. Tao, X. Li, and X. Gao, "Sparse representation for blind image quality assessment," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2012, pp. 1146–1153.

[13] C. Zhang, J. Xu, X. Huang, and S. H. Park, "No-reference image quality assessment using independent component analysis and convolutional neural network," *J. Elect. Eng. Technol.*, vol. 14, no. 1, pp. 487–496, Jan. 2019.

[14] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.

[15] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[16] J. Hu, X. Wang, F. Shao, and Q. Jiang, "TSPR: Deep network-based blind image quality assessment using two-side pseudo reference images," *Digit. Signal Process.*, vol. 106, 2020, Art. no. 102849.

[17] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2012, pp. 1098–1105.

[18] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018.

[19] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 36–47, Jan. 2020.

[20] H.-W. Chang, H. Yang, Y. Gan, and M.-H. Wang, "Sparse feature fidelity for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 4007–4018, Oct. 2013.

[21] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, Mar. 1999.

[22] A. Ahar, A. Barri, and P. Schelkens, "From sparse coding significance to perceptual quality: A new approach for image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 879–893, Feb. 2018.

[23] A. Hyvärinen, "Statistical models of natural images and cortical visual representation," *Top. Cogn. Sci.*, vol. 2, no. 2, pp. 251–264, Apr. 2010.

[24] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," 2009. [Online]. Available: http://www.cs.toronto.edu/∼kriz/cifar.html

[25] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 206–220, Jan. 2017.

[26] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[27] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, Mar. 2010, Art. no. 011006.

[28] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process.-Image Commun.*, vol. 30, pp. 57–77, 2015.

[29] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Proc. Conf. Rec. 46th Asilomar Conf. Signals, Syst. Comput.*, 2012, pp. 1693–1697.

[30] H. Lin, V. Hosu, and D. Saupe, "KADID-10k: A large-scale artificially distorted IQA database," in *Proc. 11th Int. Conf. Qual. Multimedia Experience*, 2019, pp. 1–3.

[31] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[32] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.

[33] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Signal Process.-Image Commun.*, vol. 40, pp. 1–15, 2016.

[34] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipIQ: Blind image quality assessment by Learning-to-Rank discriminable image pairs," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3951–3964, Aug. 2017.

[35] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Trans. Image Process.*, vol. 30, pp. 3474–3486, 2021.

[36] C. Yang, X. Zhang, P. An, and L. Shen, "Blind image quality assessment based on multi-scale KLT," *IEEE Trans. Multimedia*, vol. 23, pp. 1557–1566, 2021.

[37] J. Ma, J. Wu, L. Li, W. Dong, and X. Xie, "Blind image quality assessment with active inference," *IEEE Trans. Image Process.*, vol. 30, pp. 3650–3663, 2021.

[38] S. Bosse, D. Maniry, K.-R. Mueller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.