Explicit-implicit dual stream network for image quality assessment

Guangyi Yang1, Xingyu Ding1, Tian Huang1, Kun Cheng1 and Weizheng Jin1,2*

Abstract
Communications industry has remarkably changed with the development of fifth-generation cellular networks. Image, as an indispensable component of communication, has attracted wide attention. Thus, finding a suitable approach to assess image quality is important. Therefore, we propose a deep learning model for image quality assessment (IQA) based on explicit-implicit dual stream network. We use frequency domain features of kurtosis based on wavelet transform to represent explicit features and spatial features extracted by convolutional neural network (CNN) to represent implicit features. Thus, we constructed an explicit-implicit (EI) parallel deep learning model, namely, EI-IQA model. The EI-IQA model is based on the VGGNet that extracts the spatial domain features. On this basis, the number of network layers of VGGNet is reduced by adding the parallel wavelet kurtosis value frequency domain features. Thus, the training parameters and the sample requirements decline. We verified, by cross-validation of different databases, that the wavelet kurtosis feature fusion method based on deep learning has a more complete feature extraction effect and a better generalisation ability. Thus, the method can simulate the human visual perception system better, and subjective feelings become closer to the human eye. The source code about the proposed EI-IQA model is available on github https://github.com/jacob6/EI-IQA.

Keywords: CNN, Feature fusion, IQA, Wavelet feature extraction, EI dual stream network

1 Introduction
The emergence of 5G [1] period has brought great innovation to communications industry, and the demand for information transmission has increased under the bombardment of high-speed information streams. Evidently, different regions have different abilities to receive information due to their geographic location and other factors. Image is the main carrier of visual information [2] because it can intuitively reflect information, which is particularly important in the information transmission process. During image acquisition and transmission, different degrees of distortion [3] are caused by various factors, such as processing system and environmental noise. Then, distortions affect people’s visual effects. Image quality directly affects the subjective perception of the human eye and the acquisition of image information. Therefore, the research on IQA has arouse widespread concern.

In accordance with whether the human eye is required for classification, IQA methods can be roughly divided into two types, namely, (1) subjective assessment and (2) objective assessment [4]. The subjective assessment method uses human visual assessment [5] as a standard to assess images based on human intuitive visual experience. In the subjective assessment method, the distorted image [6] and the original image are assessed. Although the subjective assessment method works well, it entails heavy workload; thus, it does not satisfy the requirements of practical application [7, 8]. By contrast, the objective assessment is simpler than subjective assessment; it has strong controllability. It uses the mathematical model to directly score the image without the need for an assessor, thereby saving time and resources. Thus, the subjective assessment has better application prospect and has become the main method in the IQA field [9].
Objective assessment methods can be further divided into full-reference, semi-reference and non-reference IQA [10] (or blind IQA). Their main differences lie in the presence or absence of the reference image. Non-reference IQA (NR-IQA) method has a broader application prospect because the original image is often unavailable in practical applications.

NR-IQA can be divided into two categories in accordance with the different methods of extracting features. One type is an explicit NR-IQA method [11]. The image feature which will be tested is initially extracted. Then, the extracted features will be input into the shallow regression network to obtain the final quality score. Represented by shallow machine learning, the BIQI [12] algorithm proposed by Moorthy et al. extracts statistical features in the wavelet domain of distorted images based on a two-level framework and uses support vector machines to classify image distortions. It calculates the probability of the existing distortion type and the quality corresponding to each distortion. The final quality is the weighted sum of the distortion probability and the corresponding quality. On the basis of the BIQI model, Moorthy et al. proposed an image authenticity and integrity assessment model, DIVINE [13] algorithm based on distortion type identification. This algorithm uses a controllable pyramid [14] to perform wavelet decomposition in the direction and scale, extracts the statistical characteristics of the wavelet coefficients after separating normalisation [15], and then uses support vector machines to build a feature model. Saad et al. proposed the BLIINDS [16] algorithm and the BLIINDS-II [17] improved algorithm. The image was divided into blocks, and then for each block, the statistical characteristics of the discrete cosine transform (DCT) [18] coefficients were extracted in the DCT domain of the image to establish a support vector regression model. The model is based on the statistics of the local discrete cosine transform coefficients to achieve performance that satisfies the requirements of real-time systems. The improved BLIINDS-II algorithm uses a simple expression, a low-dimensional feature space and a simple Bayesian prediction framework, in the sparse DCT domain. Mittal et al. proposed the BRISQUE [19] algorithm, which established a regression model by extracting statistical features of the image's spatial normalisation [20] coefficients. After calculating the mean subtracted contrast normalised (MSCN) [21], it was modelled by a symmetric generalised Gaussian distribution model (GGD) [12] and an asymmetric GGD (AGGD [22]) model to obtain statistical features. Then, it used the nearest neighbour algorithm for downsampling, extracted features on another scale, and finally obtained 36 features in the training image. On the basis of the BRISQUE method, Mittal et al. proposed a “completely blind” image quality analyser NIQE [23]. After calculating the MSCN of the image to be tested, the block is partitioned in a non-interval manner, and then the feature vector is extracted by the same method of the BRISQUE algorithm for the block. Finally, the multivariate Gaussian (MVG) model was fitted to the extracted feature vector to obtain the final quality assessment score. Zhang et al. proposed NIQE’s upgraded algorithm IL-NIQE [24] based on NIQE, which measures the image quality by calculating the distance between the distorted image and the undistorted image by the multivariate Gaussian distribution model, and performs principal component analysis (PCA) dimensionality reduction on the extracted feature vectors. Then, the MVG calculation was performed on the feature obtained from each image block to finally obtain the feature quality. Zhang et al. proposed the DESIQUE [25] algorithm, which extracts features from the spatial and frequency domains, then downsamples the image, and finally obtains a quality score through a shallow regression network. Liu et al. proposed the SSEQ [26] algorithm to block the input pictures and calculate the local entropy average, local entropy skewness, local spectral domain entropy mean, and local spectral domain entropy skewness of each region. Then, 12 image features were obtained by performing two downsampling. Finally, a quality score through a shallow regression network was obtained. Explicit NR-IQA algorithm performance comparison is shown in Table 1.

The early extraction of explicit features cannot establish an end-to-end model [27] due to the limited computational power. Therefore, it uses hierarchical extraction method to obtain feature set initially, and then performs corresponding operations on the feature set to achieve the image score. With the advancement of computing power and the emergence of deep networks, the implicit NR-IQA is introduced. Implicit NR-IQA [11] inputs the image into the algorithm, establishes an end-to-end model, and directly obtains the final image quality score. Represented by deep networks, the DeepIQA [28] model proposed by Gao Xinbo et al. performs qualitative assessment through machine learning and outputs numerical scores. The image is represented by the statistical characteristics of natural scenes [2]. The deep model is trained, the classification framework is established, the extracted features are graded to correspond to different subjective feelings, and then the qualitative labels are converted into image quality assessment scores by pooling. The CNNIQA [29] algorithm proposed by Kang et al. is based on the BIQA model of the CNN [30]. It takes the image patch as the input and uses back propagation and other methods for training. CNN works in the spatial domain, and the feature extraction and regression are integrated into the CNN, thereby deepening the network depth to improve learning ability. Kang et al. proposed a follow-up algorithm CNNIQA++ [31] to increase the number of convolutional layers on the basis of CNNIQA. It modified the full
Table 1 Explicit NR-IQA algorithm performance comparison

| Algorithm   | Parameters used | No. of parameters | SROCC | PLCC | Training method | Year |
|-------------|-----------------|-------------------|-------|------|-----------------|------|
| BIQI        | GGD ($\alpha, \theta^2$) | 18                | 0.8271 | 0.8417 | SVM              | 2010 |
| DIVINE      | GGD ($\alpha, \theta^2$) | 88                | 0.9160 | 0.9170 | SVM              | 2011 |
| BLINDS-II   | DCT             | 16                | 0.9124 | 0.9164 | Model            | 2012 |
| BRISQUE     | GGD ($\alpha, \theta^2$), AGGD ($\eta, \nu, \theta^2, \theta^2$) | 36                | 0.9393 | 0.9424 | SVM              | 2012 |
| DESIQUE     | GGD ($\alpha, \theta^2$) | 60                | 0.9437 | 0.9465 | SVM              | 2013 |
| NIQE        | GGD ($\alpha, \theta^2$), AGGD ($\eta, \nu, \theta^2, \theta^2$) | 36                | 0.9153 | 0.9147 | Model            | 2013 |
| SSEQ        | entropy         | 12                | 0.9348 | 0.9383 | SVM              | 2014 |
| IL-NIQE     | GGD ($\alpha, \theta^2$), AGGD ($\eta, \nu, \theta^2, \theta^2$) | 468               | 0.9025 | 0.9085 | Model            | 2015 |

Table 2 Implicit NR-IQA algorithm performance comparison

| Algorithm   | Depth | SROCC | PLCC | Training method | Year |
|-------------|-------|-------|------|-----------------|------|
| CNNIQ       | 5     | 0.956 | 0.953 | CNN             | 2014 |
| DeepBIQ     | 3     | 0.889 | 0.908 | CNN             | 2017 |
| RankIQA     | 16    | 0.981 | 0.982 | CNN             | 2017 |
| dipIQ       | 5     | 0.960 | 0.969 | RankNet         | 2017 |
| BIECON      | 6     | 0.961 | 0.962 | CNN             | 2017 |
| deepIQA     | 12    | 0.972 | 0.960 | VGGnet          | 2018 |
| MEON        | 13    | 0.974 | 0.972 | CNN             | 2018 |
| BLINDER     | 37    | 0.966 | 0.959 | VGGnet          | 2018 |
| DQA         | 7     | 0.970 | 0.971 | CNN             | 2019 |
| DeepVR-IQA  | 7     | 0.882 | 0.888 | ResNet-50       | 2020 |
increasing model parameters and the difficulty of machine learning. The explicit NR-IQA has good physical meaning and interpretability. Its advantage lies in its interpretability, which is convenient for insight into the relationship between features, and selective combination can obtain better stackability. Hence, people use the feature set of explicit feature extraction to enhance other machine learning algorithms as the basis of training. However, the final extracted feature information is incomplete. Figure 1 shows a comparison of explicit and implicit network structures.

For this reason, the proposed EI-IQA combines explicit features and implicit features and puts forward an experimental scheme. We combine the explicit and implicit features to describe the image characteristics, and then develop an effective deep learning model. After training and model parameters adjustment, we can finally obtain a reliable image quality score and distortion type. Compared with the traditional IQA algorithm, the proposed EI-IQA combines the advantages of explicit features and implicit features, extracts mixed features, and combines the CNN model to establish an end-to-end model. It uses deep neural networks to extract spatial features representing implicit features, and then uses wavelet transform to extract frequency domain features representing explicit features. Frequency domain features supplement the spatial features, reducing the number of deep neural network layers, improving generalisation capabilities, reducing the training difficulty, and avoiding the loss of information during feature extraction. Compared with traditional methods, the proposed EI-IQA comprehensively proposes combining two features to complement each other. The implicit feature makes up for the shortcoming of insufficient explicit feature; the explicit feature makes up for the shortcoming of the physical meaning of the implicit feature extracted by the deep network. At the same time, the number of network layers, network parameters and complexity are reduced.

• The proposed EI-IQA comprehensively proposes combining two features to complement each other. The implicit feature makes up for the shortcoming of insufficient explicit feature; the explicit feature makes up for the shortcoming of the physical meaning of the implicit feature extracted by the deep network. At the same time, the number of network layers, network parameters and complexity are reduced.

• We design a deep network that combines the advantages of explicit features and implicit features, effectively reducing the dependence of the deep network on large training samples.

• Our results prove that only when the number of network layers reaches a certain depth, the implicit features extracted by the deep network are sufficient, that is, when the number of layers of the deep network is insufficient, the extracted features are redundant and insufficient.

The remaining chapters of this paper are arranged as follows. Section 2 details the proposed EI-IQA method and parameters training process. Section 3 provides the experimental results and analysis. Finally, Section 4 concludes with a summary of our work and describes the future outlook.

2 The proposed EI-IQA framework

2.1 System solutions

With the complication of research in the field of images, feature extraction has become an important part of many algorithm design processes. In many cases, the shallow features extracted by traditional methods cannot satisfy the requirements of the algorithm. To ensure the sufficient extraction of information, most researchers use deep networks to automatically learn features from big data. However, a great deal of information redundancy resulted from a lack of evident physical significance, thereby increasing model parameters and the difficulty of machine learning.

---

**Fig. 1** Comparison of explicit and implicit network structures
The deep neural network is used to extract the implicit features represented by the spatial features of the image, and then input into the regression network to obtain the quality score and distortion type. To solve the problem on feature redundancy resulting in a large number of model parameters, we propose a new scheme that combines the explicit features represented by manually extracted wavelet features and the implicit features represented by the spatial features extracted by the deep learning network. The proposed scheme is shown in Fig. 2.

For the input image, the spatial features of the image are extracted through the VGG13 deep network model, as well as the frequency domain features of the image in the wavelet domain using wavelet transform. Image spatial domain features are used to represent implicit features, and frequency domain features are used to represent explicit features. Explicit features are combined with implicit features, and mixed features are obtained through feature fusion. Then, the mixed feature is input to perform multitask learning in the established regression network to obtain the final image quality score and distortion type. That is, adding the extraction of frequency domain information on the basis of deep network, using the implicit features extracted by the deep network model VGGNet and the explicit features extracted by wavelet transform, we propose our deep image quality assessment [47] method based on EI-IQA. We use a combination of explicit and implicit features to describe image features and deep network models to achieve effective learning and to finally obtain the quality score and distortion type.

Compared with the previous solution, we add frequency domain information extraction, which has more accurate and comprehensive features. The combination of explicit and implicit features reduces information redundancy. Explicit features are related to implicit features. The depth of the deep neural network is reduced to extract implicit features, thereby reducing parameter of training and the difficulty of machine learning. Thus, our algorithm has a better generalisation effect. In addition, because we are extracting a mixture of explicit and implicit features, which greatly reduces our demand for samples, our algorithm can obtain satisfactory results even for small sample libraries.

2.2 Feature extraction

The frequency domain features extracted by wavelet transform represent explicit features, and the spatial domain features extracted by deep network represent implicit features. To solve the shortcomings of insufficient explicit features and unclear physical meaning of implicit features, we propose a parallel deep learning model EI-IQA based on EI-IQA. The implicit features extracted by VGGNet and the explicit features represented by the wavelet kurtosis [48] frequency domain features are used as the underlying feature vector X. The explicit and implicit features are extracted in parallel. The combination of explicit features and implicit features is used as input of the regression network.

Different from the traditional deep network, the proposed EI-IQA model combines the explicit features represented by frequency domain features and the implicit features represented by spatial domain features to form a mixed feature. It solves the shortcomings of insufficient explicit features and unclear physical meaning of implicit features. These features have complementary advantages and disadvantages. In addition, the EI-IQA model can achieve an experimental effect similar to the original deep model by reducing the number of deep neural networks.
in the model, reducing parameter training and adding frequency domain information.

### 2.2.1 Explicit feature extraction based on kurtosis value in wavelet domain

According to the existing research, compared with the original image, the distribution of the distorted image is flat, the peak value is low and the tail is long. Its kurtosis value has frequency scale invariance, which can be used as a metric and a feature to distinguish images with different degrees of distortion.

In the explicit feature extraction represented by the kurtosis value in the wavelet domain, discrete wavelet transform (DWT) [49] is performed on the image, and then 38 Daubechies filters from low frequency to high frequency scale are used. Thus, wavelet sub-band coefficients of $t$ groups ($t = 38$) are obtained in low frequency, horizontal, vertical and diagonal directions, which are $L_t$, $H_t$, $V_t$ and $D_t$ ($t \in 1, 2, 3, \cdots, 38$), respectively. They are merged into 38 new matrices $J_t$ ($t \in 1, 2, 3, \cdots, 38$).

$$J_t = [L_t, H_t, V_t, D_t]$$ (1)

The kurtosis value of the matrix $J_t$ is recorded as $K(J_t)$, as follows:

$$K(J_t) = \frac{k_4(J_t)}{k_2(J_t)^2} = \frac{\mu_4(J_t)}{\sigma(J_t)^4} - 3$$ (2)

where $k_i(J_t)$ represents the $i$th cumulative function of the matrix $J_t$, and $\mu_i(J_t)$ represents the $i$th central moment of the matrix $J_t$.

The 38 kurtosis values of an image are combined into a 38-dimensional feature vector, denoted as $E$, as follows:

$$E = [K(J_1), K(J_2), \cdots, K(J_{38})]$$ (3)

The wavelet frequency domain feature extraction process is shown in Fig. 3.

### 2.2.2 Implicit feature extraction based on deep CNN

The deep neural network can avoid the loss of feature information, directly input the original image into the network model for training and combine feature learning with training. The proposed EI-IQA uses the classic deep neural network model VGG to segment the input image according to $32 \times 32$ to extract features. After the convolutional layer of $n$ ($n = 13$) layers, the final implicit features represented by the spatial features are obtained. In the equation, $n$ is the number of convolutional layers in the deep neural network model VGG.

### 2.3 Model training

In the model training process, the image in the database is initially normalised, and then divided into $32 \times 32$ size. Then, the multitask CNN is used to predict the image block. Finally, the prediction value combined with the original image is used to obtain the final results. Assuming that the original image is uniformly distorted, the resulting quality score is the average of the quality scores of all image blocks, and the type of distortion of the final image is determined by voting for most image blocks.

---

**Fig. 3**: Schematic of explicit feature extraction structure
2.3.1 Image normalisation

The essence of neural network learning is the distribution of learning data. The difference between the distribution of training data and test data greatly reduces the generalisation ability of the model. Moreover, the neural network learning process is complicated. Once a slight change occurs in a certain middle layer of the network, it is gradually enlarged in the subsequent process. This network layer needs to be learned to adapt to the new data distribution in each iteration process, greatly reducing the speed of the neural network training. Therefore, the normalised data preprocessing is an indispensable part of the model training process.

The image is cut into 32 $\times$ 32 noncoincident image blocks, and then the local normalisation operation is performed as follows:

$$
\hat{l}(i,j) = \frac{l(i,j) - \mu(i,j)}{\sigma(i,j) + C}
$$

(4)

where $l(i,j)$ represents a locally normalised brightness image matrix, $i \in 1, 2, \cdots, M, j \in 1, 2, \cdots, N$, and $M$ and $N$ represent the height and width of the image, respectively. $C$ is a constant that prevents the denominator from reaching zero. The calculation formulas for $\mu(i,j)$ and $\sigma(i,j)$ are as follows:

$$
\mu(i,j) = \frac{1}{K} \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} l_{k,l}(i,j)
$$

(5)

$$
\sigma(i,j) = \left( \frac{1}{K} \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} (l_{k,l}(i,j) - \mu(i,j))^2 \right)^{1/2}
$$

(6)

where $w_{k,l}(k \in -K, \cdots, K, l \in -L, \cdots, L)$ represents a 2D circular symmetric Gaussian weighting function, sampled to three standard deviations, and then readjusted to unit volume, that is, $K = L = 3$.

2.3.2 Loss function

Stochastic gradient descent and backpropagation are used in the model to approximately minimise the loss during training. Gradient weighting is used, and network parameters are updated. $w_i$ is the $i$th network parameter, and $p_i$ represents the learning efficiency of the $i$th network. $D_{m}^{l}$ represents the gradient of task $m$ corresponding to $w_i$, and $\alpha_m$ represents the relative weight of task $m$. The update rules during the iteration process are as follows:

$$
w_i \leftarrow w_i - p_i \sum_{m=1}^{m=2} \alpha_{m} D_{m}^{l}
$$

(7)

The loss function describes the average absolute error between the predicted value and the target value. The LOSS function can be described as follows:

$$
L(x,y) = L = [l_1, \cdots, l_N]^T
$$

(8)

where $N$ is the number of input samples, $l_N$ is the average absolute error between the predicted value and the target value when the input sample number is $N$, and the specific expression is as follows:

$$
L_N = \frac{1}{N} \sum_{i=1}^{N} \| X_1^{(i)} - X_2^{(i)} \|
$$

(9)

where $X_1^{(i)}$ and $X_2^{(i)}$ are two input vectors, and $N$ is the number of input samples.

2.4 Methods

Apparently, explicit NR-IQA represented by shallow machine network extracts features incompletely, but its physical significance is evident. On the contrary, implicit NR-IQA represented by the deep network extracts features adequately, but its unobvious significance causes information redundancy. Therefore, we combine both of them to extract features and propose the EI-IQA method. The proposed EI-IQA has solved the shortcomings of insufficient explicit features and unclear physical meaning of implicit features. The frequency domain features extracted by wavelet transform represent explicit features, and the spatial domain features extracted by deep network represent implicit features. Extract the explicit and implicit features are in parallel and input the combination of explicit feature and implicit feature to the regression network. Then, we directly obtain the final image quality score.

3 Results and discussion

To verify the performance of the algorithm, we conduct experiments on the LIVE [51], categorical subjective image quality (CSIQ [52]), and TID2013 [53] databases. The basic comparison of the three databases is shown in Table 3. These databases provide a data source for image quality assessment and play an important role. To ensure the consistency of training and testing, when conducting cross-database tests, we only select the same five distortion categories (JP2K, JPEG, Fast-fading, Gaussian noise and Gaussian blur) to carry out the test. Spearman’s rank correlation coefficient (SROCC), Pearson correlation coefficient (PLCC), and distortion type classification Acc indicators are used to assess the performance of the algorithm given that the output of the model involves distortion types and quality scores. To enhance the model comparison effect, in the selection of feature extraction [54] models, SROCC, Kendall rank correlation coefficient (KROCC), root mean square error (RMSE), MSE, exit rate OR, PLCC, and distortion Acc index of type classification are used to assess the performance of the feature extraction model.
Table 3 Comparison of basic database conditions

| Database name | Issuing time | Number of images | Resolution | Types of distortion | No. of distorted images | Quality Index |
|---------------|--------------|------------------|------------|---------------------|------------------------|---------------|
| LIVE          | 2006         | 29               | 438 × 634  | 5                   | 779                    | DMOS          |
| CSIQ          | 2010         | 30               | 512 × 512  | 6                   | 886                    | DMOS          |
| TID2013       | 2013         | 25               | 384 × 512  | 24                  | 3000                   | MOS           |

During the experiment, the implicit feature extraction method was first verified. The VGGNet and ResNet models were used to extract the implicit features, and two models with different implicit features were established, namely, EI-VIQA and EI-RIQA. The comparative training results finally proved that VGGNet can more effectively extract the implicit features when the same explicit features are extracted. Then, to determine the optimal depth of VGGNet in the EI-VIQA model, we conduct EI-VIQA ablation experiment. Finally, EI-VIQA is used as the implicit feature extraction model of the EI-IQA model. At the same time, the ablation supplement experiment is performed on EI-RIQA, which further proved the accuracy of the experimental idea. The experimental results similar to the original deep network can be achieved by adding explicit features to supplement implicit features to reduce the network depth. Thus, the structure of the EI-IQA model has been established.

Then, the assessment index of this experimental algorithm is compared with the classic algorithm. The emphasis of different assessment indexes is slightly different. The comparison shows that a considerable part of the assessment indexes has higher scores of the proposed EI-IQA, proving the superiority of this algorithm. Finally, the established EI-IQA model is used for cross-database training, which is carried out for different distortion types. Among them, JP2K, JPEG, noise, blur, and other distortion types have been better assessed in multiple databases. The results prove that the model has good generalisation performance.

3.1 Model selection of the feature extraction

To extract better implicit features, during the experiment, VGGNet and ResNet models were selected, and the implicit features extracted by VGGNet and ResNet models under the LIVE library were combined with the same explicit features as feature inputs. The implicit feature extraction models EI-VIQA and EI-RIQA are established and compared to obtain a better implicit feature extraction model. The result showed that the implicit features extracted by EI-VIQA achieve better results. Table 4 shows the comparison of the implicit features extracted by the VGGNet and ResNet models under the LIVE library.

In Table 4, the comparison between the implicit feature input extracted by VGGNet and the implicit feature input extracted by the ResNet model evidently shows that the implicit feature extraction of VGGNet has the advantage of small depth and excellent index.

3.2 Quantitative test results

The same library training of the LIVE library is performed on the EI-IQA model determined by the experimental plan. EI-IQA and 18 classic algorithms, including shallow network and deep network, involved in Section 1 are trained together under the LIVE library. To prove the generalisation performance of the EI-IQA algorithm, the algorithm trained under the LIVE library was cross-library trained under the CSIQ and TID2013 databases. The method was also compared with the 18 classic algorithms involved in Section 1. Table 5 shows the IQA performance index comparison table tested on the LIVE database [55]. Table 6 shows the cross-database comparative results of the algorithm across TID2013. Table 7 shows the cross-database comparative results of the algorithm across CSIQ. Figure 4 shows the algorithm in CSIQ [52] and TID2013 [53] cross-library training results.

Tables 5, 6, and 7 show that the proposed EI-IQA still has considerable advantages over the classic algorithm under the CSIQ and TID2013 libraries. It is superior to some algorithm indicators, has good generalisation performance, and satisfies the expected requirements. In Fig. 4, the assessment index results of the EI-IQA algorithm in different databases are compared. The histogram clearly shows that the assessment indexes of the EI-IQA algorithm under different databases are not significantly different. The EI-IQA algorithm has been further proven to have effective generalisation performance and greatly enhanced reliability.

Table 4 Comparison between VGGNet and ResNet extracted features

|       | VGG16  | VGG13  | ResNet50 | ResNet34 |
|-------|--------|--------|----------|----------|
| Acc   | 98.12% | 98.75% | 87.50%   | 92.50%   |
| SROCC | 0.9643 | 0.9585 | 0.9375   | 0.9501   |
| KROCC | 0.8472 | 0.8403 | 0.7939   | 0.8187   |
| PLCC  | 0.9633 | 0.9555 | 0.9343   | 0.9473   |
| RMSE  | 7.2622 | 8.0145 | 9.3961   | 9.4640   |
| MAE   | 5.2481 | 5.5931 | 6.9643   | 7.3811   |
| OR    | 41.88% | 44.38% | 56.25%   | 63.12%   |
Table 5 Test results on LIVE database

| IQA algorithm | ACC  | SROCC | PLCC |
|---------------|------|-------|------|
| BIQI          | -    | 0.827 | 0.842|
| DIVINE        | 0.838| 0.916 | 0.917|
| BLIINDS-II    | -    | 0.912 | 0.916|
| BRISQUE       | 0.886| 0.940 | 0.942|
| DESIQUE       | -    | 0.944 | 0.947|
| NIQE          | -    | 0.915 | 0.915|
| SSEQ          | -    | 0.935 | 0.938|
| IL-NIQE       | -    | 0.903 | 0.909|
| CNNIQA        | -    | 0.956 | 0.953|
| DeepBIQ       | -    | 0.889 | 0.908|
| RankIQA       | -    | 0.981 | 0.982|
| dipIQA        | -    | 0.960 | 0.969|
| BIECON        | -    | 0.961 | 0.962|
| deepIQA       | -    | 0.972 | 0.960|
| MEON          | -    | 0.974 | 0.972|
| BLINDER       | -    | 0.966 | 0.959|
| DIQA          | -    | 0.970 | 0.971|
| DeepVR-IQA    | -    | 0.882 | 0.888|
| **EI-IQA**    | **0.944** | **0.957** | **0.955**|

Table 6 Test results on TID2013 database

| IQA algorithm | ACC  | SROCC | PLCC |
|---------------|------|-------|------|
| BIQI          | -    | 0.809 | 0.809|
| DIVINE        | 0.838| 0.795 | 0.794|
| BLIINDS-II    | -    | 0.980 | 0.935|
| BRISQUE       | 0.886| 0.883 | 0.901|
| DESIQUE       | -    | 0.894 | 0.886|
| NIQE          | -    | 0.900 | 0.819|
| SSEQ          | -    | 0.904 | 0.922|
| IL-NIQE       | -    | 0.881 | 0.890|
| CNNIQA        | -    | 0.920 | 0.903|
| DeepBIQ       | -    | 0.868 | 0.938|
| RankIQA       | -    | 0.780 | 0.770|
| dipIQA        | -    | 0.877 | 0.894|
| BIECON        | -    | 0.721 | 0.765|
| deepIQA       | -    | 0.885 | 0.913|
| MEON          | -    | 0.912 | 0.912|
| BLINDER       | -    | 0.894 | 0.864|
| DIQA          | -    | 0.825 | 0.850|
| DeepVR-IQA    | -    | 0.878 | 0.880|
| **EI-IQA**    | **0.917** | **0.900** | **0.911**|

Table 7 Test results on CIQ database

| IQA algorithm | ACC  | SROCC | PLCC |
|---------------|------|-------|------|
| BIQI          | -    | 0.809 | 0.809|
| DIVINE        | 0.838| 0.835 | 0.855|
| BLIINDS-II    | -    | 0.935 | 0.935|
| BRISQUE       | 0.886| 0.909 | 0.937|
| DESIQUE       | -    | 0.910 | 0.935|
| NIQE          | -    | 0.883 | 0.900|
| SSEQ          | -    | 0.922 | 0.914|
| IL-NIQE       | -    | 0.887 | 0.914|
| CNNIQA        | -    | 0.869 | 0.866|
| DeepBIQ       | -    | 0.890 | 0.908|
| RankIQA       | -    | 0.960 | 0.966|
| dipIQA        | -    | 0.930 | 0.949|
| BIECON        | -    | 0.825 | 0.838|
| deepIQA       | -    | 0.871 | 0.891|
| MEON          | -    | 0.932 | 0.944|
| BLINDER       | -    | 0.894 | 0.765|
| DIQA          | -    | 0.884 | 0.915|
| DeepVR-IQA    | -    | 0.883 | 0.884|
| **EI-IQA**    | **0.917** | **0.903** | **0.911**|

3.3 Model testing for specific distortion types

To test the generalisation ability of the assessment model for different samples, we use the entire LIVE database as the training set and select the same type of distortion as the training sample in the CIQ, and TID2013 databases (JPEG2000, JPEG compression, fast-fading, white noise and Gaussian blur) are used as the test set to obtain the performance index of the algorithm. The SROCC assessment index is used to quantify the adaptability of the algorithm to different types of distortion; it is compared...
with the classic IQA algorithm. The adaptability of the proposed EI-IQA to different types of distortion is different. The comparative results under the LIVE library are shown in Table 8 and that under the CSIQ library are shown in Table 9. The comparative results under the TID2013 library are shown in Table 10.

These tables show that the proposed EI-IQA is more sensitive to distortion types, such as JP2K, JPEG, noise, and blur and slightly less sensitive to fast-fading distortion.

### 3.4 Ablation experiment

The depth of the deep network is reduced by adding explicit features, changing the number of network layers of VGGNet and combining the frequency domain features extracted by wavelet transform. The number of network layers of VGGNet is reduced by adding explicit features to supplement implicit features. The experimental results are similar to the original deep network. In addition, cross-library testing of the model, compared with other experiments, obtained good results. Table 11 shows the performance index of the algorithm tested on the LIVE library, and Fig. 5 shows a histogram of the algorithm performance indicator tested on the LIVE library. The SROCC, PLCC, and Acc type of the classification of the proposed EI-IQA are the best among similar algorithms.

In Table 11, the depth of the VGG network is reduced, and then combined with the explicit features of different degrees. No evident fluctuation in the assessment index results was found. This experiment aims to reduce the depth of the deep network by adding explicit features. In Fig. 5, part of the results of Table 11 are visualised, indicating that the depth of the VGG network is reduced. In addition, combined with the explicit features, the results of the assessment indicators do not change significantly.

To further prove the experimental idea, the implicit feature extraction model is replaced. The ResNet model is used to extract implicit features. The implicit features extracted by the ResNet model are supplemented by adding frequency domain features, which can achieve similar final indicators whilst reducing the depth of the ResNet model. This finding further proves the experimental idea, that is, adding frequency-domain features to supplement implicit features reduces the depth of the model and the difficulty of parameter training. Table 12 shows the performance index of the ResNet model under the LIVE library, and Fig. 6 shows the histogram of the performance index of the ResNet model under the LIVE library.

| Table 8 Test results of specific distortion types under the LIVE database |
|---------------------|--------|--------|-------|-------|-------|--------|
| Algorithm         | JP2K   | JPEG   | Noise | Blur  | FF    | All    |
| PSNR              | 0.868  | 0.885  | 0.943 | 0.761 | 0.891 | 0.866  |
| SSIM              | 0.938  | 0.947  | 0.964 | 0.907 | 0.956 | 0.913  |
| VIF               | 0.952  | 0.910  | 0.984 | 0.972 | 0.963 | 0.952  |
| BICQ              | 0.802  | 0.874  | 0.958 | 0.821 | 0.730 | 0.824  |
| DIVINE            | 0.913  | 0.910  | 0.984 | 0.921 | 0.863 | 0.916  |
| BLIINDS-II        | 0.951  | 0.941  | 0.978 | 0.944 | 0.927 | 0.920  |
| BRISQUE           | 0.947  | 0.925  | 0.989 | 0.951 | 0.903 | 0.947  |
| SSSEQ             | 0.946  | 0.951  | 0.978 | 0.948 | 0.904 | 0.935  |
| **EI-IQA**        | **0.974** | **0.986** | **0.992** | **0.985** | **0.874** | **0.958** |

| Table 9 Test results of specific distortion types under the CSIQ database |
|---------------------|--------|--------|-------|-------|-------|--------|
| Algorithm         | JP2K   | JPEG   | Noise | Blur  | FF    | All    |
| PSNR              | 0.941  | 0.901  | 0.943 | 0.936 | 0.915 | 0.928  |
| SSIM              | 0.962  | 0.956  | 0.912 | 0.965 | 0.937 | 0.935  |
| VIF               | 0.967  | 0.970  | 0.957 | 0.975 | 0.963 | 0.972  |
| BICQ              | 0.975  | 0.962  | 0.954 | 0.968 | 0.936 | 0.927  |
| DIVINE            | 0.831  | 0.800  | 0.866 | 0.872 | 0.891 | 0.864  |
| BLIINDS-II        | 0.895  | 0.900  | 0.801 | 0.892 | 0.868 | 0.892  |
| BRISQUE           | 0.867  | 0.909  | 0.925 | 0.903 | 0.859 | 0.916  |
| SSSEQ             | 0.977  | 0.975  | 0.983 | 0.961 | 0.904 | 0.955  |
| **EI-IQA**        | **0.954** | **0.938** | **0.969** | **0.911** | **0.957** | **0.936** |

| Table 10 Test results of specific distortion types under the TID2013 database |
|---------------------|--------|--------|-------|-------|-------|--------|
| Algorithm         | JP2K   | JPEG   | Noise | Blur  | FF    | All    |
| PSNR              | 0.898  | 0.929  | 0.942 | 0.965 | 0.865 | 0.924  |
| SSIM              | 0.950  | 0.935  | 0.896 | 0.969 | 0.945 | 0.924  |
| VIF               | 0.970  | 0.931  | 0.913 | 0.958 | 0.953 | 0.946  |
| BICQ              | 0.974  | 0.925  | 0.833 | 0.847 | 0.810 | 0.889  |
| DIVINE            | 0.853  | 0.629  | 0.855 | 0.834 | 0.898 | 0.880  |
| BLIINDS-II        | 0.911  | 0.838  | 0.715 | 0.826 | 0.899 | 0.893  |
| BRISQUE           | 0.904  | 0.910  | 0.823 | 0.874 | 0.848 | 0.896  |
| SSSEQ             | 0.846  | 0.866  | 0.801 | 0.835 | 0.853 | 0.846  |
| **EI-IQA**        | **0.925** | **0.934** | **0.934** | **0.895** | **0.895** | **0.936** |

| Table 11 Comparison of VGGNet degrading results |
|---------------------|--------|--------|--------|--------|--------|--------|
| VGG16               | **98.12%** | **98.75%** | **92.50%** | **94.38%** | **94.38%** |
| VGG13               | **98.75%** | **99.15%** | **94.38%** | **95.96%** | **95.96%** |
| VGG11               | **99.15%** | **99.37%** | **95.96%** | **96.55%** | **96.55%** |
| VGG9                | **99.37%** | **99.37%** | **95.96%** | **96.55%** | **96.55%** |

| **Acc**            | **98.12%** | **98.75%** | **92.50%** | **94.38%** |
| **SROCC**          | **0.9643** | **0.9585** | **0.9596** | **0.9566** |
| **KROCC**          | **0.8472** | **0.8403** | **0.8414** | **0.8341** |
| **PLCC**           | **0.9633** | **0.9555** | **0.9558** | **0.9554** |
| **RMSE**           | **7.2622** | **8.0145** | **7.9754** | **8.0061** |
| **MAE**            | **5.2481** | **5.5931** | **5.5643** | **5.5833** |
| **OR**             | **41.88%** | **44.38%** | **46.25%** | **47.50%** |
In Table 12, the depth of the ResNet network is reduced, and then different levels of explicit features are combined. No evident fluctuation is found in the assessment index results. As a supplementary experiment, to further prove the idea of the proposed EI-IQA, explicit features are added to reduce the depth of the deep network. In Fig. 6, part of the results of Table 6 are visualised. The image shows that the depth of the ResNet network is reduced, and then combined with the explicit features, the results of the assessment indicators do not fluctuate significantly.

Since then, the structure of the EI-IQA algorithm has been established, that is, VGGNet is used to extract implicit features, and wavelet transform is used to extract wavelet domain and frequency domain features to represent explicit features. Explicit features supplement implicit features, thereby appropriately reducing the depth of VGGNet model. The experimental results achieved are close to the original network depth. The extracted explicit features and the implicit features are subjected to feature fusion to extract mixed features, and then the mixed features are input into the regression network [42] established for multitask learning to obtain the final quality score and distortion category.

3.5 Discussion

Most of the IQA methods have a large demand for samples. Compared with classical, the proposed EI-IQA combines explicit and implicit features. The explicit features complement the implicit features, reducing the network depth and solving the problem of large sample demand for traditional algorithms. The proposed EI-IQA constructs two different approaches to extract implicit features and chooses the VGGNet as the better one. In future work, we hope to further optimise the extraction of implicit features and combine with explicit features to obtain better mixed features. The generalisation ability of the model is improved by the model structure; it attempts to be closer to the subjective feeling of human eyes.

4 Conclusion

We build an EI-IQA method to extract features. From two different approaches, combining explicit and implicit features to describe image features makes up for the shortcomings of insufficient explicit features and unclear physical meaning of implicit features. The proposed EI-IQA model avoids the loss of feature information. The original image is directly fed into the model for training, and the generalisation ability of the model is effectively improved. Explicit features complement implicit features, considering mixed features as input. Thus, the dependence of deep networks on large training samples is effectively reduced.

We also construct two different approaches to extract implicit features. VGGNet and ResNet are used to extract implicit features. Our results suggest that VGGNet extracts features better. Then, we test the proposed EI-IQA over three different databases. Compared with some classical, the proposed EI-IQA obtains better scores in some components. We believe that the proposed EI-IQA has more effective generalisation performance and the reliability is greatly enhanced.
Abbreviations
AGGD: Asymmetric generalised Gaussian distribution; BIOA: Blind image quality assessment; CNN: Convolutional neural network; CSIQ: Categorical subjective image quality; DCT: Discrete cosine transform; DeepPR-IQA: Virtual reality IQA method based on deep learning; DMOS: Differential mean opinion score; DWT: Discrete wavelet transform; E: Explicit-implicit; E-IQA: Explicit-implicit dual stream network for image quality assessment; GGD: Generalised Gaussian distribution; IQA: Image quality assessment; KROCC: Kendall rank correlation coefficient; LIVE: Laboratory for image and video engineering; MOS: Mean opinion score; MSCN: Mean subtracted contrast normalised; MVG: Multivariate Gaussian; NR-IQA: Non-reference IQA; PCA: Principal component analysis; PLCC: Pearson correlation coefficient; RMSE: Root mean square error; SRCC: Spearman’s rank correlation coefficient

Acknowledgements
The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University. The authors would like to thank Prof. Hongyan Zhang and Prof. Jiang Hao for the valuable opinions they have offered during our heated discussions.

Authors’ contributions
XY conducted the experiments and drafted the manuscript. HT and CK implemented the core method and performed the statistical analysis. GY designed the methodology. WZ modified the manuscript. All authors read and approved the final manuscript.

Funding
This study is partially supported by National Natural Science Foundation of China (NSFC) (No. 61871298, 61671333) and National Key Research and Development Program of China (No. 2018YF 80504001, No. 2018YF B1201602).

Availability of data and materials
The Python source code of E-IQA can be downloaded at https://github.com/jacob6/EI-IQA for public use and evaluation. You can change this program as you like and use it anywhere, but please refer to its original source.

Competing interests
The authors declare that they have no competing interests.

Received: 10 August 2020 Accepted: 12 October 2020
Published online: 27 October 2020

References
1. C. Luo, J. Ji, Q. Wang, X. Chen, P. Li, Channel state information prediction for 5G wireless communications: a deep learning approach. IEEE Trans. Netw. Sci. Eng. 7(1), 227–236 (2020)
2. C. Yan, B. Shao, H. Zhao, R. Ning, Y. Zhang, F. Xu, 3D room layout estimation from a single RGB image. IEEE Trans. Multimedia, 1–1 (2020)
3. K. Gu, G. Zhai, X. Yang, W. Zhang, M. Liu, in 2013 IEEE International Conference on Image Processing (No. 2018YF B1201602).
4. Z. Wang, A. C. Bovik, Modern image quality assessment (2006)
5. Fan Zhang, Yuli Xu, in 2009 Chinese Control and Decision Conference, vol. 14. Image quality evaluation based on human visual perception, (2009), pp. 1487–1490
6. A. C. Bovik, Automatic predition of perceptual image and video quality. Proc. IEEE. 101(9), 2008–2024 (2013)
7. Z. Wang, A. C. Bovik, Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Signal Proc. Mag. 26(1), 98–117 (2009)
8. Z. Wang, Applications of objective image quality assessment methods [applications corollary]. IEEE Signal Proc. Mag. 28(6), 137–142 (2011)
9. G. J. Katuwal, J. Kerekes, R. Ramchandran, C. Sisson, N. Rao, in 2013 IEEE Western New York Image Processing Workshop (WNYIPW), vol. 37. Automatic explisit-implicit image field correction and quality assessment, (2013), pp. 9–13
10. Zhou Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004)
11. C. Luo, J. Ji, H. Zeng, D. Ghassemi, S. Lee, L. Zhang, A. C. Bovik, Deep convolutional neural models for picture-quality prediction: challenges and solutions to data-driven image quality assessment. IEEE Signal Proc. Mag. 34(6), 130–141 (2017)
12. A. K. Moorthy, A. C. Bovik, A two-step framework for constructing blind image quality indices. IEEE Signal Proc. Lett. 17(5), 513–516 (2010)
13. A. K. Moorthy, A. C. Bovik, Blind image quality assessment from natural scene statistics to perceptual quality. IEEE Trans. Image Process. 20(12), 3350–3364 (2011)
14. M. Unser, N. Chenouard, D. Van De Ville, Steerable pyramids and tight wavelet frames in $l_2(R^d)$. IEEE Trans. Image Process. 20(10), 2705–2721 (2011)
15. M. Wainwright, O. Schwartz, E. Simoncelli, Natural image statistics and divisive normalization: modeling nonlinearity and adaptation in cortical neurons. Stat. Theor. Brain, 203–222 (2002)
16. M. A. Saad, A. C. Bovik, C. Charrier, A DCT statistics-based blind image quality index. IEEE Signal Proc. Lett. 17(6), 583–586 (2010)
17. M. A. Saad, A. C. Bovik, C. Charrier, in 2011 18th IEEE International Conference on Image Processing, vol. 11. DCT statistics model-based blind image quality assessment, pp. 3093–3096
18. N. Ahmed, T. Natarajan, K. R. Rao, Discrete cosine transform. IEEE Trans. Comput. C-23(1), 90–93 (1974)
19. A. Mittal, A. K. Moorthy, A. C. Bovik, A non-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. 21(12), 4665–4708 (2012)
20. D. L. Ruderman, Statistics of natural images. Netw. Comput. Neural Syst. 5(4), 517–548 (1994)
21. K. Gu, G. Zhai, X. Yang, W. Zhang, Using free energy principle for blind image quality assessment. IEEE Trans. Multimed. 17(1), 50–63 (2015)
22. A. Mittal, A. K. Moorthy, A. C. Bovik, Non-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. 21(12), 4665–4708 (2012)
23. A. Mittal, Fellow, IEEE, R. Soundararajan, A. C. Bovik, Making a ‘completely blind’ image quality analyzer. IEEE Signal Proc. Lett. 20(3), 209–212 (2013)
24. L. Zhang, L. Zhang, A. C. Bovik, A feature-enriched completely blind image quality evaluator. IEEE Trans. Image Process. 24(8), 2579–2591 (2015)
25. Z. Yi, D. M. Chandler, No-reference image quality assessment based on log-derivative statistics of natural scenes. J. Electron. Imaging. 22(4), 043025–104302522 (2013)
26. Y. Liu, B. Liu, H. Huang, A. C. Bovik, Noise-reference image quality assessment based on spatial and spectral entropies. Sig. Process. Image Commn. (2014)
27. K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, W. Zuo, End-to-end blind image quality assessment using deep neural networks. IEEE Trans. Image Process. 27(3), 1202–1213 (2018)
28. W. Hou, X. Gao, D. Tao, X. Li, Blind image quality assessment via deep learning. IEEE Trans. Neural Netw. Learn. Syst. 26(6), 1275–1286 (2015)
29. L. Kang, P. Ye, Y. Li, D. Doermann, in 2014 IEEE Conference on Computer Vision and Pattern Recognition, vol. 14. Convolutional neural networks for no-reference image quality assessment, pp. 1733–1740
30. R. Girshick, J. Donahue, T. Darrell, J. Malik, in 2014 IEEE Conference on Computer Vision and Pattern Recognition, vol. 34. Rich feature hierarchies for accurate object detection and semantic segmentation, (2014), pp. 580–587
31. L. Kang, P. Ye, Y. Li, D. Doermann, in 2015 IEEE International Conference on Image Processing (ICIP). Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks, (2015), pp. 2791–2795
32. S. Bianco, L. Celona, P. Napoletano, in Signal, Image and Video Processing, vol. 12. On the use of deep learning for blind image quality assessment, pp. 355–362
33. X. Liu, J. Van De Weijer, A. D. Bagdanov, in 2017 IEEE International Conference on Computer Vision (ICCV), vol. 18. Rankiq: learning from rankings for no-reference image quality assessment, (2017), pp. 1040–1049
34. S. Bosse, D. Mansy, K. Müller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment. IEEE Trans. Image Process. 27(1), 206–219 (2018)
35. J. Kim, S. Lee, Fully deep blind image quality predictor. IEEE J. Sel. Top. Sig. Process. 11(1), 206–220 (2017)
36. J. Kim, A. Nguyen, S. Lee, Deep CNN-based blind image quality predictor. IEEE Trans. Neural Netw. Learn. Syst. 30(1), 11–24 (2019)
37. K. Ma, W. Liu, T. Liu, Z. Wang, D. Tao, dipic: blind image quality assessment by learning-to-rank discriminable image pairs. IEEE Trans. Image Process. 26(8), 3951–3964 (2017)
38. C. Burges, T. Shaked, E. Renshaw, Learning to rank using gradient descent. Proc. 22nd Int. Conf. Mach. Learn. 11(10), 89–96 (2005)
39. Z. Wang, E. P. Simoncini, A. C. Bovik, in The Thirty-Seventh Asilomar Conference on Signals, Systems Computers, 2003, vol. 2. Multiscale structural similarity for image quality assessment, (2003), pp. 1398–14022
40. H. R. Sheikh, A. C. Bovik, Image information and visual quality. IEEE Trans. Image Process. 15(2), 430–444 (2006)
41. W. Xue, L. Zhang, X. Mou, A. C. Bovik, Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. IEEE Trans. Image Process. 23(2), 684–695 (2014)
42. H. R. Sheikh, M. F. Sabir, A. C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Process. 15(11), 3440–3451 (2006)
43. K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, W. Zuo, End-to-end blind image quality assessment using deep neural networks. IEEE Trans. Image Process. 27(3), 1202–1213 (2018)
44. G. Fei, J. Yu, S. Zhu, Q. Huang, T. Qi, Blind image quality prediction by exploiting multi-level deep representations. Pattern Recog. 81, 432–442 (2018)
45. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. Proc. ImageNet Chal. 1409(15), 1–10 (2014)
46. H. G. Kim, H. Lim, Y. M. Ro, Deep virtual reality image quality assessment with human perception guider for omnidirectional image. IEEE Trans. Circ. Syst. Video Technol. 30(4), 917–928 (2020)
47. C. Yan, Z. Li, Y. Zhang, Y. Liu, X. Ji, Y. Zhang, Depth image denoising using nuclear norm and learning graph model (2020)
48. M. Bournahd, J. Lacoume, in 1995 International Conference on Acoustics, Speech, and Signal Processing, vol. 3. Blind identification using the kurtosis: results of field data processing, (1995), pp. 1980–19833
49. M. J. Shensa, The discrete wavelet transform: wedding the a trous and Mallat algorithms. IEEE Trans. Sig. Process. 40(10), 2464–2482 (1992)
50. D. J. Field, Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A-optics Image Sci. Vision. 4(12), 2379–2394 (1987)
51. H. R. Sheikh, A. C. Bovik, L. Cormack, No-reference quality assessment using natural scene statistics: JPEG2000. IEEE Trans. Image Process. 14(11), 1918–1927 (2005)
52. E. C. Larson, D. M. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy. J. Electron. Imaging. 19(1), 011006 (2010)
53. N. Ponomarenko, L. Jin, O. Ieremeiev, Image database TID2013: Peculiarities, results and perspectives. Sig. Process. Image Commun. 30, 57–77 (2015)
54. C. Yan, B. Gong, Y. Wei, Y. Gao, Deep multi-view enhancement hashing for image retrieval. arXiv. 29(10), 1–1 (2020)
55. H. R. Sheikh, Live image quality assessment database release 2.online (2005). http://live.ece.utexas.edu/research/quality

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.