REGULAR PAPER

Full reference image quality assessment based on dual-space multi-feature fusion

Xinrong Wu[1] · Zhiming Shi[1]

Abstract

At present, the majority of techniques for assessing image quality are limited to extracting features from an image in a single space. This paper proposes a new dual-space multi-feature fusion based method for full-reference image quality assessment. This method involves simultaneously extracting features from both the YIQ and $L^*a^*b^*$ color spaces. First, we extract the luminance, slope, chroma, and gradient features in the spatial domain of the image to describe the salient differences in the image. Second, based on contrast sensitivity characteristics, we extract spatial frequency features in the spatial domain of the image to represent frequency differences in the image. Next, merge the features extracted in the dual space to construct a quality perception feature vector. Finally, the feature vector is input into the Random Forest model for regression prediction to obtain the predicted score of the image. Many experiments have been carried out on the four public datasets, and contrasted with other methods. The experimental confirm that the proposed method predicts image quality more accurately. The MATLAB source code and dataset of this paper will be published on GitHub, and the corresponding author can be contacted if necessary.

1 Introduction

Recently, with the rapid development of internet technology, many kinds of media formats have filled people's daily lives, and the sharing and transmission of images and videos are more frequent [1]. However, during the process of image capture and transmission, factors such as transmission protocols and signal interference can lead to image distortion, resulting in a decrease in image quality that affects user perception. To improve the consistency of visual perception of image quality and meet the needs of observers, it is of great value to study the degradation of images generated during different processing stages [2]. In terms of image processing, image quality assessment (IQA) is important. Subjective and objective methods are the two categories of IQA methods, which vary depending on the assessment criteria and application circumstances.

The subjective method assesses the quality of image directly by observers. Subjective assessment is the most reliable assessment method. The most commonly used subjective evaluation methods include double-stimulus impairment scale (DSIS), double-stimulus continuous quality scale (DSCQS), and single-stimulus continuous quality scale (SSCQS). However, it is frequently costly and time-consuming [3]. Thus, in order to quantify visual quality efficiently, objective methods that have a strong correlation with subjective scores must be developed [4].

The objective method uses the mathematical models to assess the image quality. Objective methods include full-reference (FR), reduced-reference (RR), and no-reference (NR) image quality assessment.

The FR-IQA methods fully utilize the original images to assess the quality of distorted images [5]. Wen Sun et al. proposed an FR-IQA method based on superpixel similarity index, it uses three metrics to assess the quality of an image: pixel gradient similarity, superpixel luminance similarity, and superpixel chrominance similarity [6]. Although it improves calculation accuracy, its computational complexity

✉ Zhiming Shi
szmi_2007@126.com

1 School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang, China

is high. Junfeng Yang et al. proposed a diffusion speed structure similarity index for FR-IQA. This method calculates image similarity by considering both intra-block structures and inter-block textures, thereby deriving the quality score of the image [7]. Kyohoon Sim et al. introduced a method for assessing image quality, named the deep and local similarity method. This technique uses 2D full-reference analysis to evaluate the similarity between original and deformed deep feature maps obtained from convolutional neural networks. The mean and deviation of these similarity measures are reported as key findings [8]. This method uses the mean and standard deviation to reflect the influence of visual saliency and the distribution of image distortions on quality, respectively [9]. It has a good pooling method, but its computational complexity is high. Zihan Zhou et al. proposed an FR-IQA method which constructed a kernel dictionary and introduced nonlinear sparse coding to IQA. While this method is proficient in analyzing various types of distortions in a higher-dimensional feature space, its generalization capability is limited [10]. Keyan Ding et al. proposed a method that combines the correlations of spatial averages with correlations of the feature maps [11]. This technique explains human perception scores on texture datasets as well as traditional image quality datasets. Dong Wu et al. proposed an FR-IQA method based on multi-scale and multi-directional visibility differences. This method considers the visibility differences and contrast sensitivity functions in the discrete and non-separable shear transform domain, as well as visual masking effects. It evaluates all sub-bands of the shear wavelet transform and combines the perceptual errors of these sub-bands to obtain an objective quality metric for distorted images [12]. This method maintains moderate computational complexity, but its generalization capability needs to be improved. Ke Gu et al. introduced a novel method for assessing the quality of perceptual images by leveraging the capabilities of the human visual system (HVS). This method efficiently applies convolution operations across several scales, taking into account factors such as gradient magnitude, similarity of color information, and perception-based pooling [13]. The advantage of FR-IQA lies in its higher accuracy, as it is based on the comparative analysis between the original image and the deformed image being evaluated, enabling an accurate assessment of the degree of image distortion.

The RR-IQA method refers to using partial original image information as a reference for quality assessment. Mengzhu Yu et al. introduced a novel perceptual hashing method that incorporates the use of complementary color wavelet transform (CCWT) and compressed sensing (CS) for the purpose of RR-IQA. The CCWT is used to decompose the input color image into various sub-bands, while the block-based CS technique is utilized to extract relevant features from these CCWT sub-bands [14]. Wenhan Zhu et al. proposed a free energy based RR-IQA metric inspired by the principle of free energy. This method involves decomposing the image using wavelet transformation, extracting the free energy features of sub-band images using coefficient matrices, and finally using support vector regression to assess image quality [15].

The NR-IQA refers to the method of evaluating image quality solely by analyzing the deformed image's features, without relying on information from the original image [16]. It has strong current development potential and will be obtained in the future [17]. Xiaohan Yang et al. proposed a new transfer learning method for NR-IQA, which can effectively alleviate the overfitting issue [18]. Lixiong Liu et al. introduced a novel NR-IQA metric that considers the influence of pre-attention and spatial dependency on the perceived quality of distorted images. The proposed model, known as the pre-attention and spatial-dependency driven quality assessment predictor, incorporates the pre-attention theory to simulate early phase visual perception by enhancing luminance-channel data [19]. Yang Wen et al. proposed an unsupervised image deblurring method for fuzzy images, which is based on multi-adversarial optimization cycles to uniformly generate adversarial networks. This method enhances the structure and detail retention capability of multi-adversarial networks by introducing a sensing mechanism [20]. Guanghui Yue et al. proposed a new NF-IQA method, named TANet. This method embeds a texture enhancement module in the shallow layers to evaluate facial images by considering texture artifacts. Experimental results on the constructed SZU-RFD benchmark dataset show that the method achieves high accuracy [21]. To address the current lack of fair comparisons in assessing the performance of LFI stitching methods, Yueli Cui et al. built the first stitched WLFI dataset and proposed a blind stitching WLFI quality metric to assess the visual quality. Compared to other quality metrics, the proposed metric demonstrates excellent performance [22]. Zhewei Fang et al. proposed a robust blind metric. This method captures local statistical features to capture local texture variations due to significant local texture degradation engendered by the DIBR procedure. Additionally, global features of the image are extracted to characterize overall blurriness. This metric outperforms the newly developed 3D-synthesized image metrics [23]. The NR-IQA typically involves the use of machine learning or deep learning techniques to classify or predict based on the features of distorted images. Nowadays, due to the availability of computational power and large sets of labeled training images, a lot of researches are going in the field of image quality assessment which utilizes Deep Learning methods [24]. Learning-based methods do achieve better performance than hand-crafted-based methods in certain fields [25]. In contrast, partial reference

assessment methods only require partial reference image information to assess quality by analyzing partial image features. The accuracy of NR-IQA is typically lower than that of FR-IQA because it does not rely on original image information and may be influenced by the type and degree of distortion.

The results of FR-IQA are compared based on the original image, thus allowing for comparability across different times, locations and devices. The FR-IQA can more accurately assess the degree of image distortion. Therefore, this paper introduces a dual-space multi-feature fusion-based method for FR-IQA.

The representation of images in different spaces has distinct features and application scenarios. Simultaneously extracting image features from these two color spaces can comprehensively utilize their advantages, construct more feature representation, and improve the precision and robustness of image quality assessment. Firstly, extracting the luminance, slope, chroma, gradient, and spatial frequency features of the image in both the YIQ space and $L^*a^*b^*$ space. Next, the extracted features are combined to form a feature vector. Finally, the Random Forest regression model is used to predict the image quality.

In this paper, a method based on dual-space multi-feature fusion is proposed to full-reference image quality assessment. The method calculates the similarity of two images by extracting the chroma, luminance, slope, gradient, and spatial frequency features of the images. On one hand, compared to other color spaces of an image, the YIQ color space and $L^*a^*b^*$ color space of an image separate chroma information from luminance information. This allows for independent extraction of chroma and luminance features during image processing, increasing the flexibility in image manipulation. On the other hand, both the YIQ and $L^*a^*b^*$ color spaces exhibit high chroma uniformity, meaning that chroma changes within the same distance are relatively consistent. This characteristic enhances stability and reliability when conducting chroma analysis and feature extraction. Currently, most methods extract features in a single space, where the information displayed in a single space is limited, and images show different information feature in different spaces. Inspired by this, this paper further improves the accuracy of evaluation by extracting the relevant features of images in the dual-space of YIQ and $L^*a^*b^*$. After that, the extracted features are fused into a feature vector, which is input into the Random Forest for regression prediction. The dual-space feature extraction method and the slope feature extraction method proposed in this paper are a novel method. The main contributions of this paper are as follows:

We propose a new method for dual-space feature extraction, which goes beyond the method of feature extraction in a single space. In different spaces, images contain different information. As a result, our method is able to take full advantage of more information from the image.

We introduce a new feature of image, i.e., slope. In remotely sensed terrain images, the slope represents the undulating variation of the image. In non-topographic maps, the texture information of the image is reflected by extracting the slope features of the image.

This paper is structured in the following manner: Sect. 1 introduces the methods and relevant concepts of IQA, proposing an FR-IQA method. Section 2 provides an overview of the conceptual structure of the method model. Section 3 focuses on the necessary preprocessing of images before conducting image feature extraction, as well as the process and relevant computations of image feature extraction. Section 4 summarizes the extracted features, performs feature fusion, and introduces the Random Forest model as the primary tool for data processing. Additionally this paper introduces the validation of the proposed method in four public image datasets, comparing it with several mainstream and currently popular methods. It also includes feature analysis, model performance analysis, and sample size analysis. Section 5 provides a summary of this paper, elaborating on the innovative aspects and future work.

2 Method model

Given that features extracted in a single space often cannot fully describe all information of image and image have many features in different space, this paper proposes a dual-space multi-feature fusion based method to FR-IQA to further describe the internal in formation of the images. The model aims to extract features from two different spaces of the image simultaneously to construct a more comprehensive representation of image features. Considering human visual perception, we extract luminance and chroma features to assess color differences, and further extracts slope features based on the luminance features. In addition, distorted images often disrupt the image structure, so gradient features are extracted to describe structural differences in the image [26]. Extracting spatial frequency features reflects the visual differences in the image. The Fig. 1 illustrates the overall framework of the method proposed in this paper. Where SY and SL respectively represent the luminance similarity calculated from the YIQ space and the $L^*a^*b^*$ space. Similarly, SP1 and SP2 represent the slope similarity, SC1 and SC2 represent the chroma similarity, SG1 and SG2 represent the gradient similarity. Additionally, SH1, SM1, SL1 as well as SH2, SM2, SL2 represent the frequency similarity calculated from the YIQ space and the $L^*a^*b^*$ space. After extracting these 14 features, they are combined into a feature vector, which is then used to generate a dataset along with the subjective scores. Next, the dataset is partitioned into testing and training subsets. In this paper, decision trees

Fig. 1 Overall framework of the method in this paper
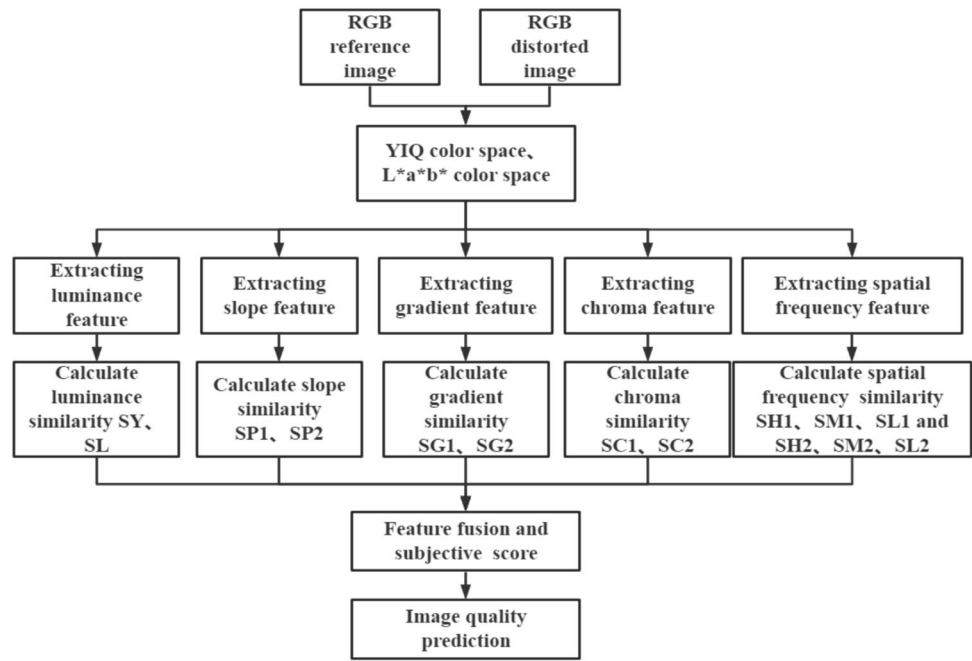
are used as regressors to construct a Random Forest model. The training set is inputted into the Random Forest for training. Finally, the proposed model is used to predicting image quality.

3 Image feature extraction

3.1 Image preprocessing

In image processing, preprocessing the input image is a common method. Xiao Lin et al. proposed the division of images into blocks and devised an encoding and decoding communication module to capture communication information among all image blocks [27]. Images have different features in different color spaces. As illustrated in the Fig. 1, to establish a more comprehensive representation of image features, the model described in this paper all perform feature extraction in both the YIQ and $L^*a^*b^*$ color spaces. Most assessment images are in the RGB color space. Therefore, before conducting feature extraction, it is necessary to convert the two images into the YIQ and $L^*a^*b^*$ color spaces, respectively. Due to the inability to directly convert RGB images to the $L^*a^*b^*$ color space, conversion via the XYZ color space is necessary. The conversion from the RGB space to YIQ space and $L^*a^*b^*$ space is detailed in formulas (1), (2), and (3):

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412 & 0.357 & 0.180 \\ 0.212 & 0.715 & 0.072 \\ 0.019 & 0.119 & 0.950 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} L^* \\ a^* \\ b^* \end{bmatrix} = \begin{bmatrix} 3.240 & -1.537 & -0.498 \\ -0.969 & 1.875 & 0.041 \\ 0.055 & -0.204 & 1.507 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3}$$

where Y and L^* represent the luminance channels of the image, while I, Q, a^* and b^* represent the image's chrominance channels.

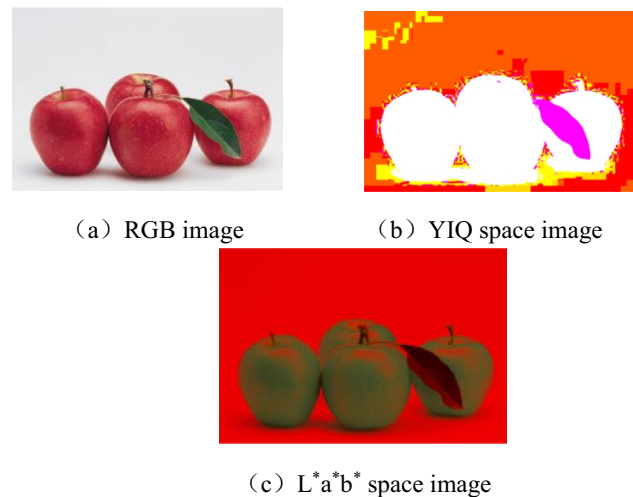（a）RGB image   （b）YIQ space image

（c）$L^*a^*b^*$ space image

Fig. 2 Images in different color spaces

Different color spaces have different features. According to the Fig. 2, (a) is an RGB image, Fig. 2(b) represents the image in the YIQ space, and Fig. 2(c) represents the image in the $L^*a^*b^*$ space.

3.2 Luminance feature

The luminance feature is one of the significant features of an image, which reflects the overall luminance and contrast, typically represented by the magnitude of pixel values. When an image experiences luminance distortion, this uneven distribution of luminance causes certain areas to appear darker or brighter, thereby affecting the overall visual impact of the image.

This paper extracts the luminance feature from the image's luminance channel to assess the degree of luminance distortion in the image. Firstly, in the image's YIQ space and $L^*a^*b^*$ space, the Y and L^* channels represent the image's luminance channels. Subsequently, the Y channel and L^* channel are used to extract the luminance features from both the deformed image and the reference image. Next, the similarity calculation formula is applied to calculate the luminance similarity SY and SL from the images in both the YIQ space and $L^*a^*b^*$ space, respectively. The luminance similarity calculation formulas are as follows:

$$SY = \frac{1}{N} \sum_x \frac{2Y_R(x) \cdot Y_D(x) + C_1}{Y_R^2(x) + Y_D^2(x) + C_1} \tag{4}$$

$$SL = \frac{1}{N} \sum_x \frac{2L_R(x) \cdot L_D(x) + C_1}{L_R^2(x) + L_D^2(x) + C_1} \tag{5}$$

where the luminance information in the Y channel of the original image is represented by Y_R, and that of the distorted image is represented by Y_D; In the L channel, L_R and L_D stand for the luminance information of both images, respectively. To avoid calculation errors due to a zero denominator, C_1 is introduced as a constant. SY and SL represent the luminance similarity between the original and deformed images in YIQ space and $L^*a^*b^*$ space, respectively.

Through the extraction of the image's luminance features and the subsequent calculation of the similarity, we can obtain the degree of resemblance between the two images. According to the Fig. 3, Fig. 3(a) is the original image, while Fig. 3(b) and (c) represent different levels of distortion under the same type of distortion (Level 1 to 5, level 1 indicates the least distortion degree, level 5 indicates the greatest distortion degree). The Fig. 3(b) represents distortion level 1, with a luminance similarity of 0.3331 in the YIQ space and 0.2623 in the $L^*a^*b^*$ space compared to the reference image. The Fig. 3(c) represents distortion level 5, with a luminance similarity of 0.3268
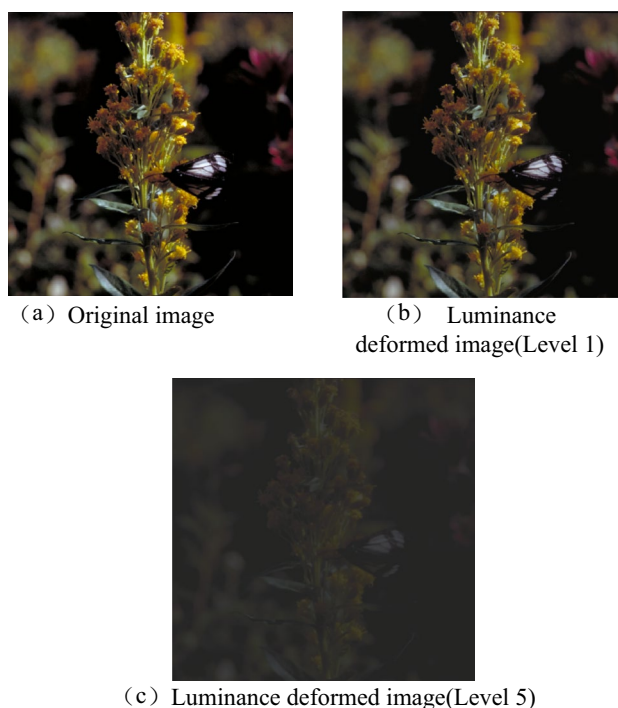

（a）Original image

（b）Luminance deformed image(Level 1)

（c）Luminance deformed image(Level 5)

Fig. 3 Comparison of luminance deformed images at different levels

in the YIQ space and 0.1852 in the $L^*a^*b^*$ space compared to the reference image. Therefore, the larger values of SY and SL, the smaller difference in luminance between the two images, indicating a higher degree of similarity in luminance and better image quality.

3.3 Slope feature

The slope feature is a method to describe local variations in the image, measuring the degree of local changes at each pixel in the image. If the image undergoes distortion, it will result in greater local variations in the image. The slope maps corresponding to the original image and the deformed image are shown in the Fig. 4. Therefore, this paper introduces the concept of slope into non-topographic images, enabling the extraction of slope features. This will enhance the assessment of image quality's accuracy.

First, after extracting the luminance features from the image, we obtain the luminance feature maps, then extract slope features from luminance feature maps. For a pixel of an input image with two dimensions: horizontal and vertical, let the function be $f(x, y)$. The slope at point (x_i, y_i) is calculated by computing the derivatives $f'(x_i)$ in the x-direction and $f'(y_i)$ in the y-direction. Then, we calculate the angle from the positive x-axis to the direction of the slope using the following formula:
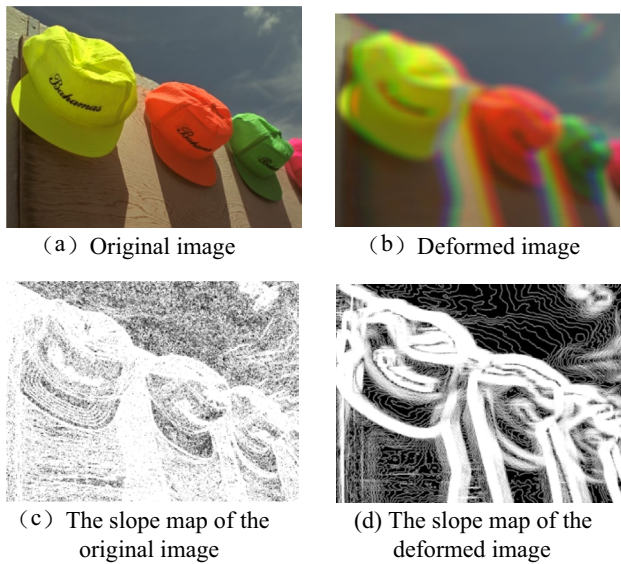
（a）Original image   （b）Deformed image

（c）The slope map of the   (d) The slope map of the
original image   deformed image

Fig. 4 Slope maps corresponding to the original image and its deformed images

Table 1 Comparison of slope similarity for different levels of distortion

Metric	1	2	3	4	5
SP_1	0.3333	0.3333	0.3331	0.3330	0.3329
SP_2	0.3308	0.3280	0.3030	0.2910	0.2854

$$Slope = \arctan\left(\frac{f\prime(y_i)}{f\prime(x_i)}\right) \tag{6}$$

where *arctan()* represents the arctangent function. Finally, the slope features extracted from the original image and the deformed image are used to compute the similarity of slopes, with the following formula:

$$SP_i = \frac{1}{N}\sum_x \frac{2S_R \cdot S_D + C_1}{S_R^2 + S_D^2 + C_1} \tag{7}$$

where S_R and S_D respectively represent the slope features of the original image and deformed image. SP_1 and SP_2 represent the slope similarity of the reference image and distorted image in the YIQ space and L*a*b* space, respectively. Table 1 lists the values of $SP_i (i = 1,2)$ for different levels of distortion in an input image in both color spaces. As the distortion level increases, the values of SP_i gradually decrease. Therefore, a higher value of SP_i denotes a lower degree of image distortion and better image quality.

3.4 Chroma feature

The color of an image is a global feature that describes the surface properties of objects or regions within the image, playing a crucial role in human visual perception. Therefore, this paper converts RGB images to YIQ and L*a*b* color spaces for feature extraction, which are more perceptually uniform by the human eye.

In the YIQ color space, the image's color information is described through the I and Q channels, and the formula for computing the color similarity between the reference image and the distorted image is as follows:

$$SI = \frac{1}{N}\sum_x \frac{2I_R(x) \cdot I_D(x) + C_1}{I_R^2(x) + I_D^2(x) + C_1} \tag{8}$$

$$SQ = \frac{1}{N}\sum_x \frac{2Q_R(x) \cdot Q_D(x) + C_1}{Q_R^2(x) + Q_D^2(x) + C_1} \tag{9}$$

$$SC_1 = \frac{1}{N}\sum_x \left(\frac{2I_R(x) \cdot I_D(x) + C_1}{I_R^2(x) + I_D^2(x) + C_1} \cdot \frac{2Q_R(x) \cdot Q_D(x) + C_1}{Q_R^2(x) + Q_D^2(x) + C_1} \right) \tag{10}$$

where I_R, I_D and Q_R, Q_D represent the original image's and deformed image's color information in the I and Q channel, respectively. The similarity between the original image and deformed image in the I and Q channels is denoted by SI and SQ, respectively. N denotes the number of pixels. C_1 is a constant, and SC_1 represents the image's color fidelity in the YIQ space.

In the L*a*b* color space, the image's color information is described via the a* and b* channels, and the formula for computing the color similarity between the original image and deformed image is as follows:

$$Sa = \frac{1}{N}\sum_x \frac{2a_R(x) \cdot a_D(x) + C_1}{a_R^2(x) + a(x) + C_1} \tag{11}$$

$$Sb = \frac{1}{N}\sum_x \frac{2b_R(x) \cdot b_D(x) + C_1}{b_R^2(x) + b_D^2(x) + C_1} \tag{12}$$

$$SC_2 = \frac{1}{N}\sum_x \left(\frac{2a_R(x) \cdot a_D(x) + C_1}{a_R^2(x) + a_D^2(x) + C_1} \cdot \frac{2b_R(x) \cdot b_D(x) + C_1}{b_R^2(x) + b_D^2(x) + C_1} \right) \tag{13}$$

where a_R, a_D and b_R, b_D refer to the color information of the original image and deformed image in the a* and b* channel, respectively. The similarity between the original image and deformed image in the a* and b* channels is denoted by Sa and Sb. N denotes the number of pixels. C_1 is a constant, and

SC_2 represents the color fidelity of the image in the $L^*a^*b^*$ color space.

Table 2 lists the values of SC_i calculated for an input image at distortion levels 1 to 5 in two color spaces. As the distortion level increases, the value of SC_i decrease. Therefore, larger values of SC_1 and SC_2 denote smaller color differences between the original image and deformed image, resulting in a lesser degree of color distortion in the image.

3.5 Gradient features

The edges of images contain rich information, making the extraction of edge features crucial. In edge detection, a commonly used method is gradient-based edge detection. Commonly used gradient-based edge detection operators include the Sobel operator, Prewitt operator, and Roberts operator. Compared to the Sobel and Roberts operators, the Prewitt operator is better at suppressing noise, making it less susceptible to interference from noise when detecting edges. It yields relatively good edge localization results. Therefore, in this paper, the Prewitt operator is used to extract the gradient features to represent the edge information of the image. The Fig. 5 represents gradient maps extracted from different color spaces, with some detailed differences annotated.

Firstly, to acquire the image's horizontal and vertical gradients, the Prewitt operator is applied to the image through convolution in both the horizontal and vertical directions, resulting in the horizontal gradient and vertical gradient of the image. The gradient components of the horizontal and vertical directions are as follows:

$$G_h(x) = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} * f(x) \tag{14}$$

$$G_v(x) = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} * f(x) \tag{15}$$

where $f(x)$ represents the test image, the gradient component is denotes in the horizontal direction by $G_h(x)$ and in the

vertical direction by $G_v(x)$. Thus, the formula for computing the image gradient $G(x)$ is as follows:

$$G(x) = \sqrt{G_h^2 + G_v^2} \tag{16}$$

Then, the gradient features of the original image and deformed image are used to calculate the gradient similarity based on the similarity calculation formula, which is expressed as follows:

$$SG_i = \frac{1}{N} \sum_x \frac{2G_R(x) \cdot G_D(x) + C_1}{G_R^2(x) + G_D^2(x) + C_1} \tag{17}$$

where $G_R(x)$ represents the gradient feature of the reference image, $G_D(x)$ represents the gradient feature of the distorted image. C_1 is a constant. The gradient similarity between two images is computed by $SG_i(x)$ $(i = 1,2)$, where the value of i denotes different spaces.

Finally, the gradient similarities computed in the YIQ space and $L^*a^*b^*$ space are denoted as SG_1 and SG_2, respectively. Table 3 lists the values of SG_i calculated in the two spaces for input images with different distortion levels. As the distortion level increases, the values of SG_i
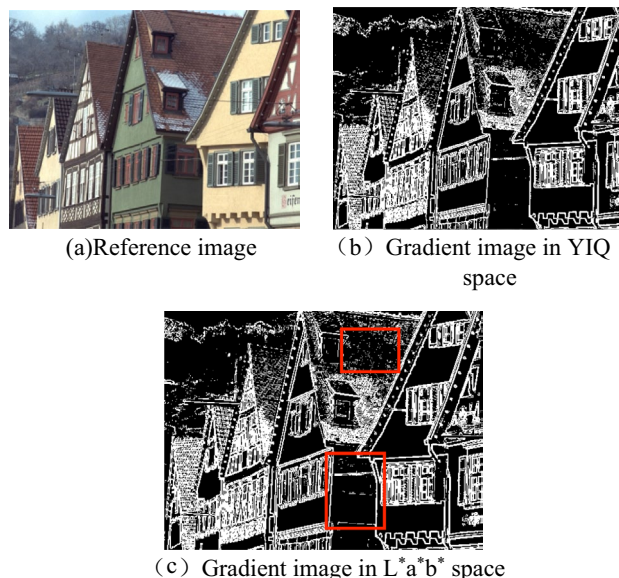

(a)Reference image   （b）Gradient image in YIQ space

（c）Gradient image in $L^*a^*b^*$ space

Fig. 5 Gradient images extracted from different color spaces

Table 2 Comparison of chromatic similarity for different distortion levels

Metric	1	2	3	4	5
SC_1	0.3332	0.3329	0.3327	0.3323	0.3318
SC_2	0.3040	0.2827	0.2783	0.2557	0.2235

Table 3 Comparison of gradient similarity corresponding to different levels of distortion

Metrics	1	2	3	4	5
SG_1	0.9913	0.9818	0.9485	0.9407	0.9405
SG_2	0.9914	0.9838	0.9523	0.9421	0.9416

gradually decrease. Therefore, a larger SG_i value indicates lower image distortion and better image quality.

3.6 Spatial frequency feature

The spatial frequency feature of an image refers to the frequency of changes in pixel grayscale values within the image. In an image, the changes in grayscale values can be seen as a variation in spatial information, and spatial frequency describes the frequency of such changes. The Contrast Sensitivity Function (CSF) measures the sensitivity of the Human Visual System (HVS) to different spatial frequencies. Therefore, the CSF can be simulated by extracting an image's spatial frequency feature. In this paper, the CSF of the HVS is simulated using the Y channel of the image in both the YIQ space and the $L^*a^*b^*$ space. Firstly, the spatial frequency of the image is partitioned into high-frequency, mid-frequency and low-frequency sub-bands. Then, the energy values of these three sub-bands are calculated within 4×4 discrete cosine transform (DCT) blocks of the image.

The formula for calculating the energy in the image's high-frequency region is as follows:

$$E_H = \sum_{(u,v) \in R_H} P(u, v) \tag{18}$$

where R_H represents the high-frequency region, the normalized DCT coefficient of the image at the frequency domain (u,v) is represented by $P(u,v)$, and E_H represents the energy value of the high-frequency region.

Similarly, the formulas for calculating the energy in the mid-frequency and low-frequency regions of the image are as follows:

$$E_M = \sum_{(u,v) \in R_M} P(u, v) \tag{19}$$

$$E_L = \sum_{(u,v) \in R_L} P(u, v) \tag{20}$$

where R_M and R_L respectively represent the mid-frequency and low-frequency regions.

Next, the energy similarity of the three frequencies is calculated based on the similarity formula, which are as follows:

$$SH_i = \frac{1}{N} \sum_{(u,v) \in R_H} \frac{2E_{RH} \cdot E_{DH} + C_2}{E_{RH}^2 + E_{DH}^2 + C_2} \tag{21}$$

$$SM_i = \frac{1}{N} \sum_{(u,v) \in R_M} \frac{2E_{RM} \cdot E_{DM} + C_3}{E_{RM}^2 + E_{DM}^2 + C_3} \tag{22}$$

$$SL_i = \frac{1}{N} \sum_{(u,v) \in R_L} \frac{2E_{RL} \cdot E_{DL} + C_4}{E_{RL}^2 + E_{DL}^2 + C_4} \tag{23}$$

where E_{RH} and E_{DH}, E_{RM} and E_{DM}, E_{RL} and E_{DL} respectively represent the energy with frequencies of high, mid, and low of the reference and distorted images. C_2, C_3 and C_4 are a constant. The frequency similarity of the image in the high, mid, and low domains of the two spaces is represented by the variables SH_i, SM_i and $SL_i(i = 1,2)$.

Finally, the spatial frequency similarities calculated in the YIQ space and $L^*a^*b^*$ space are denoted as SH_1, SM_1, SL_1 and SH_2, SM_2, SL_2 respectively. According to the Fig. 6, (a) represents the original image, while Fig. 6(b) and (c) depict different levels of distortion for the same distortion type (graded from level 1 to 5, with 1 indicating minimal distortion and 5 representing maximum distortion). Figure 6(b) is a level 1 deformed image with SH_1, SM_1 and SL_1 values of 0.0026, 0.0026 and 0.0026 respectively in the YIQ space; SH_2, SM_2 and SL_2 values of 0.0012, 0.0024 and 0.0022 respectively in the $L^*a^*b^*$ space. On the other hand, Fig. 6(c) is a level 5 deformed image with SH_1, SM_1 and SL_1 values of 0.0026, 0.0026 and 0.0024 respectively in the YIQ space; SH_2, SM_2, and SL_2 values of 0.0012, 0.0023 and 0.0019 respectively in the $L^*a^*b^*$ space. Therefore, the values of SH_i, SM_i and SL_i can reflect the degree of image distortion.

(a)Original image   (b）Deformed image(Level 1)
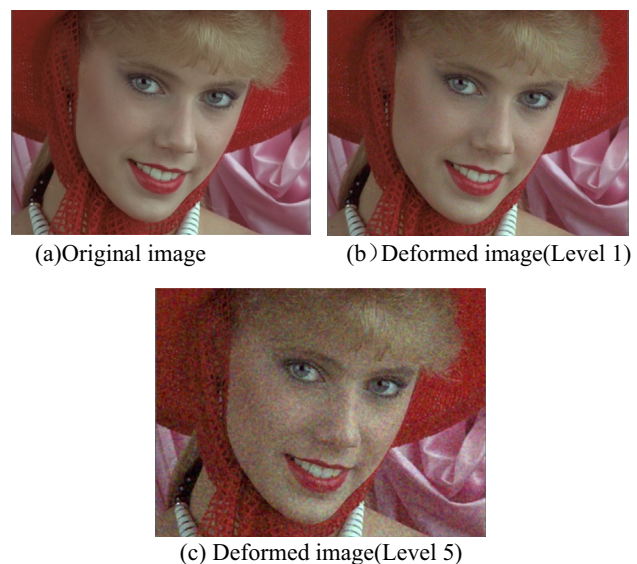
(c) Deformed image(Level 5)

Fig. 6 The original image and deformed images of different levels

4 Model and experimental analysis

4.1 Feature fusion

Through the aforementioned process, we can extract 2 luminance features, 2 slope features, 2 chroma features, 2 gradient features, and 6 spatial frequency features from the image in both spaces, yielding a total of 14 features. Subsequently, these 14 features are fused to build a new feature vector F.

$$F = [SY, SL, SP_1, SP_2, SC_1, SC_2, SG_1, SG_2, SH_1, SM_1, SL_1, SH_2, SM_2, SL_2]$$

4.2 Random forest regression prediction model

The Random Forest (RF) is an ensemble learning method used for both classification and regression tasks. It consists of multiple decision trees, each functioning as a classifier or regressor. In regression tasks, Random Forest predicts results by averaging values. Compared to a single decision tree, Random Forest demonstrates higher accuracy and robustness, making it capable of handle high-dimensional data, large datasets, and complex feature spaces. Therefore, this paper uses the Random Froest regression model to predict image quality. The parameters chosen for the Random Forest model in this paper are 500 decision trees and a minimum leaf node size of 2, i.e., (ntree, mtry) = (500, 2).

Firstly, during the training and testing stages of the model, we need to merge the feature vector F and the subjective scores of corresponding images to generate a new dataset. Subsequently, the dataset is inputted into the model for training model and prediction image quality.

4.3 Feature analysis

During the feature fusion process, each feature has a different impact on the predictive performance of the model. One of the methods to calculate feature importance in Random Forests is by manipulating out-of-bag (OOB) data. The formula to compute feature importance is as follows:

$$I = \sum \frac{(errOOB2 - errOOB1)}{N} \tag{24}$$

where *errOOB1* represents the out-of-bag error calculated for each decision tree, *errOOB2* represents the recalculated out-of-bag error after introducing noise to each feature, and N is the number of decision trees.
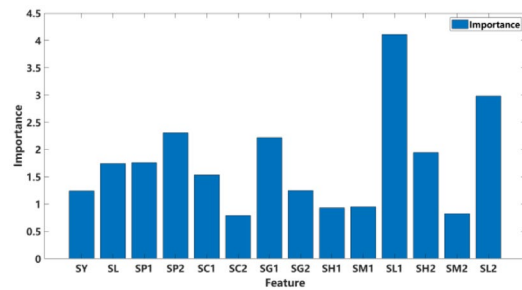
Therefore, to evaluate the contribution of each feature to the model, this paper plots the importance of the 14 extracted features in the four datasets. As shown in Fig. 7, the same feature demonstrates different importance in
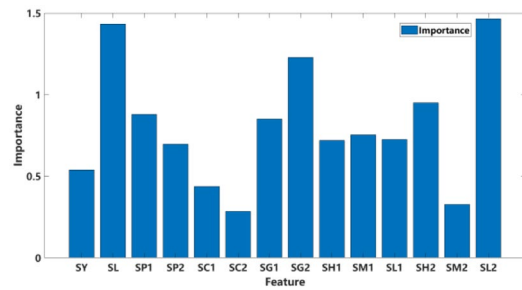
（a）Feature importance on the LIVE dataset
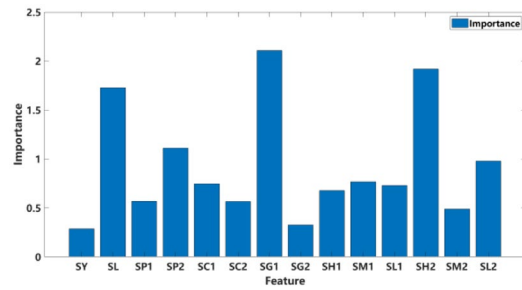
（b）Feature importance on the CSIQ dataset

（c）Feature importance on the TID2008 dataset
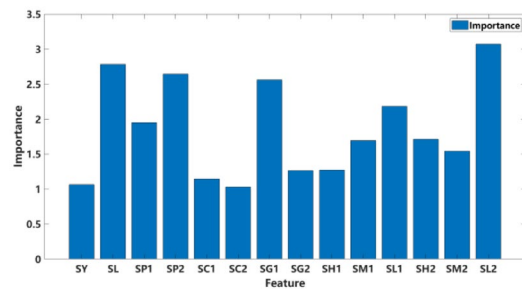
（d）Feature importance on the TID2013 dataset

Fig. 7 Feature importance on different datasets

different datasets, with a higher value indicating greater importance of the feature.

The analysis of the four figures reveals that in the LIVE dataset, SL, SL2, and SG2 emerge as the top three significant features. Similarly, in the CSIQ dataset, SG1, SH2, and SL are denoted as the most important features. In the TID2008 dataset, SL1, SL2, and SP2 hold prominence as the

first three crucial features. Lastly, in the TID2013 dataset, SL2 takes precedence followed by SL and SP2. Hence, it is evident that a newly introduced contrast feature, namely slope feature SP, makes a notable contribution to the accuracy of IQA in this paper.

4.4 Image datasets

We conducted experimental tests on four publicly available image datasets, which are LIVE, CSIQ, TID2008, and TID2013. Each image dataset contains multiple types of distortion and corresponding original images.

LIVE: The LIVE dataset comprises of 779 distorted images, obtained by subjecting the reference images to five distinct computer distortion operations at five different levels. These operations include structural distortions (Gaussian blur), image-related distortions (JPEG compression, JPEG2000 compression, and JPEG2000 fast scale fading distortion), random noise (white noise), along with 29 high-resolution 24-bit/pixel RGB color reference images. Each type of distortion is present at five different levels.

CSIQ: The CSIQ dataset includes 30 pristine images and 866 composite distorted images. It cover six types of distortions: Gaussian blur, additive color Gaussian noise, additive white Gaussian noise, global contrast attenuation, JPEG compression, and JPEG2000 compression. Each distortion type generates 866 degraded versions of the original image across 4 to 5 different levels of degradation.

TID2008: The TID2008 contains 25 reference images and 1700 distorted images. The types of distortion in the data set are: additive Gaussian noise, additive noise with a color component stronger than the illumination component, spatial position correlation noise, mask noise, high frequency noise, pulse noise, quantization noise, Gaussian blur, image noise, JPEG compression, JPEG2000 compression, JPEG transmission errors, JPEG2000 transmission errors, non-eccentric noise, local block distortion of different intensification, intensity mean shift and contrast change. Each type of distortion is present at four different levels.

TID2013: The TID2013 dataset contains 25 reference images and 3000 distorted images. Compared to TID2008, TID2013 introduces seven additional types of distortions: changing color saturation, multiple Gaussian noise, comfort noise, lossy compression, color image quantization, color difference, and sparse sampling. Each type of distortion is present at five different levels.

4.5 Experimental analysis

The proposed method in this paper has been assessed using public datasets: LIVE, CSIQ, TID2008, and TID2013. The evaluation involved conducting 1000 experiments to assess the performance of the method. This assessment is based on

calculating the average values of four evaluation metrics: Root Mean Squared Error (RMSE), Spearman rank-order correlation coefficient (SROCC), Pearson linear correlation coefficient (PLCC), and Kendall rank-order correlation coefficient (KROCC).

Before conducting the experiments, this paper sets the parameters C_1, C_2, C_3 and C_4 in the above formulas as 1, 1.7, 2000 and 0.6, respectively. In calculating the similarity formulas (4), (5), (7) to (13) and (17), C1 is set to a value of 1. In calculating the mid-frequency similarity of images, C3 is set to 2000 because mid-frequency represents details and structures with moderate changes in the image, aiming to better match the human eye's sensitivity to mid-frequency components. On the other hand, through multiple experiments and comparisons with other values settings, the constant values set in the paper have achieved optimal results. This paper adopts five-fold cross-validation to assess the generalization ability of the Random Forest regression model, determine the optimal combination of hyperparameters, and avoid overfitting issues. The image dataset is divided into 80% for training the model and 20% for evaluating image quality prediction. Furthermore, to confirm the benefits of the proposed method, a comparison is made with four metrics of several other method. As shown in Tables 4, 5, 6 and 7, the comparison includes traditional image quality assessment methods such as SPSIM [6], DSSIM [7], MaD-DLS [8], KSCM [10], DISTS [11], MMVD [12], PSIM [13], PSNR [28], VIF [29], SSIM [30], FSIM [31], GSM [32] and VSI [33]. It also includes deep learning-based image quality assessment methods such as LLF-ELM [34], with the optimal results highlighted in bold.

From Table 4 to Table 7, according to the comparison results, the method proposed in this paper outperforms the

Table 4 Comparison of RMSE metrics for different methods

Methods	LIVE	CSIQ	TID2008	TID2013
SPSIM	7.6288	0.0934	0.6046	0.5165
DSSIM	8.1397	0.0829	0.6261	0.5502
MaD-DLS	6.6111	0.0690	–	0.5225
KSCM	8.3316	0.0794	0.6473	0.5859
DISTS	–	–	–	–
MMVD	7.4964	0.0845	**0.4279**	0.5157
PSIM	7.7996	0.0696	0.5628	0.5193
PSNR	13.3597	0.1575	1.0994	0.8832
VIF	7.6137	0.0980	0.7899	0.7880
SSIM	8.9455	0.1334	0.8511	0.7608
FSIM	7.6742	0.1078	0.6527	0.6349
GSM	8.4326	0.1146	0.7235	0.6603
VSI	8.6817	0.0979	0.6466	0.5404
LLF-ELM	–	**0.0659**	0.5159	0.4949
Proposed	**4.2173**	0.0683	0.5268	**0.4443**

Table 5 Comparison of SROCC metrics for different methods

Methods	LIVE	CSIQ	TID2008	TID2013
SPSIM	0.9620	0.9440	0.9104	0.9044
DSSIM	0.9584	0.9516	0.9028	0.8758
MaD-DLS	0.9724	0.9634	–	0.8849
KSCM	0.9585	0.9519	0.8790	0.8603
DISTS	0.954	0.929	–	0.830
MMVD	09557	0.9351	0.9347	0.9106
PSIM	0.9622	0.9621	0.9120	0.8925
PSNR	0.8756	0.8058	0.5531	0.6394
VIF	0.9636	0.9195	0.7491	0.6769
SSIM	0.9479	0.8756	0.7749	0.7417
FSIM	0.9634	0.9240	0.8804	0.8022
GSM	0.9561	0.9108	0.8504	0.7946
VSI	0.9524	0.9423	0.8979	0.8965
LLF-ELM	–	**0.9701**	0.9212	0.9027
Proposed	**0.9769**	0.9695	**0.9355**	**0.9273**

Table 7 Comparison of KROCC metrics for different methods

Methods	LIVE	CSIQ	TID2008	TID2013
SPSIM	0.8271	0.7880	0.7303	0.7251
DSSIM	0.8204	0.8013	0.7205	0.6978
MaD-DLS	–	–	–	–
KSCM	–	–	–	–
DISTS	0.811	0.767	–	0.639
MMVD	0.8501	0.7674	0.7574	0.7426
PSIM	0.7161	0.7395	**0.8273**	**0.8294**
PSNR	0.6865	0.6084	0.4027	0.4696
VIF	0.8282	0.7573	0.5860	0.5147
SSIM	0.7963	0.6907	0.5768	0.5588
FSIM	0.8335	0.7561	0.6945	0.6269
GSM	0.8150	0.7374	0.6569	0.6255
VSI	0.8058	0.7857	0.7123	0.7183
LLF-ELM	–	0.8470	0.7587	0.7321
Proposed	**0.8820**	**0.8509**	0.7691	0.7719

Table 6 Comparison of PLCC metrics for different methods

Methods	LIVE	CSIQ	TID2008	TID2013
SPSIM	0.9559	0.9344	0.8927	0.9091
DSSIM	0.9545	0.9401	0.8844	0.8941
MaD-DLS	0.9703	0.9648	–	0.9068
KSCM	0.9524	0.9531	0.8760	0.8812
DISTS	0.954	0.928	–	0.855
MMVD	0.9567	0.9365	0.9271	0.9171
PSIM	0.9584	0.9642	0.9078	0.9080
PSNR	0.8723	0.8000	0.5734	0.7017
VIF	0.9604	0.9277	0.8084	0.7720
SSIM	0.9449	0.8613	0.7732	0.7859
FSIM	0.9597	0.9118	0.8738	0.8589
GSM	0.9512	0.8964	0.8422	0.8463
VSI	0.9482	0.9279	0.8762	0.9000
LLF-ELM	–	0.9695	0.9224	0.9169
Proposed	**0.9808**	**0.9696**	**0.9279**	**0.9350**

represents the predicted values. The horizontal axis denotes the samples, and the vertical axis denotes the image scores. From the figure, it can be visually observed that the predicted values of each image on the model have a good fit with the true values, as well as the overall prediction error of the model. Smaller errors between the predicted and true values indicate higher accuracy of the proposed method in predicting image quality.

4.6 Ablation experiments

In this paper, five features of the image are extracted in the dual-space, namely: luminance, slope, chrominance, gradient, and spatial frequency features. In order to verify that feature extraction in dual-space for image quality assessment is more accurate than in single space, as well as to measure the effectiveness of the proposed "slope" metric, we conducted two ablation experiments. In the first ablation experiment, we compared the differences in image quality prediction between feature extraction in the single YIQ space, $L^*a^*b^*$ space, and dual-space. The experimental data is written in Tables 8 and 9. In the second ablation experiment, we plotted the Tables 10 and 11 for image quality prediction with and without the "slope" feature. From Tables 8 to Table 11 respectively demonstrate the SROCC and PLCC values between the predicted results and subjective scores for single space, dual-space, and with/ without "slope" feature.

The best results highlighted in bold. It can be observed from Table 8 to Table 11 that the dual-space method and the "slope" feature proposed in this paper both contribute to improving the accuracy of image quality prediction.

previous methods. Compared with the deep learning-based LLF-ELM method, the method proposed in this paper only slightly lower RMSE and SROCC values on the CSIQ dataset, and lower RMSE values on the TID2008 dataset. Apart from this, the overall performance of this paper is superior to the LLF-ELM method. Overall, the proposed method demonstrates higher accuracy in predicting image quality.

Additionally, comparison plots are generated to visually depict the predicted values and true values of the proposed method on the test sets of the LIVE, CSIQ, TID2008 and TID2013 datasets. As shown in the Fig. 8, the test sample sizes for the four test sets are 785, 600, 1360 and 2400, respectively. The red represents the true values, and the blue

（a）Predicted values vs. truth values on the LIVE dataset

（b）Predicted values vs. truth values on the CSIQ dataset

（c）Predicted values vs. truth values on the TID2008 dataset

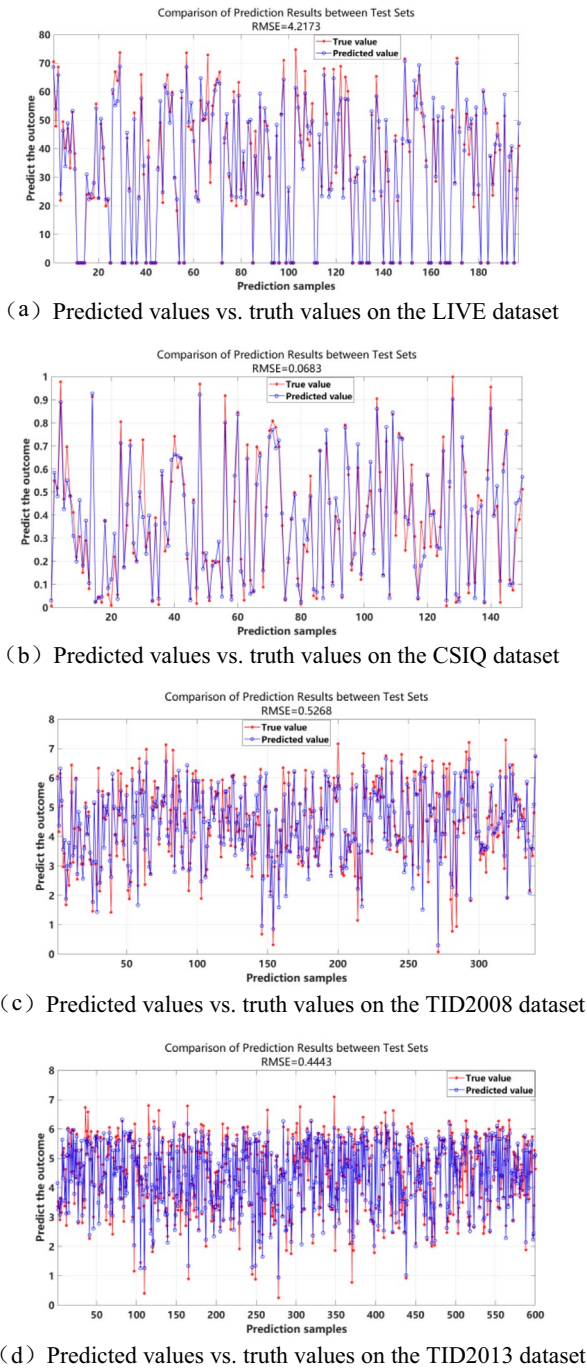（d）Predicted values vs. truth values on the TID2013 dataset

Fig. 8 Comparison of predicted values and truth values on the test sets of the four datasets

Table 8 Comparison of SROCC metrics for different spaces

Spaces	LIVE	CSIQ	TID2008	TID2013
YIQ space	0.9542	0.9503	0.8825	0.8828
L*a*b* space	0.9508	0.9469	0.8756	0.9083
Dual-space (Proposed)	**0.9769**	**0.9695**	**0.9355**	**0.9273**

Table 9 Comparison of PLCC metrics for different spaces

Spaces	LIVE	CSIQ	TID2008	TID2013
YIQ space	0.9659	0.9519	0.8857	0.8971
L*a*b* space	0.9632	0.9544	0.8700	0.9083
Dual-space (Proposed)	**0.9808**	**0.9696**	**0.9279**	**0.9350**

Table 10 Comparison of SROCC metrics with/without "slope" feature

Compare	LIVE	CSIQ	TID2008	TID2013
With "slope"	**0.9769**	**0.9695**	**0.9355**	**0.9273**
Without "slope"	0.9721	0.9524	0.9175	0.9046

Table 11 Comparison of PLCC metrics with/without "slope" feature

Compare	LIVE	CSIQ	TID2008	TID2013
With "slope"	**0.9769**	**0.9695**	**0.9355**	**0.9273**
Without "slope"	0.9670	0.9588	0.9150	0.9130

4.7 Analysis of model performance

In Random Forests, the training data for each decision tree is obtained through bootstrap sampling, which means that some samples may not appear in the training set of a certain tree. For these unsampled samples, their performance can be evaluated by the number of times they do not appear in any tree, which is known as the out-of-bag error. The formula for calculating the out-of-bag error is as follows:

$$OOB = \frac{\sum |T - P|}{N} \tag{25}$$

where T and P represent the true values and predicted values for each sample respectively, and N represents the number of samples.

To measure the model's generalization performance, which enables it to adapt well to new data, we plot the model error curve under different numbers of decision trees. This facilitates a clearer understanding of how the quantity of decision trees affects the outcomes of Random Forests predictions. As shown in the Fig. 9, the error curve experiences a rapid decline within the range of 0–10 decision trees. With the increase in the number of decision trees, the out-of-bag error gradually decreases and eventually stabilizes at around 0.0030. Therefore, the model demonstrates strong generalization performance.
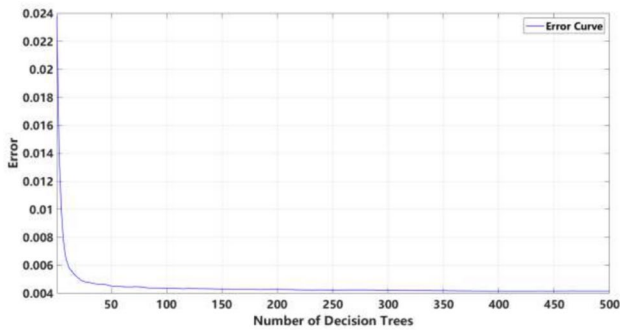
Fig. 9 The impact of the number of decision trees on the model's generalization performance

Table 12 The relationship between sample quantity and prediction performance in the LIVE dataset

Training sample quantity	RMSE	SROCC	PLCC	KROCC
786 (80%)	4.6857	0.9727	0.9795	0.8654
687 (70%)	4.7852	0.9700	0.9785	0.8650
589 (60%)	4.7593	0.9695	0.9760	0.8645
491 (50%)	5.1568	0.9682	0.9753	0.8539
393 (40%)	5.2716	0.9668	0.9743	0.8482
295 (30%)	5.3876	0.9638	0.9721	0.8425
196 (20%)	5.4961	0.9598	0.9711	0.8318

Table 13 The relationship between sample quantity and prediction performance in the CSIQ dataset

Training sample quantity	RMSE	SROCC	PLCC	KROCC
600 (80%)	0.0710	0.9654	0.9696	0.8386
525 (70%)	0.0732	0.9667	0.9688	0.8437
450 (60%)	0.0737	0.9630	0.9688	0.8264
375 (50%)	0.0748	0.9623	0.9601	0.8250
300 (40%)	0.0784	0.9600	0.9589	0.8236
225 (30%)	0.0814	0.9539	0.9541	0.8118
150 (20%)	0.0889	0.9484	0.9459	0.8021

4.8 Analysis of sample quantity

In addition, this paper conducted further experiments by adjusting the number of training and testing samples on four datasets. Specifically, the experiments involved calculating four assessment metrics at each stage as the training samples decreased from 80 to 20% in increments of 10%, while the corresponding testing set increased by 10%. Tables 12, 13, 14, and 15 display the experimental outcomes.

Table 14 The relationship between sample quantity and prediction performance in the TID2008 dataset

Training sample quantity	RMSE	SROCC	PLCC	KROCC
1360 (80%)	0.5079	0.9223	0.9279	0.7691
1190 (70%)	0.5387	0.9203	0.9202	0.7659
1020 (60%)	0.5362	0.9126	0.9138	0.7543
850 (50%)	0.5666	0.9045	0.9103	0.7460
680 (40%)	0.5822	0.8936	0.9020	0.7278
510 (30%)	0.6166	0.8842	0.8918	0.7154
340 (20%)	0.6534	0.8661	0.8739	0.6893

Table 15 The relationship between sample quantity and prediction performance in the TID2013 dataset

Training sample quantity	RMSE	SROCC	PLCC	KROCC
2400 (80%)	0.4443	0.9273	0.9350	0.7719
2100 (70%)	0.4622	0.9201	0.9301	0.7595
1800 (60%)	0.4809	0.9174	0.9231	0.7539
1500 (50%)	0.4823	0.9128	0.9199	0.7434
1200 (40%)	0.4948	0.9074	0.9176	0.7402
900 (30%)	0.5154	0.8999	0.9094	0.7283
600 (20%)	0.5367	0.8958	0.9033	0.7210

From Table 12 to Table 15, it can be visually observed that by reducing the model training sample size in the four datasets, the four evaluation metrics of the method RMSE, SROCC, PLCC and KROCC have all decreased. The experimental results indicate that as the quantity of training samples increases, the accuracy of the proposed image quality prediction method in this paper also gradually improves.

5 Conclusion and outlook

5.1 Conclusion

This paper introduces a dual-space multi-feature fusion based method to FR-IQA. The method concurrently extracts luminance, slope, chroma, gradient and spatial frequency features from both spaces. These features are then integrated into a feature vector, which is combined with average subjective scores to generate a dataset. Finally, the Random Forest model is used to predict regression using the dataset.

The innovations of this method are show:

1. Not limited to extracting features in a single space, it can effectively avoid the problem of insufficient feature extraction caused by spatial limitations. By extracting

features from multiple spaces, it is possible to obtain a richer set of feature information, which better describes the content and structure of the image;

2. This paper introduces a method that applies the slope feature of remote sensing images to extract the slope feature of non-terrain images. In remote sensing images, slope is an important terrain feature that can reflect the undulations and changes in the terrain. However, non-terrain images often lack terrain undulations, and extracting the image's slope, which reflects the texture information.

This paper is the first to propose extracting image features in dual-space and introduces a new metric called "slope". Experimental results on four datasets demonstrate that the proposed method exhibits good predictive performance and generalization ability compared to mainstream methods.

5.2 Outlook

The proposed is an FR-IQA method, which typically requires the availability of the original image as a reference. However, in numerous instances, obtaining the original image is difficult, which limits the applicability of FR-IQA methods. In subsequent research endeavors, it would be beneficial to explore the integration of dual-space methods into NR-IQA. Additionally, proposing better feature pooling strategies can also contribute to improving the accuracy of model predictions.

Despite our method has achieved promising performance and improved the accuracy of IQA, there is still room for overall improvement. Limitations of our work: 1. Image feature extraction is a crucial process in IQA. In this paper, we extract universal features of images in two spaces for image quality prediction. Therefore, we will conduct in-depth research on image feature extraction to extract more useful features. 2. In terms of feature fusion strategy, we only fuse features into a feature vector based on empirical methods. We are aware that the strategy of feature fusion can affect prediction results to a certain extent. Therefore, in future work, we will design a more efficient feature fusion strategy to enhance prediction accuracy. 3. This paper uses the Random Forest model for regression prediction of image quality scores. With the continuous development of deep learning, we will conduct research in this area and utilize deep learning methods for more accurate predictions.

5.3 Future work

Unlike traditional two-dimensional images, light field images contain information about the direction and intensity of light propagation in space, typically involving more information dimensions. Light field images possess unique characteristics, such as additional angular and depth information, which may require higher quality standards and more specialized evaluation metrics. When assessing the quality of light field images, their uniqueness may necessitate the use of more targeted evaluation metrics and methods to evaluate their quality, and these metrics may differ from those used for traditional images.

In terms of light field image super-resolution, the development of quality assessment methods for light field images is relatively mature at present. For instance, a new evaluation metric designed in [22], as well as a tensor-based method for light field image quality assessment proposed in [35], have advanced the field to a certain extent and made significant contributions. Although light field images differ from traditional two-dimensional images, they are essentially based on a color space. Therefore, it may be worthwhile to explore the application of the dual-space method mentioned in this paper to this field in our future work.

Author contributions X.W wrote the main manuscript text, Z.S provide data support

Data availability All data generated or analyses during this study are included in this published paper. The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Declarations

Conflicts of interests There is no conflict in this paper.

Ethical approval The main content of this paper is the study of IQA methods, and there is very close relevance between the topic of the revised manuscript and the scope of the Computer Graphics Society's Journals.

References

1. Yang, F., Yang, C., An, P., et al.: 360 video quality assessment based on saliency-guided viewport extraction. Multimedia Syst. **30**, 89 (2024). https://doi.org/10.1007/s00530-024-01285-0
2. No-Reference Screen Content Image Quality Assessment With Unsupervised Domain Adaptation: B. Chen, H. Li, H. Fan and S. Wang. IEEE Trans. Image Process. **30**, 5463–5476 (2021). https://doi.org/10.1109/TIP.2021.3084750
3. S. Aghapour Maleki and H. Ghassemian.: Spatial Quality Assessment of Pansharpened Images Based on Gray Level Co-Occurrence Matrix. 2022 International Conference on Machine Vision and Image Processing (MVIP), Ahvaz, Iran, Islamic Republic of, 1–6 (2022). https://doi.org/10.1109/MVIP53647.2022.9738763
4. Kim, D., Lee, S., Kim, C.: Contextual Information Based Quality Assessment for Contrast-Changed Images. IEEE Signal Process. Lett. **26**(1), 109–113 (2019). https://doi.org/10.1109/LSP.2018.2882916

5. Seo, S., Ki, S., Kim, M.: A Novel Just-Noticeable-Difference-Based Saliency-Channel Attention Residual Network for Full-Reference Image Quality Predictions. IEEE Trans. Circuits Syst. Video Technol. **31**(7), 2602–2616 (2021). https://doi.org/10.1109/TCSVT.2020.3030895

6. Sun, W., Liao, Q., Xue, J.-H., Zhou, F.: SPSIM: A Superpixel-Based Similarity Index for Full-Reference Image Quality Assessment. IEEE Trans. Image Process. **27**(9), 4232–4244 (2018). https://doi.org/10.1109/TIP.2018.2837341

7. Yang, J., Zhang, W., Li, X., Zhou, T., Ou, B.: Full Reference Image Quality Assessment by Considering Intra-Block Structure and Inter-Block Texture. IEEE Access **8**, 179702–179715 (2020). https://doi.org/10.1109/ACCESS.2020.3028282

8. Sim, K., Yang, J., Lu, W., Gao, X.: MaD-DLS: Mean and Deviation of Deep and Local Similarity for Image Quality Assessment. IEEE Trans. Multimedia **23**, 4037–4048 (2021). https://doi.org/10.1109/TMM.2020.3037482

9. Kim, W., Nguyen, A.-D., Lee, S., Bovik, A.C.: Dynamic Receptive Field Generation for Full-Reference Image Quality Assessment. IEEE Trans. Image Process. **29**, 4219–4231 (2020). https://doi.org/10.1109/TIP.2020.2968283

10. Zhou, Z., Li, J., Quan, Y., Xu, R.: Image Quality Assessment Using Kernel Sparse Coding. IEEE Trans. Multimedia **23**, 1592–1604 (2021). https://doi.org/10.1109/TMM.2020.3001472

11. Ding, K., Ma, K., Wang, S., Simoncelli, E.P.: Image Quality Assessment: Unifying Structure and Texture Similarity. IEEE Trans. Pattern Anal. Mach. Intell. **44**(5), 2567–2581 (2022). https://doi.org/10.1109/TPAMI.2020.3045810

12. Dong, W., Bie, H., Lu, L., Li, Y.: Image Quality Assessment by Considering Multiscale and Multidirectional Visibility Differences in Shearlet Domain. IEEE Access **7**, 78715–78728 (2019). https://doi.org/10.1109/ACCESS.2019.2922011

13. Gu, K., Li, L., Lu, H., Min, X., Lin, W.: A Fast Reliable Image Quality Predictor by Fusing Micro- and Macro-Structures. IEEE Trans. Industr. Electron. **64**(5), 3903–3912 (2017). https://doi.org/10.1109/TIE.2017.2652339

14. Yu, M., Tang, Z., Zhang, X., Zhong, B., et al.: Perceptual Hashing With Complementary Color Wavelet Transform and Compressed Sensing for Reduced-Reference Image Quality Assessment. IEEE Trans. Circuits Syst. Video Technol. **32**(11), 7559–7574 (2022). https://doi.org/10.1109/TCSVT.2022.3190273

15. Zhu, W., et al.: Multi-Channel Decomposition in Tandem With Free-Energy Principle for Reduced-Reference Image Quality Assessment. IEEE Trans. Multimedia **21**(9), 2334–2346 (2019). https://doi.org/10.1109/TMM.2019.2902484

16. Shen, L., Zhao, B., Pan, Z., Peng, B., Kwong, S., Lei, J.: Channel Recombination and Projection Network for Blind Image Quality Measurement. IEEE Trans. Instrum. Meas. **71**, 1–12 (2022). https://doi.org/10.1109/TIM.2022.3190029

17. Jin, C., Zhao, X., Xiong, Q., et al.: Blind Image Quality Assessment for Multiple Distortion Image. Circuits Syst Signal Process **41**, 5807–5826 (2022). https://doi.org/10.1007/s00034-022-02055-x

18. Yang, X., Li, F., Liu, H.: TTL-IQA: Transitive Transfer Learning Based No-Reference Image Quality Assessment. IEEE Trans. Multimedia **23**, 4326–4340 (2021). https://doi.org/10.1109/TMM.2020.3040529

19. Liu, L., Wang, T., Huang, H.: Pre-Attention and Spatial Dependency Driven No-Reference Image Quality Assessment. IEEE Trans. Multimedia **21**(9), 2305–2318 (2019). https://doi.org/10.1109/TMM.2019.2900941

20. Wen, Y., Chen, J., Sheng, B., et al.: Structure-Aware Motion Deblurring Using Multi-Adversarial Optimized CycleGAN. IEEE Trans. Image Process. **30**, 6142–6155 (2021). https://doi.org/10.1109/TIP.2021.3092814

21. Yue, G., Wu, H., Jiang, Q., et al.: Perceptual Quality Assessment of Retouched Face Images. IEEE Trans. Multimedia (2023). https://doi.org/10.1109/TMM.2023.3338412

22. Cui, Y., Jiang, G., Yu, M.Y., et al.: Stitched Wide Field of View Light Field Image Quality Assessment: Benchmark Database and Objective Metric. IEEE Trans. Multimedia **26**, 5092–5107 (2024). https://doi.org/10.1109/TMM.2023.3330096

23. Fang, Z., Cui, Y., Yu, M., et al.: Blind 3D-Synthesized Image Quality Measurement by Analysis of Local and Global Statistical Properties. IEEE Trans. Instrum. Meas. **72**, 1–15 (2023). https://doi.org/10.1109/TIM.2023.3306527

24. Baig, M.A., Moinuddin, A.A., Khan, E.: A simple spatial domain method for quality evaluation of blurred images. Multimedia Syst. **30**, 28 (2024). https://doi.org/10.1007/s00530-023-01223-6

25. Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu and G. Zhai.: A Full-Reference Quality Assessment Metric for Fine-Grained Compressed Images. 2021 International Conference on Visual Communications and Image Processing (VCIP), Munich, Germany. 1–4 (2021). https://doi.org/10.1109/VCIP53242.2021.9675389

26. Liu, Y., Li, X.: No-Reference Quality Assessment for Contrast-Distorted Image. IEEE Access **8**, 84105–84115 (2020). https://doi.org/10.1109/ACCESS.2020.2991842

27. Lin, X., Sun, S., Huang, W., et al.: EAPT: Efficient Attention Pyramid Transformer for Image Processing. IEEE Trans. Multimedia **25**, 50–61 (2023). https://doi.org/10.1109/TMM.2021.3120873

28. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. IEEE Trans. Image Process. **15**(11), 3440–3451 (2006). https://doi.org/10.1109/TIP.2006.881959

29. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. **15**(2), 430–444 (2006). https://doi.org/10.1109/TIP.2005.859378

30. Zhou, W., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861

31. Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: A Feature Similarity Index for Image Quality Assessment. IEEE Trans. Image Process. **20**(8), 2378–2386 (2011). https://doi.org/10.1109/TIP.2011.2109730

32. Liu, W.L., Narwaria, M.: Image Quality Assessment Based on Gradient Similarity. IEEE Trans. Image Process. **21**(4), 1500–2151 (2012). https://doi.org/10.1109/TIP.2011.2175935

33. Zhang, L., Shen, Y.: VSI A Visual Saliency-Induced Index for Perceptual Image Quality Assessment. IEEE Trans. Image Process. **23**(10), 4270–4281 (2014). https://doi.org/10.1109/TIP.2014.2346028

34. Ma, Y., Hao, F., Liu, G., et al.: Full-Reference Color Image Quality Assessment Method Based on Extreme Learning Machine for Low-Level Features. Journal of Nanjing Normal University (Natural Science Edition) **45**(4), 91–101 (2022). https://doi.org/10.3969/j.issn.1001-4616.2022.04.013

35. Zhou, W., Shi, L., Chen, Z., et al.: Tensor Oriented No-Reference Light Field Image Quality Assessment. IEEE Trans. Image Process. **29**, 4070–4084 (2020). https://doi.org/10.1109/TIP.2020.2969777