

Perceptual Quality Assessment of Cartoon Images

Hangwei Chen , Xiongli Chai , Feng Shao , Member, IEEE, Xuejin Wang, Qiuping Jiang , Member, IEEE, Xiangchao Meng , Member, IEEE, and Yo-Sung Ho , Fellow, IEEE

Abstract—In the animation industry, automatically predicting the quality of cartoon images based on the inputs of general distortions and color change is an urgent task, while the existing no-reference (NR) methods usually measure the perceptual quality of the natural images. In this paper, based on the observation that structure and color are the main factors affecting cartoon images quality, we proposed a new NR quality prediction metric for cartoon images, which fully takes gradient and color information into account. The experimental results on our newly constructed NBU-CIQAD dataset with color change and other existing cartoon image dataset demonstrate that the proposed method significantly outperforms existing no-references methods for the task of cartoon image quality assessment. The database and code will be released at <https://github.com/1010075746/NBU-CIQAD>.

Index Terms—Cartoon images, color change, no-reference image quality assessment, structural measure, color measure.

I. INTRODUCTION

CARTOON, as a unique art form, plays a vital role in many fields, such as film, entertainment and education, etc. Fig. 1 shows an example of a real-world application, in which viewers can use video website or Photoshop to adjust the color of the scene. Therefore, it is desirable to study a quality evaluation metric for cartoon images with color change, which can make contributions to providing necessary guidance in devising, monitoring and optimizing the animation industry chain. Image quality assessment (IQA) is a basic solution to assess the perceptual quality of an image close to human vision. The cartoon images may encounter general distortions as in natural

Manuscript received 24 May 2021; revised 7 September 2021 and 12 October 2021; accepted 18 October 2021. Date of publication 26 October 2021; date of current version 13 January 2023. This work was supported in part by the Natural Science Foundation of China under Grants 62071261, 61901236, 41801252, and 61622109, and in part by the Natural Science Foundation of China under Grant R18F010008. It was also sponsored by K. C. Wong Magna Fund in Ningbo University. The Associate Editor coordinating the review of this manuscript and approving it for publication was Dr. Mai Xu. (*Hangwei Chen and Xiongli Chai contributed equally to this work.*) (*Corresponding author: Feng Shao.*)

Hangwei Chen, Xiongli Chai, Feng Shao, Qiuping Jiang, and Xiangchao Meng are with the Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: 1010075746@qq.com; 747866472@qq.com; shaofeng@nbu.edu.cn; jiangqiuping@nbu.edu.cn; mengxiangchao@nbu.edu.cn).

Xuejin Wang is with the School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou 350118, China (e-mail: 1020468620@qq.com).

Yo-Sung Ho is with the School of Information and Communications, Gwangju Institute of Science and Technology (GIST), Gwangju 500-712, Korea (e-mail: hoyo@gist.ac.kr).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TMM.2021.3121875>.

Digital Object Identifier 10.1109/TMM.2021.3121875



Fig. 1. Example of a real-world application.

images, e.g., blur, noise, compression artifacts, and other specific cartoon-related distortions caused by different production techniques, e.g., celluloid cartoons, computer animation [1]. Although with escalating research interesting in IQA, the research on cartoon image quality assessment (CIQA) is still lacking.

Over the past decade, IQA is undergoing an increasing popularity in the field of image processing, which can not only measure the accuracy of image information, but also optimize the processes of image fusion [2], [3], enhancement [4], or denoising [5]. Depending on the amount of information used in the task, objective IQA measures can be divided into three categories: full-reference (FR), reduced-reference (RR), and no-reference (NR). Many well-known FR-IQA methods, e.g., Structural Similarity (SSIM) [6], Visual Information Fidelity (VIF) [7], Edge Similarity (EISM) [8], have been proposed. In RR-IQA methods, the quality of the distorted image is estimated based on partial information of the reference images [9], [10]. Particularly, considering the unavailability of reference images in many cases, it is more meaningful to design NR-IQA methods in practice. More comprehensive introduction can be found in [11].

According to the prior knowledge of distortion type, the existing studies on NR-IQA can be broadly classified into two categories: distortion-specific and general-purpose metrics. The distortion-specific quality assessment metrics predict the visual quality of images degraded by specific distortions (e.g., blurriness, blockiness, contrast, rendering or tone-mapping artifacts) [12]–[20]. As a new type of distortion caused by low-quality imaging, IQA for contrast distortion has attracted broad attentions. Jiang *et al.* [16] proposed a no-reference image contrast evaluator by generating bidirectional pseudo references. Gu *et al.* [19] used the concept of information maximization to design a method (NIQMC) for contrast-distorted images from both local and global perspectives. In [20], features from five perceptual factors were adopted to design a no-reference method (BIQME) for contrast-adjusted images. To cope with the CIQA task, color change caused by contrast distortion

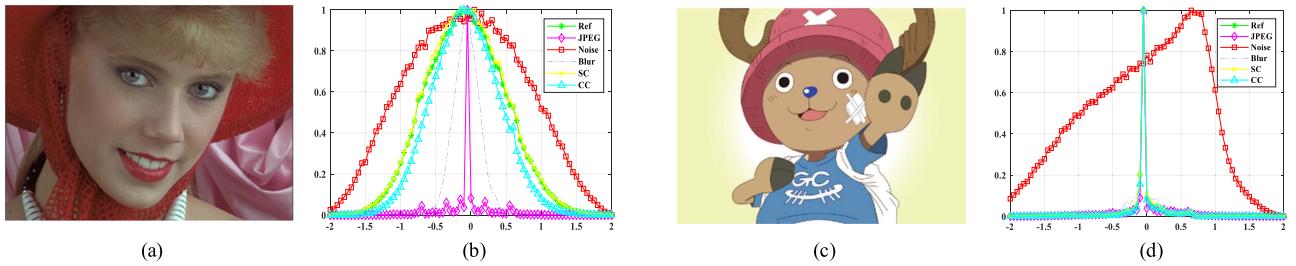


Fig. 2. Distribution of MSCN coefficients for natural and cartoon images. (a) and (c) are reference natural and cartoon images, respectively. (b) Distribution of MSCN coefficients of natural images. (d) Distribution of MSCN coefficients of cartoon images.

should be considered even in the general-purpose measures. The general-purpose NR-IQA does not necessarily embody specific distortion type, which can fall into two categories: natural scene statistics (NSS)-based metrics and learning-based metrics. The NSS-based metrics often transform a raw image into a compact representation, so that statistical regularities can be revealed using common probability density functions, e.g., generalized Gaussian distribution in Blind Image Spatial Quality Evaluator (BRISQUE) [5], Blind Image Quality Indices (BIQI) [20], NR Free Energy-Based Robust Metric (NFERM) [23], or Distortion Identification-based Image Verity and Integrity Evaluation (D-IVINE) [26]. The learning-based approaches aim to train shallow networks (e.g., Support Vector Machine, and Random Forest) or deep networks (e.g., deep neural network, convolutional neural network, and residual neural network) to establish the mapping relationship between image features and human's visual perception. Zhang *et al.* [27] propose a Unified No-reference Image Quality and Uncertainty Evaluator (UNIQUE) based on deep neural network to confront the cross-distortion-scenario challenge. Xu *et al.* [28] propose a viewport-based convolutional neural network (V-CNN) approach for visual quality assessment (VQA) on 360° video. Zhu *et al.* [29] presents a no-reference IQA metric based on deep meta-learning. However, traditional NR-IQA approaches are usually designed for natural images.

Recently, there has been increasing interest in assessing the visual quality of other types of image/video [30]–[39], such as graphic and screen content. Min *et al.* [30] proposed a unified content-type adaptive (UCA) blind IQA measure that is applicable across content types, including compressed natural, graphic and SCIs. Gu *et al.* [31] developed a NR-IQA model for screen content images (SCIs) via the analysis of structural variation (SVQ). Li *et al.* [32] firstly established a brand new Compressed Screen Content Video Quality Database (CSCVQ), and developed a MutiScale Relative Standard Deviation Similarity (MS-RSDS) model for screen content video quality evaluation. Li *et al.* [33] proposed a NR-VQA model via using a multi-scale approach. Fang *et al.* [34] proposed a NR-IQA method by incorporating statistical luminance and texture features (NRLT) for SCIs. Shao *et al.* [35] proposed a blind quality predictor for SCI (BLIQUP-SCI), which conducts local sparse representation for the textual and pictorial regions, and global sparse representation for the global SCIs. Cartoon image can be regarded as a special type of screen content image, but lack of natural scene and textural areas. Chen *et al.* [1] firstly developed a special NR quality estimation method (CBIQA) for cartoon images by

building sharpness and local statistic prior models in edge and non-edge areas, respectively. However, the feature components designed for natural scene and textural areas in the above discussed SCIs methods may be redundant for cartoon images.

The cartoon and natural images are different in appearance and the characteristics of structure and color. For example, the cartoon images often consist of clear edges, smooth color shading and relatively simple textures, while the natural images contain more complex textures. To demonstrate the statistic difference of the natural and cartoon images, we demonstrate the distributions of Subtracted Contrast Normalized (MSCN) coefficients [5] for natural and cartoon images in Fig. 2. It is clearly shown that the coefficients of natural images follow a Gaussian distribution, while cartoon images do not obey natural scene statistics (NSS). Since the cartoon images are different from natural images in structure and color distributions, the traditional NR-IQA methods may be not effective for cartoon images. Fig. 3 presents the scores predicted by different NR-IQA metrics on two sets of high-quality and low-quality cartoon images. The low-quality image in (b) obviously has multiple blocking and noise distortions and in (d) apparently loses edge and texture details caused by the contrast and brightness distortion. However, almost all NR-IQA methods generate high scores for the low-quality cartoon images, which is obviously inconsistent with human's visual perception. From the observations, the traditional NSS models appear to be inadequate for predicting the quality of cartoon images, which indicates that simple parametric models (e.g., generalized Gaussian distribution and asymmetric generalized Gaussian distribution) cannot well fit the distribution of cartoon images. Hence, it is desired to apply non-parametric statistical characteristics (e.g., intensity and distribution) in the form of the histogram to extract quality-aware features without making any assumption about the feature distribution of cartoon images.

To our best knowledge, there is only one publicly available cartoon images database (HFUT-CID) [1], which degrades images with different degradation processes to obtain corresponding low-quality images, including JPEG compression, Gaussian noise, salt and pepper noise, Gaussian blur and AVS2 compression. Our work also considers the similar CIQA issue from the subjective and objective perspective, but aims to solve the following challenges that have not well addressed in these approaches: 1) The HFUT-CID database mainly focuses on the effect of image production, storage, and compression on different stages, but the color-related distortions have been relatively

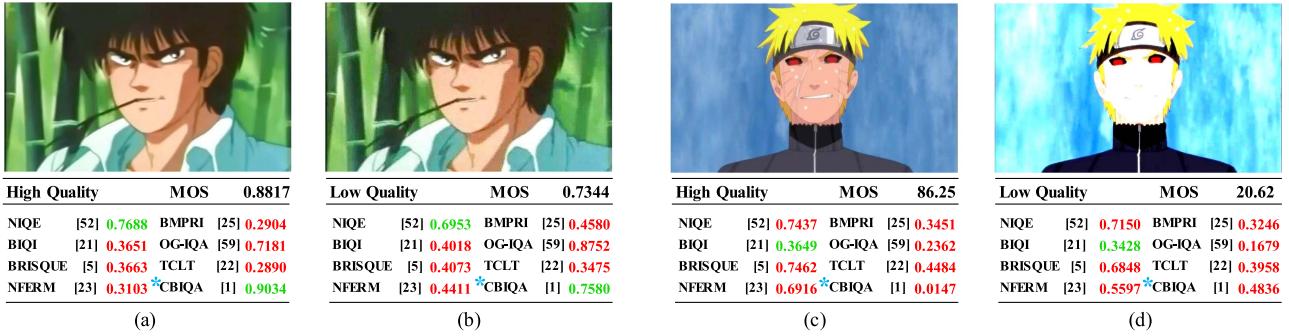


Fig. 3. Traditional NR-IQA methods are ineffective for cartoon images with different types of distortion. Inconsistent scores are marked in red. (a) and (b) are distortions of blur, compression and noise, which are selected from the HFUT-CID. (c) and (d) are distortion of contrast, saturation and brightness, which are selected from the NBU-CIQAD. '*' indicating a lower score for a worse image quality.

ignored in the database, while the quantization of color components in compression, color printing, and color setting of the screen may cause the color change of cartoon images; 2) The CBIQA relies on the assumption that the statistical histograms of high-quality cartoon images obey certain distributions, but the assumption is not always satisfied for low-quality images. In addition, the color information as the basic element in the cartoon image is ignored.

Considering the above challenges, it is necessary to construct a new cartoon image dataset with color change and further investigate the issue of color change for CIQA. Since our database is to provide a benchmark in investigating perceptual quality evaluation for cartoon images with color change, we construct an objective model using statistical features in gradient domain and color features in HSV color space without any prior information. The major contributions of our work are summarized as follows:

1. To carry out in-depth study on perceptual quality assessment of cartoon images from both subjective and objective aspects, we build a new cartoon images database named NBU-CIQAD, which consists of 1800 singly-changed cartoon images (SCCIs) corrupted by one distortion and 800 multiply-changed cartoon images (MCCIs) successively corrupted by two or three distortions. To our knowledge, it is the first large-scale cartoon image database designed for color change with human opinion scores.
2. Based on the observation that structure details and color saturation are the main factors affecting cartoon images quality, we proposed a NR metric, which fully takes gradient and color information into account. In particular, for the purpose of reflecting structure variations, texture and edge information are extracted in gradient domain to give a reasonable estimation. To capture the color variations, we propose to extract moment and entropy features in HSV color space. Extensive experiments on two cartoon databases demonstrate that the proposed method significantly outperforms the existing NR IQAs for the task of CIQA.

The rest of this paper is organized as follows. In section II, we introduce the new NBU-CIQAD database. Section III introduces the proposed method in detail. The experimental results

are shown and discussed in Section IV, and Section V concludes the paper.

II. NBU-CIQAD

To investigate the quality evaluation of cartoon images, we construct a new cartoon image quality assessment database (NBU-CIQAD) with different degrees of brightness, saturation and contrast changes. The database includes 100 reference images, 1800 SCCIs degraded only by one distortion on six degradation levels, and 800 MCCIs degraded by two or three types of distortion on four degradation levels. Subjective evaluation for these MCCIs and SCCIs is conducted to obtain the human option scores. The database will be introduced in details as follows.

A. Construction of the NBU-CIQAD

1. *Source Contents:* In the database, 100 typical cartoon images containing different characters and sceneries are collected from online animation website in different years. In particular, all 100 cartoon images are used as the sources for SCCIs, and the partial 20 cartoon images are selected as the sources for MCCIs. Examples of the selected source images in the database are given in Fig. 4 with the scenes of character shown in the first row and scenery in the second row.
2. *Distortion Stimuli:* Different with the existing image databases degraded by Gaussian blurring, JPEG compression, White Noise or other distortions [40], [41], brightness change (BC), saturation change (SC) and contrast change (CC) are the distortion stimuli in the database. The degradation of distortion stimuli is varied by controlling the parameters within the pre-defined ranges, as reported in Table I. The SCCIs have six degradation levels, and the MCCIs have four degradation levels. Specifically, we can generate 1800 SCCIs from 100 reference images corrupted by independent BC, SC or CC on six degradation levels. In addition, by corrupting with two or three distortions in order on four degradation levels, e.g., SC&CC, CC&BC, SC&BC, and SC&CC&BC, we can obtain eight combinations for two distortions and sixteen combinations for three distortions, respectively. As a result, we have total 40



Fig. 4. Examples of the selected source images in the database.



Fig. 5. Examples of changed images.

TABLE I
CONTROL PARAMETERS IN THE DATABASE

Change	SCCIs		MCCIs	
	Range	Level	Range	Level
Saturation	$\pm 60, \pm 40, \pm 20$	6	$\pm 50, \pm 30$	4
Contrast	$\pm 60, \pm 40, \pm 20$	6	$\pm 50, \pm 30$	4
Brightness	$\pm 60, \pm 40, \pm 20$	6	$\pm 50, \pm 30$	4

combinations for all degradation levels and generate 800 MCCIs from 20 reference images. Overall, the database includes 1800 SCCIs and 800 MCCIs. In the actual implementation, refer to [4], we add a shifting value to change the contrast and brightness of the image. In addition, we adopt linear and nonlinear adjustments to expand the range of saturation. The detailing operations are illustrated in Table II. Examples of the adjusted cartoon images are shown in Fig. 5.

B. Subjective Testing Methodology

The subjective tests were conducted in a laboratory designed for subjective quality evaluation. All images are displayed on a 23-inch true color (32bits) LCD monitor with a resolution of 1920×1080 . The viewing conditions meet the recommendation of ITU-R BT.500-13 [42]. Twenty-five graduate students (twelve females and thirteen males) are participated in the subjective evaluation. In the subjective test, all testing images are randomly presented, and the participants are asked to rate the quality of cartoon images on a five-level scale: Excellent, Good, Fair, Poor,

TABLE II
SPECIFIC FORMULA DESCRIPTION

BC	Enhancement	Using the components V of HSV color space $I_{out} = I_{in}(1+i)$
	Diminishment	Using the components V of HSV color space $I_{out} = I_{in}(1-i)$
CC	Enhancement	$I_{out} = 127 + (I_{in} - 127) \cdot (1+i)$
	Diminishment	$I_{out} = 127 + (I_{in} - 127) \cdot (1-i)$
SC	Enhancement	$\alpha = \begin{cases} \frac{1-s}{s} & \text{if } i+s > 1 \\ \frac{i}{1-i} & \text{if } i+s < 1 \end{cases}$ $I_{out} = I_{in} + (I_{in} - 255 \cdot B) \cdot \alpha$
	Diminishment	$\alpha = i$ $I_{out} = B \cdot I_{in} + (I_{in} - 255 \cdot B) \cdot (1+\alpha)$

* I_{out} denotes the output image, I_{in} denotes the input image. i denotes the degradation level, normalized to [01]. B denotes the brightness of image, s denotes the saturation of image and α denotes adjustment coefficient.

Bad, which meet the ITU-R absolute category rating (ACR). The rating criteria focus on details, edges, and color in subjective test.

After obtaining all raw subjective scores from 25 participants and removing 4 outlier subjects (within 95% confidence interval), the normalized z-score is calculated as:

$$z_{i,j} = \frac{s_{i,j} - \mu_i}{\sigma_i} \quad (1)$$

where $s_{i,j}$ denotes the score assigned by the subject i to the image j , and μ_i and σ_i denotes the mean and standard deviation

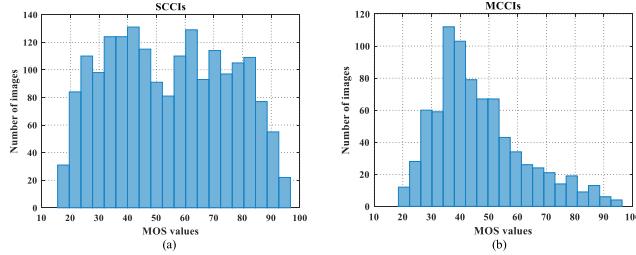


Fig. 6. Histogram of MOS in SCCIs and MCCIs sessions. (a) SCCIs, and (b) MCCIs.

calculated from all scores assigned by subject i . Then, the normalized z-score is mapped to the range [0100] to obtain a mean opinion scores (MOS) [43]:

$$z'_{i,j} = \frac{100(z_{i,j} + 3)}{6} \quad (2)$$

$$\text{MOS}_j = \frac{1}{U} \sum_i z'_{i,j} \quad (3)$$

where U denotes the number of subjects after subject rejection. Distributions of MOS scores for the two individual SCCIs and MCCIs sessions are exhibited in Fig. 6.

C. Analyses of Subjective Scores

It is expected that the measured subjective rating scales of the SCCIs and MCCIs in the database should exhibit good separation of perceptual quality. As exhibited in Fig. 6, the MOS distributions of the two sub-datasets span over a wide range of visual quality (from the low quality to high quality), which indicate that the database exhibits good separation of perceptual quality. Compared with the MOS distribution of the SCCIs, the distribution of MCCIs mainly concentrates upon the low-quality level, in which the discrimination of the MCCIs is comparatively poor than the SCCIs with a wide quality range. To distinguish the individual influences of saturation, contrast and brightness, we show the MOS values of all SCCI against the saturation level, contrast level or brightness level respectively in Fig. 7, in which piecewise linear interpolation is used to enhance the visibility of the trends [44]. It can be seen that the MOS value declines with the increasing degree of adjustment, but the fact is not always true because a small increase in saturation may have positive influence on quality (increasing MOS). In addition, to investigate the subjective rating of each individual subject, we calculate the Spearman Rank-order Correlation Coefficient (SRCC) indicator based on the MOS values of each subject. Fig. 8 gives the SRCC performance of individual subjects for SCCIs and MCCIs sessions. As observed, each individual subject can achieve comparatively consistent rating for different image content.

To quantitatively investigate the effects of different types and degrees of distortion on the perceptual quality, we show the distributions of mean and standard deviation of MOS values for different sub-sections classified by the introduced distortion types in Fig. 9(distinguished by different color bars in the histogram). From the figure, we have the following observations: 1) Although the subjects are in agreement on the quality of the

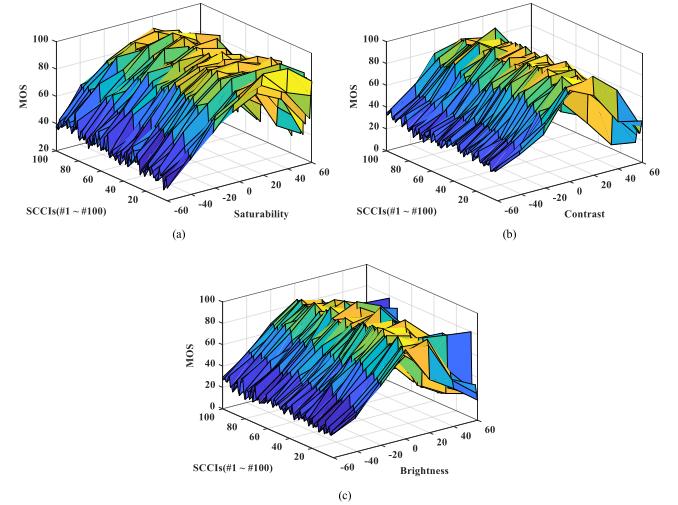


Fig. 7. The MOS values of all SCCIs against the (a) saturation level, (b) contrast level, and (c) brightness level.

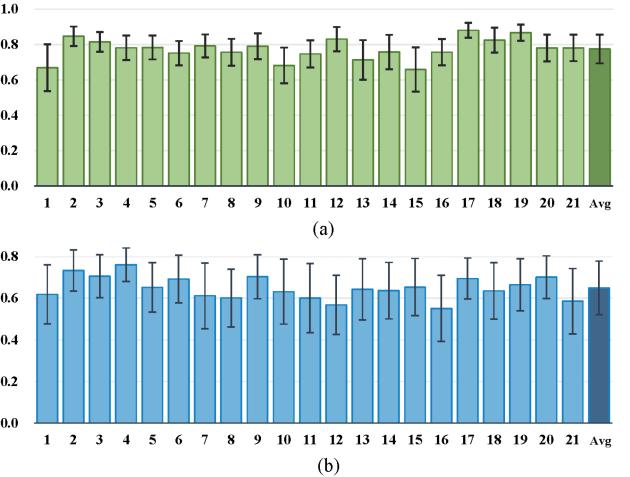


Fig. 8. SRCC performances of individual subjects for (a) SCCIs and (b) MCCIs sessions.

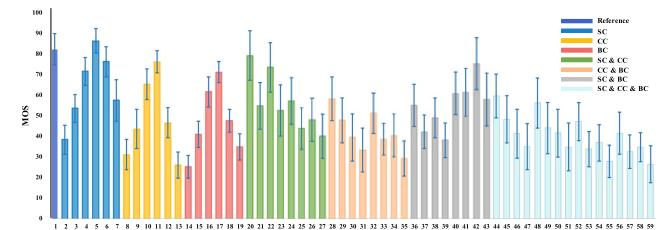


Fig. 9. Mean and standard of MOS values across all sub-sections.

degraded cartoon images in the same category, the standard deviation of MCCIs (denoted by the indexes from 20 to 59) is obviously higher than SCCIs (denoted by the indexes from 2 to 19), which suggests that the subjects show greater divergence and diversity when rated on the MCCIs. Therefore, evaluating the MCCIs is much challenging than the SCCIs. 2) Compared with the reference, saturation enhancement might be useful to create perceptually appealing results on cartoon images, especially when adding 20 on saturation shown in the 5-th index.

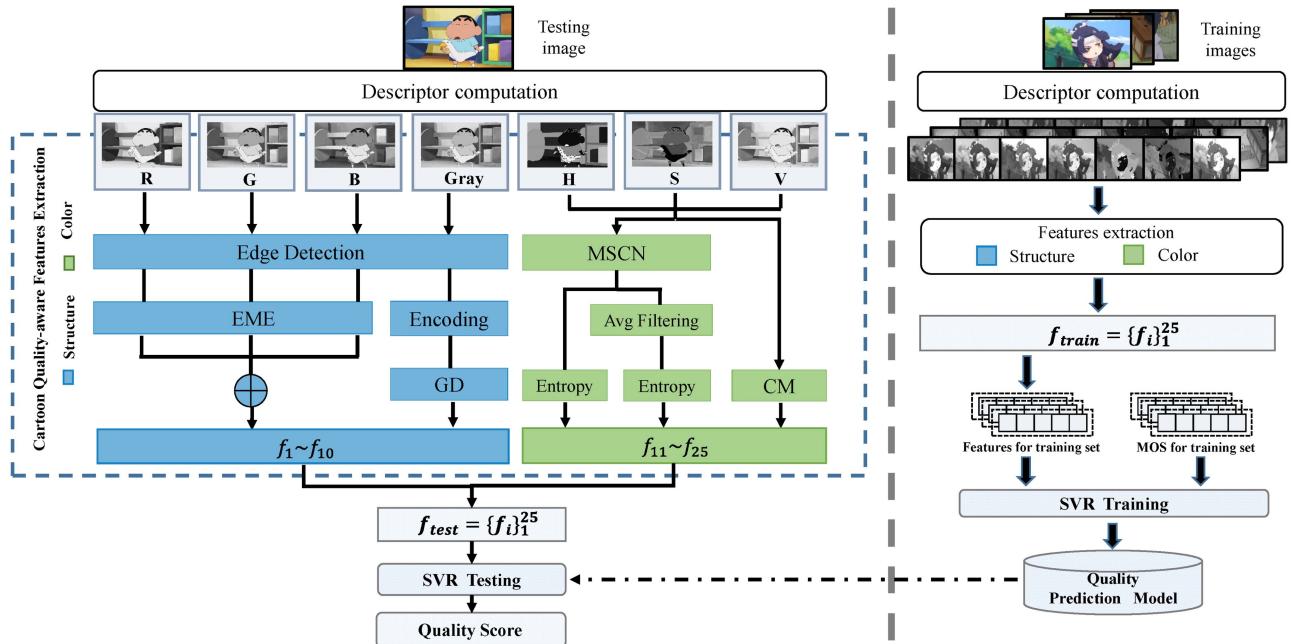


Fig. 10. Framework of our scheme.

The reason may be that the subjects prefer to the loss of color information rather than the detail information.

III. IMAGE QUALITY METRICS

In this paper, we propose a new NR quality prediction model for cartoon images, as shown in Fig. 10. The framework is mainly composed of two parts: feature extraction and quality regression. The structure and color features are extracted to measure the quality of the degraded cartoon images, and SVR is used to establish the quality predictor. The details of each component in the proposed model are described as follows.

A. Structure Feature Extraction

As known, human visual perception is highly adapted for extracting structural information from a scene [17]. The cartoon images usually contain sharp lines, clear outline edges, and locally simple details. The distortions induced by compression, noise, blur, contrast and brightness are bound to affect the sharpness and local details of cartoon images. For this, we extract the structure features from the gradient to perceive visual distortion of cartoon images.

First, we adopt Sobel filters [45] to compute the intensity of a gray image. The gradient map G can be computed by:

$$G = \sqrt{G_x^2 + G_y^2} \quad (4)$$

where G_x and G_y denote the intensity of gradient maps along the horizontal and vertical directions, computed as:

$$G_x = s_x * G = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * G \quad (5)$$

$$G_y = s_y * G = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * G \quad (6)$$

where s_x and s_y are Sobel convolution factors along the horizontal and vertical directions.

Further, according to the distribution of gradient maps, the LBP operator [46] is used to measure the distribution difference between a pixel and its neighbors in the gradient domain [47]:

$$LBP = \sum_{t=0}^{t=n-1} \psi(G_t - G_c) \quad (7)$$

$$\psi(X) = \begin{cases} 1, & \text{if } X \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where n denotes the number of the equi-spaced neighbors around a center pixel, G_c and G_t denote the gradient values of a center pixel and its surrounding pixels, respectively, and $\psi(X)$ is a thresholding function.

By the local binary encoding, the number of pixels in an area whose gradient values of the surrounding pixels are greater than the center pixel is counted. In order to combine intensity and distribution in the gradient domain, a combined metric model is computed as follows [17]:

$$GD_i = \sum_{y=1}^M \sum_{x=1}^N G(x, y) \delta(i - LBP(x, y)) \quad (9)$$

where M and N represent the size of the input image, δ is the Kronecker delta function, and $i \in [0, 8]$ denotes the index of pattern in the LBP map.

As a result, we can obtain a 9-dimensional feature vectors, denoted as $\mathbf{F}_{GD} = [GD_0, \dots, GD_8]$, to give a reasonable estimation of texture variation.

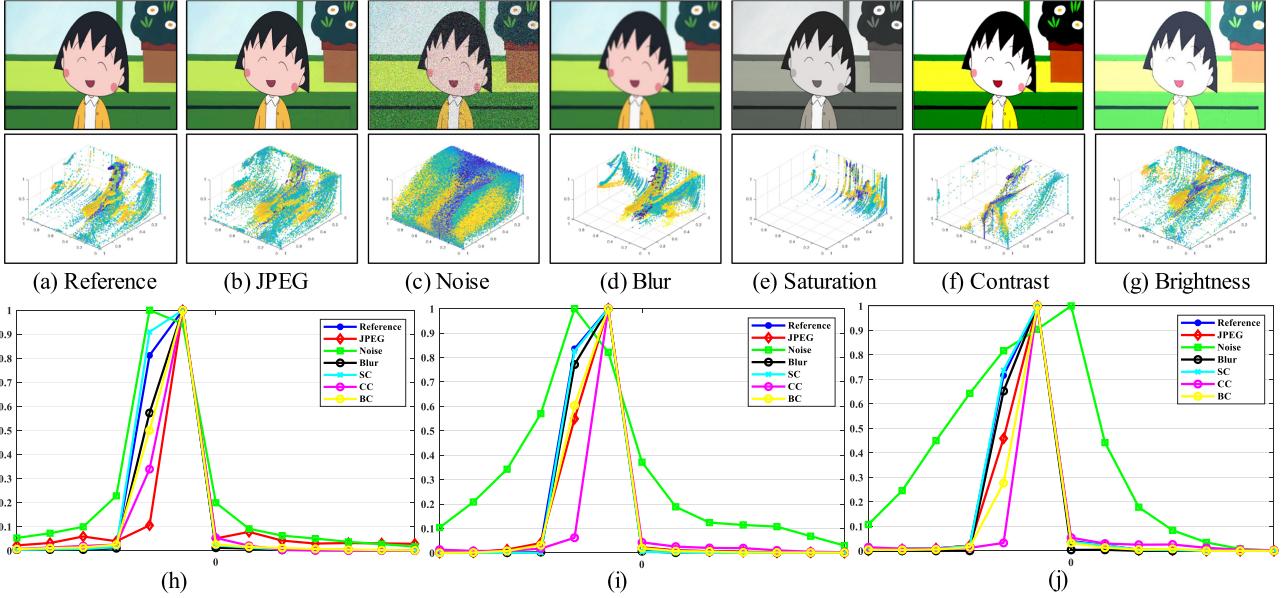


Fig. 11. The color features of the different distorted images. (a)-(g) Different distorted cartoon images with their corresponding similarity maps. First row: the source and different distorted images. Second row: the three-dimensional color space established by the values of H , S , V of the first row. (h)-(j): Distribution of MSCN coefficients of images with different distortion types in each channel map. (h) H channel. (i) S channel. (j) V channel.

To measure the edge quality in the gradient domain, we also implement Sobel filters on each RGB color component, and the binary edge maps are multiplied by the original values to obtain three grayscale edge maps on RGB components. Subsequently, we compute Weber contrast based Measure of Enhancement (EME) on the grayscale edge map [48], defined as:

$$EME = \frac{2}{k_1 k_2} \sum_{l=1}^{k_1} \sum_{k=1}^{k_2} \log \left(\frac{I_{\max,k,l}}{I_{\min,k,l}} \right) \quad (10)$$

where k_1 and k_2 are the number of blocks, $k_1 = M/5$, $k_2 = N/5$, $I_{\max,k,l}$ and $I_{\min,k,l}$ denote the local maximum and minimum in a block.

Finally, the edge quality score EQ is obtained by summing the $EMEs$ with different weights:

$$EQ = \sum_{c=1}^3 \lambda_c EME(\text{grayscale edge}_c) \quad (11)$$

where λ_1 , λ_2 and λ_3 denote the weights assigned for RGB components, respectively. In the experiment, we set $\lambda_1 = 0.299$, $\lambda_2 = 0.587$ and $\lambda_3 = 0.114$, similar to the weights in RGB to YUV conversion. As a result, we can obtain a 1-dimensional feature vector, denoted as $\mathbf{F}_{EQ} = [EQ]$, to give a reasonable estimation of sharpness on edges.

Overall, we can obtain 10-dimensional features to represent structure information, denoted by $\mathbf{F}_S = [\mathbf{F}_{GD}, \mathbf{F}_{EQ}]$.

B. Color Feature Extraction

In general, cartoon images are composed of large flat color blocks, and color information in these blocks is subject to be affected by different distortions. HSV color space has shown to be consistent with the characteristics of human vision system.

As shown in Fig. 11, it can be seen from the 3-dimensional color space established by the values of H , S , V that different distortion types will significantly change the values of H , S , V channels. Thus, in this work, HSV color space is used to extract color information from distorted cartoon images.

1) Color moments: Considering that color moments [49] are scale and rotation invariant, and color information is contained in low-order moments, we select the first-order to third-order central color moments to describe the color changes caused by different distortions. The first, second and third color moments are represented by mean, standard deviation and skewness of each channel map in a cartoon image, respectively. In particular, the mean represents the holistic tendency, the standard deviation captures the fluctuation of local contrasts, and the skewness reflects the information of global asymmetry, described as:

$$m(I) = \mathcal{M}(I) \quad (12)$$

$$d(I) = \sqrt{\mathcal{M}[(I - \mathcal{M}(I))^2]} \quad (13)$$

$$s(I) = \frac{\mathcal{M}[(I - \mathcal{M}(I))^3]}{d(I)^3} \quad (14)$$

where I denotes each channel map and \mathcal{M} denotes the mean operator. In this work, by extracting three moments in three channels, the features on all channel maps can be written as $\mathbf{F}_{CM} = [m_H, d_H, s_H, m_S, d_S, s_S, m_V, d_V, s_V]$.

2) Color entropy: The entropy measures the amount information contained in an image signal. Inspired by [50], we propose to use the color entropy to quantify the pixel characteristics of the HSV color space. Specifically, given a cartoon image, the MSCN coefficients are first used to eliminate the mutual interference of image neighborhood information in HSV color

TABLE III
THE USED FEATURES AND THEIR DEFINITIONS

Feature Group	Dim.	Feature Elements	Feature Descriptions
\mathbf{F}_S	\mathbf{F}_{GD}	GD_0, GD_1, \dots, GD_8	Structure features in gradient domain
	\mathbf{F}_{EQ}	EQ	
\mathbf{F}_C	\mathbf{F}_{CM}	$m_H, d_H, s_H, m_S, d_S, s_S, m_V, d_V, s_V$	Color features in HSV space
	\mathbf{F}_{CE}	$CE_H^C, CE_H^{CA}, CE_S^C, CE_S^{CA}, CE_V^C, CE_V^{CA}$	

space, denoted by I_C :

$$I_C(x, y) = \frac{I(x, y) - \mu(x, y)}{\sigma(x, y) + 1} \quad (15)$$

where $x \in \{1, 2, \dots, u\}$ and $y \in \{1, 2, \dots, v\}$ represents spatial indices with u and v as the image dimension. The local mean $\mu(x, y)$ and deviation $\sigma(x, y)$ are computed as follows.

$$\mu(x, y) = \sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} L(x+k, y+l) \quad (16)$$

$$\sigma(x, y) = \sqrt{\sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} [L(x+k, y+l) - \mu(x, y)]^2} \quad (17)$$

where w is a 2D circularly-symmetric Gaussian weighting filter ($K = L = 3$).

Then, an average filter is employed to reflect the structure characteristics in different color space, denoted by I_{CA} :

$$I_{CA}(x, y) = \frac{1}{HW} \sum_{h=-H}^H \sum_{w=-W}^W \varpi(h, w) I_C(x+h, y+w) \quad (18)$$

where $\{\varpi(h, w) | h = -H, \dots, H; w = -W, \dots, W\}$ denotes an average filter, and H and W represent the height and width of the averaging filter sampled out to 3 standard deviations ($H = W = 3$).

Finally, the entropies of the I_C and I_{CA} are calculated to obtain the color entropy on each channel map, denoted by CE_i^j :

$$CE_i^j = - \sum_{u=1}^{256} p_u(I_j) \log_2 p_u(I_j) \quad (19)$$

where $P_u(\cdot)$ denotes the frequency of intensity value u , $i \in \{H, S, V\}$ represents different HSV channel, and $j \in \{C, CA\}$ represents either image $I_C(x, y)$ or image $I_{CA}(x, y)$. Similarly, the features on each channel map can be represented as $\mathbf{F}_{CE} = [CE_H^C, CE_H^{CA}, CE_S^C, CE_S^{CA}, CE_V^C, CE_V^{CA}]$.

Overall, we can obtain 15-dimensional features to represent color information, denoted by $\mathbf{F}_C = [\mathbf{F}_{CM}, \mathbf{F}_{CE}]$.

C. Feature Analysis

After the above operations, we finally obtain two types of features to represent an individual cartoon image, as shown in Table III, in which \mathbf{F}_S represents the spatial structure in gradient domain and \mathbf{F}_C is used to quantify color features in HSV color space. In order to demonstrate different characteristics of BC, SC and CC, examples of feature distributions of cartoon

images with different MOS values are demonstrated in Fig. 12, in which the second and third row show the structure and color statistical histograms, respectively. It can be seen from the sets of CC and BC that the bars in each bin of the statistical histograms in low-quality cartoon image are shorter than those in high-quality cartoon image, which demonstrates that our statistical structure features can effectively capture the detailed information caused by CC and BC. However, the low-quality and high-quality SC cartoon images cannot be effectively distinguished from the structure histograms. In such situation, the color distributions can compensate the inaccurate structure measurement. Overall, the structure and color features are combined to capture more comprehensive information for CIQA.

D. Quality Regression

With the extracted structure features \mathbf{F}_S and color features \mathbf{F}_C , we combine these features to generate an overall quality score by employing SVR to build a quality prediction model. We use the LIBSVM package [51] to implement SVR with the radial basis function (RBF) kernel. Then, the trained model is utilized to predict the quality score of an arbitrary cartoon image.

IV. EXPERIMENTAL RESULTS

A. Databases and Experiment Description

Besides the NBU-CIQAD, the performance of our metric is also evaluated on another cartoon image database: HFUT-CID [1]. The database contains 200 reference images and 2000 distorted images corresponding to five distortion types: JPEG compression, Gaussian noise (GN), salt and pepper noise (SPN), Gaussian blur (GB), and AVS2 compression, with two levels for each distortion type.

To benchmark the performance of IQA metrics, four commonly used indicators, including Pearson Linear Correlation Coefficient (PLCC), SRCC, Kendall Rank-order Correlation Coefficient (KRCC), and Root Mean Square Error (RMSE), are calculated between the predicted objective scores and subjective scores, among which the SRCC and KRCC measure the prediction monotonicity, while the PLCC and RMSE measure the prediction accuracy. A superior metric should have higher PLCC, SRCC and KRCC values (with a maximum of 1) and lower RMSE (with a minimum of 0). For nonlinear regression between the predicted scores and subjective ratings, a five-parameter logistic function is used:

$$Q_p = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(Q - \beta_3)}} \right) + \beta_4 + \beta_5 \quad (20)$$

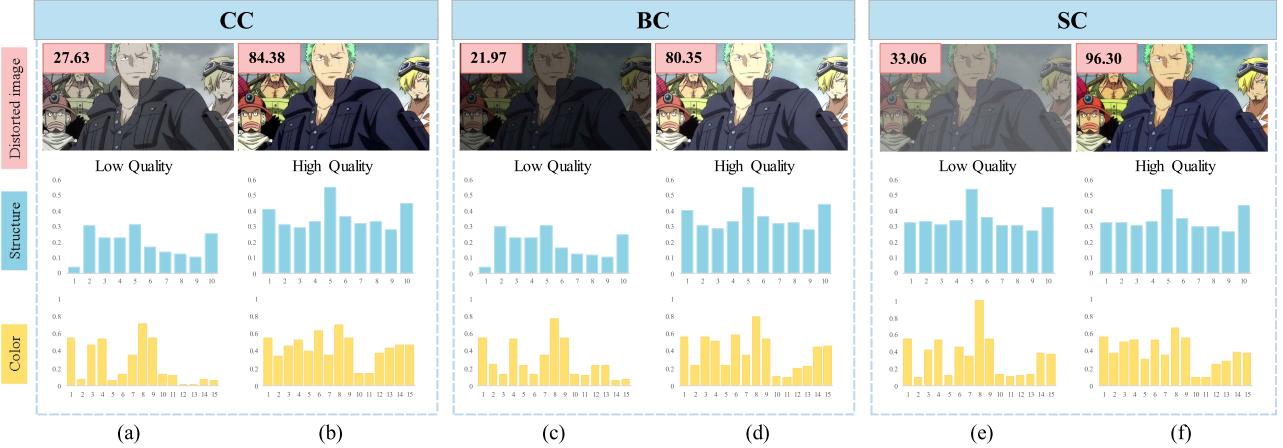


Fig. 12. Visual samples of statistical feature distributions. (a) and (b) are CC, (c) and (d) are BC, (e) and (f) are SC. MOS value of each distorted image is labeled on its top-left corn. Second row and third row denote the corresponding structure and color feature distributions in the first row, respectively.

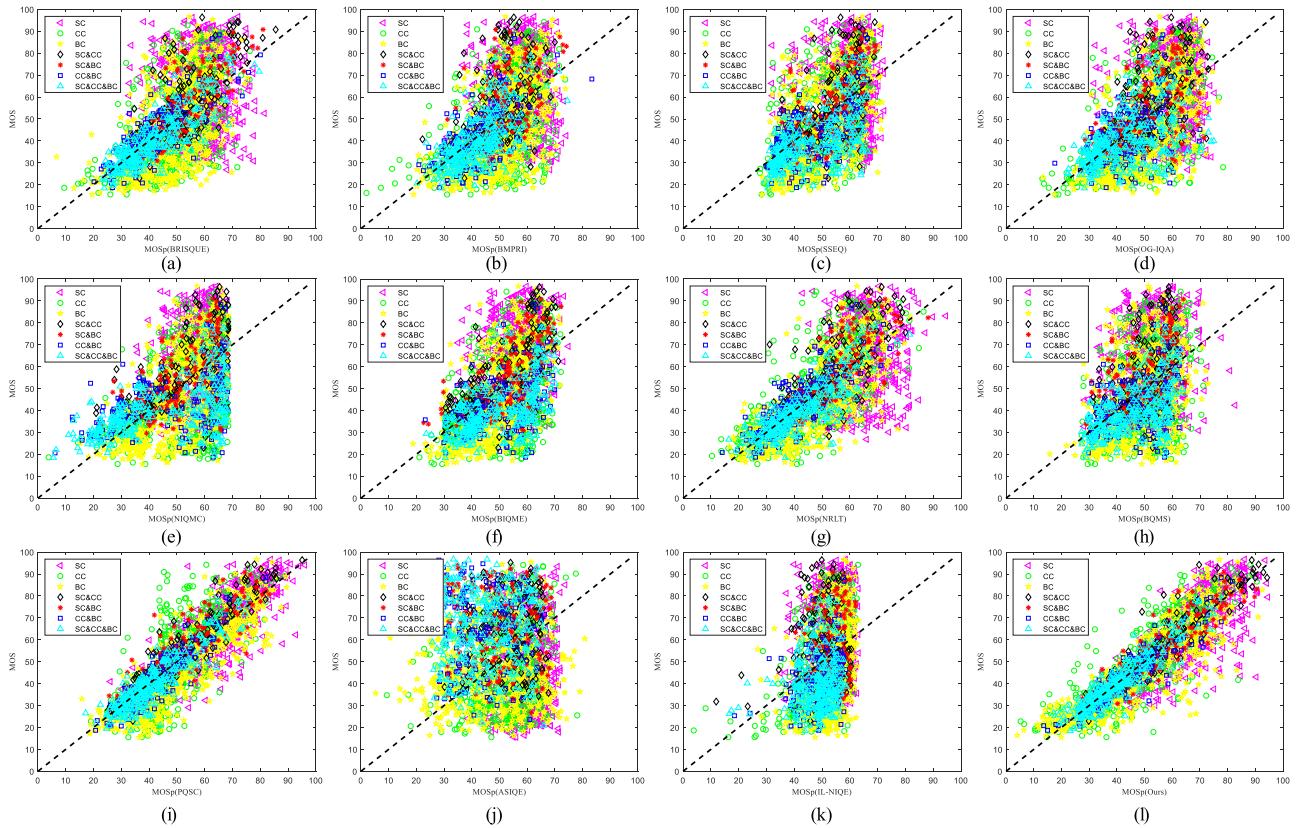


Fig. 13. Scatter of predicted quality scores by some metrics against the MOS values on the proposed database. (a) BRISQUE, (b) BMPRI, (c) OG-IQA, (d) SSEQ, (e) NIQMC, (f) BIQME, (g) NRLT, (h) BQMS, (i) PQSC, (j) ASIQUE, (k) ILNIQE, and (l) (OURS).

where $\{\beta_i | i = 1, 2, \dots, 5\}$ are five parameters determined by fitting.

B. Performance Comparison With Existing Methods

To objectively evaluate the performance of the proposed method, our metric is compared with nineteen state-of-the-art NR-IQA metrics: eleven metrics designed for natural images (e.g., BPRI [24], BMPRI [25], BIQI [21], BRISQUE [5],

NIQE [52], TCLT [22], IL-NIQE [53], OG-IQA [59], SSEQ [36], DBCNN [60], HyperIQA [61]), two metrics for contrast-changed (CC) natural images (e.g., NIQMC [19], BIQME [20]), four metrics for SCIs (e.g., BQMS [37], ASIQUE [38], NRLT [34], PQSC [39]), one metric for compressed natural, graphic, and SCIs (e.g., UCA [50]), and one metric for cartoon images (e.g., CBIQA [1]). In the implementation, similar to related studies [39], [62], we randomly divide each database into two subsets: 80% for training and the remaining 20% for testing. For a

TABLE IV
PERFORMANCE EVALUATION ON THE NBU-CIQAD DATABASE

Method	Design for	SCCIs				MCCIs				All			
		SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE	SRCC	KRCC	PLCC	RMSE
BPRI [24]	Natural Images	0.0465	0.0319	0.0458	20.4257	0.1840	-0.1342	0.1734	15.6522	0.0375	0.0267	0.0657	19.2476
BMPRI (retrain) [25]	Natural Images	0.3658	0.2517	0.3748	19.2412	0.6384	0.4831	0.6685	11.8891	0.5251	0.3648	0.5232	16.8002
BIQI [21]	Natural Images	0.0838	-0.0621	0.1132	20.3856	0.1243	0.0734	0.1678	15.3567	0.1245	0.0354	0.1425	19.2547
BRISQUE(retrain) [5]	Natural Images	0.4240	0.2912	0.4653	18.8972	0.8462	0.6585	0.8090	8.5305	0.5972	0.4215	0.5977	15.8657
NIQE [52]	Natural Images	0.1345	-0.0754	0.2057	20.1567	0.1345	-0.0734	0.1612	15.2158	0.1689	-0.0455	0.1956	19.2546
TCLT [22]	Natural Images	0.1456	-0.0458	0.1399	20.6897	0.1856	-0.1057	0.1711	15.2459	0.1456	-0.0677	0.1259	19.4879
IL-NIQE [53]	Natural Images	0.3264	-0.2058	0.3125	19.6789	0.2312	-0.1578	0.2954	15.1546	0.3450	-0.2154	0.3356	18.6789
OG-IQA (retrain) [59]	Natural Images	0.4464	0.3162	0.5123	17.8345	0.6189	0.4513	0.6173	12.1852	0.5729	0.4019	0.5771	16.1246
SSEQ (retrain) [36]	Natural Images	0.4537	0.3125	0.5012	17.9531	0.7446	0.5588	0.7350	10.4962	0.5762	0.4029	0.5768	16.1335
DBCNN (retrain) [60]	Natural Images	0.6125	0.4328	0.6312	15.9876	0.4225	0.2893	0.4464	18.5697	0.5226	0.3598	0.5321	16.8321
HyperIQA [61]	Natural Images	0.2859	0.1938	0.3014	19.8660	0.3817	0.2587	0.3447	14.5277	0.3294	0.2213	0.3383	18.6105
UCA [50]	Unified Images	0.0294	-0.0195	0.0697	20.7841	0.0018	-0.0026	0.1020	15.8945	0.0125	-0.1050	0.0865	19.5241
NIQMC [19]	Natural CC images	0.4485	0.3106	0.4846	18.1987	0.5125	0.3821	0.5356	13.2879	0.4698	0.3387	0.5112	17.0052
BIQME [20]	Natural CC images	0.5125	0.3578	0.5358	17.6587	0.5445	0.3759	0.5489	13.4025	0.5368	0.3698	0.5325	16.6657
BQMS (retrain) [37]	SCIs	0.2503	0.1718	0.3050	19.7892	0.6307	0.4518	0.6369	11.8835	0.3863	0.2663	0.4030	18.0944
ASIQUE (retrain) [38]	SCIs	0.4399	0.3056	0.4852	18.0946	0.7953	0.6030	0.7719	9.7944	0.6012	0.4200	0.6059	15.9623
NRLT (retrain) [34]	SCIs	0.6373	0.4580	0.6636	15.5388	0.8253	0.6436	0.8302	8.4555	0.7261	0.5233	0.7218	13.6523
PQSC (retrain) [39]	SCIs	0.7012	0.5123	0.7088	14.5362	0.8565	0.6720	0.8746	7.5213	0.8578	0.6721	0.8412	10.8947
CBIQA [1]	Cartoon images	0.1917	0.1420	0.1325	21.5789	0.4020	0.2915	0.3420	19.1089	0.2598	0.1864	0.1790	22.1568
Ours	Cartoon images	0.8278	0.6320	0.8289	11.6054	0.9285	0.7789	0.9050	5.8674	0.8628	0.6801	0.8511	10.3698

TABLE V
PERFORMANCE EVALUATION ON THE HFUT-CID

Method	SRCC	KRCC	PLCC	RMSE
BIQI(retrain)[21]	0.3376	0.2280	0.4638	0.1062
TCLT[22]	0.2891	0.1999	0.3909	0.1104
NFERM[23]	0.4429	0.3337	0.5317	0.1016
NFERM(retrain)	0.5330	0.3885	0.5502	0.1001
DIIVINE[54]	0.6756	0.5020	0.6394	0.0922
ILNIQE[53]	0.5804	0.4123	0.6151	0.0945
ILNIQE(retrain)	0.6612	0.4466	0.6364	0.0925
BRISQUE[5]	0.6455	0.4745	0.6171	0.0943
BRISQUE(retrain)	0.7455	0.5454	0.7075	0.0847
SISBLIM[55]	0.7585	0.5516	0.7710	0.0764
BPRI[24]	0.7004	0.4922	0.6994	0.0857
BMPRI[25]	0.5140	0.4063	0.6674	0.0893
BMPRI(retrain)	0.7977	0.5998	0.7815	0.0748
CBIQA[1]	0.8298	0.6678	0.8146	0.0696
Ours	0.8576	0.6682	0.8783	0.0572

fair comparison, the above partitioning process is repeated 1000 times, and the average value after 1000 iterations is reported as the final performance score. Note that the partitioning process is repeated ten times on the deep learning-based methods (e.g., DBCNN [60]). The metrics that involve training are directly re-trained on our database followed the same training-testing descriptions in their metrics. For the training-free methods, we directly use the models released by the authors to predict the image quality.

Table IV shows the experimental results on NBU-CIQAD. We have the following observations from the results: 1) The metrics designed for natural images are not satisfactory in evaluating the quality of cartoon images. Although the retrained BRISQUE metric performs relatively well but is still lower than

most metrics designed for SCIs in our database. 2) The metrics designed for contrast-changed natural and cartoon images are also performed worse because they are ineffective to measure the color-related distortions. 3) Some metrics designed for SCIs show great competitiveness, since the cartoon images and SCIs are similar in some characteristics. 4) It is evident that the proposed method is significantly superior to other methods designed either for natural images, cartoon images or SCIs, which validates the effectiveness of color and structure features in evaluating the SC, CC, BC distortions of cartoon images. Fig. 13 shows the scatter plots of the MOS values against the predicted quality scores by some representative IQA models (e.g., BRISQUE, BMPRI, OG-IQA, SSEQ, NIQMC, BIQME, NRLT, BQMS, PQSC, ASIQUE, ILNIQE, and the proposed method) on the NBU-CIQAD database. It is observed that the proposed metric has stronger convergence and monotonicity.

In addition, to verify our method is not only useful for the color-related distortions, we further perform the proposed method on the HFUT-CID. Table V represents the performance of the proposed methods against the existing IQA methods on the HFUT-CID. It is obvious that the proposed method is the best one among all metrics, indicating that the proposed method is also effective for the distortions of compression, blur, and noise.

C. Influence of Each Quality Component

In our metric, we extract structure and color features for quality assessment. In order to investigate the influence of each component, we separately evaluate the test performance using independent component on the two datasets, as shown in Table VI. From the table, we observe that: 1) No matter for SCCIs or MCCIs on NBU-CIQAD, color component is more effective than

TABLE VI
PERFORMANCE OF DIFFERENT COMPONENTS ON THE TWO DATABASES

Methods	NBU-CIQAD (SCCIs)			NBU-CIQAD (MCCIs)			NBU-CIQAD			HFUT-CID		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
W/ Structure	0.4476	0.4378	18.5046	0.6325	0.6544	11.8875	0.5002	0.4913	17.2014	0.9063	0.8850	0.0506
W/ Color	0.7898	0.7785	12.8654	0.8754	0.8941	7.3965	0.8025	0.8189	11.6425	0.8580	0.8362	0.0615
All	0.8289	0.8278	11.6054	0.9050	0.9285	5.8674	0.8511	0.8628	10.3698	0.8783	0.8576	0.0572

TABLE VII
PERFORMANCE (SRCC) EVALUATION ON EACH DISTORTION SET

Method	Design for	SC	CC	BC	SC&CC	SC&BC	BC&CC	SC&BC&CC
BPRI [24]	Natural Images	0.0205	0.0105	0.0017	0.0751	0.0546	0.3125	0.1956
BMPRI (retrain) [25]	Natural Images	0.0502	0.2258	0.3497	0.6510	0.5849	0.4319	0.6627
BIQI [21]	Natural Images	0.0605	0.1157	0.0952	0.0945	0.0453	0.0877	0.2602
BRISQUE (retrain) [5]	Natural Images	0.0938	0.3022	0.3737	0.7047	0.6759	0.4798	0.7386
NIQE [52]	Natural Images	0.0302	0.1752	0.2253	0.0103	0.1802	0.1705	0.1652
TCLT [22]	Natural Images	0.0468	0.0737	0.1856	0.1956	0.3185	0.1425	0.2865
IL-NIQE [53]	Natural Images	0.0102	0.4856	0.3598	0.1253	0.1314	0.3258	0.1562
SSEQ (retrain) [36]	Natural Images	0.0496	0.2919	0.2789	0.7068	0.6742	0.5004	0.7333
NIQMC [19]	Natural CC images	0.0501	0.3914	0.5856	0.6012	0.6231	0.4098	0.5501
BIQME [20]	Natural CC images	0.1512	0.4589	0.6548	0.7126	0.7322	0.4231	0.5689
BQMS (retrain) [37]	SCIs	0.0460	0.2970	0.4420	0.6771	0.7065	0.4309	0.7008
ASIQUE (retrain) [38]	SCIs	0.4341	0.5999	0.0538	0.4513	0.5912	0.0528	0.4727
NRLT (retrain) [34]	SCIs	0.0102	0.2046	0.1818	0.2973	0.4458	0.4256	0.3724
PQSC (retrain) [39]	SCIs	0.3686	0.5160	0.3923	0.7251	0.7204	0.5326	0.6725
CBIQA[1]	Cartoon images	0.1420	0.1125	0.2869	0.4356	0.5569	0.2480	0.4746
Ours	Cartoon images	0.4420	0.6792	0.5635	0.7692	0.7428	0.5498	0.7465

the structure component, and combination of two quality components achieves desirable performance on the database. 2) On the HFUT-CID, the performance of structure component is better than that of color component, but the combination weakens the performance compared with only using the structure component, because the HFUT-CID is mainly designed for the distortions of blur, compression, noise, which have a greater effect on the structure of the cartoon image than on the color of image. Our color component, although effective for the distortions on HFUT-CID, is primarily for color-related distortion (e.g., saturation, contrast). Overall, the combined metric allows for a more comprehensive prediction of various distortions.

D. Performance on Individual Distortion Types

To more comprehensively evaluate the performance of our metric, we compare fifteen metrics on different subsets of distortion types on NBU-CIQAD database. The SRCC results are listed in Table VII, where the best metric has been highlighted in boldface. One can see that the proposed metric achieves the best performance in six subsets than the other metrics, but is not very prominent for BC subset, because the structure features cannot effectively reflect the loss of detail information caused by BC. Another finding is that most existing NR-IQA methods encounter challenges to accurately predict the quality of the SC cartoon images, but they can work in multiple distortions involving SC. It could be explained by the fact that the effect of multiple distortions occurring at the same time on the cartoon

image is different from any single distortion, especially for color change. CC or BC dominates the effects of multiple distortions on the image, which can be intuitively seen from Fig. 5, the effect of SC on the image is weakened. It is apparent that the proposed method can accurately assess the multiple and single distortions with the best performance.

E. Statistical Significance

To assess the statistical significance between any two metrics, we further conduct t-test [56] to evaluate the statistical significance of the SROCC values between the proposed method and existing NR-IQA methods. Table VIII tabulates the t-test results conducted between our metric and other NR IQAs on the NBU-CIQAD dataset. The symbol “1”, “0”, or “-” means that our metric is statistically better, worst or indistinguishable than another metric, respectively. It is clearly shown that our metric is statistically better than all existing NR IQAs.

F. Ranking Performance Comparison

To comprehensively compare the performance of the IQA metrics, refer to [57], [58], we also focus on the rankings of NR-IQA according to the subjective and objective scores on our NBU-CIQAD dataset. We select 8 distorted images of the same scene from the dataset, and the rankings are obtained based on the objective scores generated by the corresponding methods. Table IX presents the rankings of eight distorted images

TABLE VIII
STATISTICAL SIGNIFICANCE COMPARISON OF OUR METRIC WITH TWELVE EXISTING NR IQAS

Ours	BPRI	BMPRI	BIQI	BRISQUE	NIQE	TCLT	IL-NIQE	NIQME	BIQME	CBIQA	NRLT	PQSC
S	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1

TABLE IX

RANKING RESULTS OF DIFFERENT EVALUATION METHODS ACCORDING TO MOS VALUE AND OBJECTIVE SCORES ON OUR NBU-CIQAD DATABASE. 1 TO 8 CORRESPOND TO THE BEST TO WORST PERFORMANCES. INCONSISTENT RANKING RESULTS ARE IN RED BOLD

MCCIs #1 ~ #8								
MOS	86.25	73.06	59.62	52.58	40.80	31.66	26.23	20.62
RANK	1	2	3	4	5	6	7	8
NIQE	1	3	5	2	6	7	8	4
BIQI	5	1	2	3	4	8	6	7
BMPRI	2	1	3	5	6	8	7	4
BPRI	7	5	4	1	8	3	2	6
BRISQUE	4	1	6	5	2	7	8	3
TCLT	3	8	7	2	6	4	5	1
OG-IQA	4	3	2	6	5	8	7	1
NFERM	2	4	5	6	3	7	8	1
CBIQA	3	4	5	6	2	7	8	1
NIQME	1	4	6	5	3	7	8	2
BIQME	1	3	5	6	2	8	7	4
NRLT	1	2	3	6	5	4	7	8
PQSC	1	2	3	4	6	5	7	8
Ours	1	2	3	4	5	6	7	8

evaluated by different metrics. It is observed that our method has more consistent ranking than other metrics compared with the ground-truth MOS distribution on the NBU-CIQAD dataset. Since one of the most important applications of IQA metric is to guide the generation of cartoon images, the proposed method is a promising tool for automatic selection of the best cartoon image.

V. CONCLUSION

In this study, we have proposed a quality prediction model to automatically predict the quality of the cartoon image. For evaluation purpose, we construct a new cartoon image database named NBU-CIQAD with color change. The model extracts texture and edge information in gradient domain to give a reasonable estimation of structural variation, and extracts moment and entropy features in HSV color space to capture color information. Our results show that our proposed model is promising in handing quality assessment problem for distorted cartoon images. In future, we would like to conduct works along the following aspects: 1) It is necessary to dig deep feature and quality representation of cartoon images. 2) With the rise of 3D animation, it is worthwhile to effectively extend the quality prediction model to 3D images.

REFERENCES

- [1] Y. Chen *et al.*, “Blind quality assessment for cartoon images,” *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 30, no. 9, pp. 3282–3288, Sep. 2020.
- [2] K. Ma, K. Zeng, and Z. Wang, “Perceptual quality assessment for multi-exposure image fusion,” *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
- [3] Y. Fang, H. Zhu, K. Ma, Z. Wang, and S. Li, “Perceptual evaluation for multi-exposure image fusion of dynamic scenes,” *IEEE Trans. Image Process.*, vol. 29, pp. 1127–1138, 2020.
- [4] K. Gu, G. Zhai, W. Lin, and M. Liu, “The analysis of image contrast: From quality assessment to automatic enhancement,” *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2016.
- [5] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [7] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
- [8] Z. Ni *et al.*, “ESIM: Edge similarity for screen content image quality assessment,” *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4818–4831, Oct. 2017.
- [9] L. Ma, S. Li, F. Zhang, and K. N. Ngan, “Reduced-reference image quality assessment using reorganized DCT-based image representation,” *IEEE Trans. Multimedia*, vol. 13, no. 4, pp. 824–829, Aug. 2011.
- [10] X. Min, K. Gu, G. Zhai, M. Hu, and X. Yang, “Saliency-induced reduced-reference quality index for natural scene and screen content images,” *Signal Process.*, vol. 145, pp. 127–136, 2018.
- [11] G. Zhai and X. Min, “Perceptual image quality assessment: A survey,” *Sci. China Ser. F: Inf. Sci.*, vol. 63, no. 11, pp. 1–52, Nov. 2020.
- [12] L. Liang *et al.*, “No-reference perceptual image quality metric using gradient profiles for JPEG2000,” *Signal Process.-Image Commun.*, vol. 25, no. 7, pp. 502–516, Aug. 2010.
- [13] Y. Zhan and R. Zhang, “No-reference JPEG image quality assessment based on blockiness and luminance change,” *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 760–764, Jun. 2017.
- [14] L. Li, H. Zhu, G. Yang, and J. Qian, “Referenceless measure of blocking artifacts by tchebichef kernel analysis,” *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 122–125, Jan. 2014.
- [15] X. Wang *et al.*, “Exploiting local degradation characteristics and global statistical properties for blind quality assessment of tone-mapped HDR images,” *IEEE Trans. Multimedia*, vol. 23, pp. 692–705, 2021.
- [16] Q. Jiang, Z. Peng, G. Yue, H. Li, and F. Shao, “No-reference image contrast evaluation by generating bi-directional pseudo references,” *IEEE Trans. Ind. Inform.*, vol. 7, no. 9, pp. 6062–6072, Sep. 2021. doi: [10.1109/TII.2020.3035448](https://doi.org/10.1109/TII.2020.3035448).

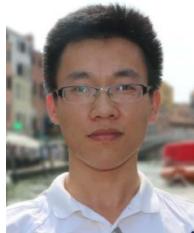
- [17] W. Lyu, W. Lu, and M. Ma, "No-reference quality metric for contrast-distorted image based on gradient domain and HSV space," *J. Vis. Commun. Image Representation*, vol. 69, May 2020, Art. no. 102797.
- [18] K. A. Panetta, E. J. Wharton, and S. S. Agaian, "Human visual system-based image enhancement and logarithmic contrast measure," *IEEE Trans. Syst., Man, Cybern., Part B (Cybern.)*, vol. 38, no. 1, pp. 174–188, Feb. 2008.
- [19] K. Gu *et al.*, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4559–4565, Dec. 2017.
- [20] K. Gu, D. Tao, J. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 1301–1313, Apr. 2018.
- [21] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.
- [22] Q. Wu *et al.*, "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 26, no. 3, pp. 425–440, Mar. 2016.
- [23] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.
- [24] X. Min *et al.*, "Blind quality assessment based on pseudo-reference image," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2049–2062, Aug. 2018.
- [25] X. Min, G. Zhai, K. Gu, Y. Liu, and X. Yang, "Blind image quality estimation via distortion aggravation," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 508–517, Jun. 2018.
- [26] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
- [27] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Trans. Image Process.*, vol. 30, no. 1, pp. 3474–3486, Mar. 2021.
- [28] M. Xu, L. Jiang, C. Li, Z. Wang, and X. Tao, "Viewport-based CNN: A multi-task approach for assessing 360° video quality," *IEEE Trans. Pattern Anal. Mach. Intell.*, Oct. 2021, doi: [10.1109/TPAMI.2020.3028509](https://doi.org/10.1109/TPAMI.2020.3028509).
- [29] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "MetaIQA: Deep meta-learning for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 14143–14152.
- [30] X. Min *et al.*, "Unified blind quality assessment of compressed natural, graphic, and screen content images," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5462–5474, Nov. 2017.
- [31] K. Gu *et al.*, "Evaluating quality of screen content images via structural variation analysis," *IEEE Trans. Visual. Comput. Graph.*, vol. 24, no. 10, pp. 2689–2701, Nov. 2017.
- [32] T. Li *et al.*, "Subjective and objective quality assessment of compressed screen content videos," *IEEE Trans. Broadcast.*, vol. 67, no. 2, pp. 438–449, Jun. 2021.
- [33] T. Li, X. Min, W. Zhu, Y. Xu, and W. Zhang, "No-reference screen content video quality assessment," *Displays*, vol. 69, Sep. 2021, Art. no. 102030.
- [34] Y. Fang, J. Yan, L. Li, J. Wu, and W. Lin, "No reference quality assessment for screen content images with both local and global feature representation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1600–1610, Apr. 2018.
- [35] F. Shao, Y. Gao, F. Li, and G. Jiang, "Toward a blind quality predictor for screen content images," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 48, no. 9, pp. 1521–1530, Sep. 2018.
- [36] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process.: Image Commun.*, vol. 29, no. 8, pp. 856–863, 2014.
- [37] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Learning a blind quality evaluation engine of screen content images," *Neurocomputing*, vol. 196, pp. 140–149, Jul. 2016.
- [38] K. Gu *et al.*, "No-reference quality assessment of screen content pictures," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4005–4018, Aug. 2017.
- [39] Y. Fang, R. Du, Y. Zuo, W. Wen, and L. Li, "Perceptual quality assessment for screen content images by spatial continuity," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 30, no. 11, pp. 4050–4063, Nov. 2020.
- [40] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, Nov. 2012, pp. 1693–1697.
- [41] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. Cormack, "Image and video quality assessment research at LIVE," 2006, [Online]. Available: <http://live.ece.utexas.edu/research/quality/>
- [42] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack, "Image and Video Quality Assessment Research at LIVE," [Online]. Available: <http://live.ece.utexas.edu/research/quality/>
- [43] A. M. van Dijk, J. B. Martens, and A. B. Watson, "Quality assessment of coded images using numerical category scaling," *Proc. SPIE*, vol. 2451, pp. 90–101, Mar. 1995.
- [44] X. Min, G. Zhai, J. Zhou, M. C. Q. Farias, and A. C. Bovik, "Study of subjective and objective quality assessment of audio-visual signals," *IEEE Trans. Image Process.*, vol. 29, pp. 6054–6068, 2020.
- [45] I. Sobel, "Camera models and machine perception," Stanford Univ., 1970.
- [46] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [47] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 541–545, Apr. 2016.
- [48] K. Panetta, A. Samani, and S. Agaian, "Choosing the optimal spatial domain measure of enhancement for mammogram images," *Int. J. Biomed. Imag.*, vol. 2014, Aug. 2014, Art. no. 937849.
- [49] X. Duanmu, "Image retrieval using color moment invariant," in *Proc. IEEE Inf. Technol.: New Generations*, Apr. 2010, doi: [10.1109/ITNG.2010.231](https://doi.org/10.1109/ITNG.2010.231).
- [50] L. Xu, Q. Chen, and Q. Wang, "Application of color entropy to image quality assessment," *J. Image Graph.*, vol. 20, no. 10, pp. 1583–1592, 2015.
- [51] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011, Art. no. 27.
- [52] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
- [53] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
- [54] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.
- [55] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *IEEE Trans. Broadcast.*, vol. 60, no. 3, pp. 555–567, Sep. 2014.
- [56] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.
- [57] S. Tian, L. Zhang, L. Morin, and O. Déforges, "NIQSV+: A no-reference synthesized view quality assessment metric," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1652–1664, Apr. 2018.
- [58] X. Wang, F. Shao, Q. Jiang, M. Chao, and Y. S. Ho, "Measuring coarse-to-fine texture and geometric distortions for quality assessment of DIBR-synthesized images," *IEEE Trans. Multimedia*, May 2020, doi: [10.1109/TMM.2020.2993942](https://doi.org/10.1109/TMM.2020.2993942).
- [59] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Signal Process.: Image Commun.*, vol. 40, pp. 1–15, 2016.
- [60] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 30, no. 1, pp. 36–47, Jan. 2020.
- [61] S. Su *et al.*, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3664–3673, 2020.
- [62] L. Shi, W. Zhou, Z. Chen, and J. Zhang, "No-reference light field image quality assessment based on spatial-angular measurement," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 30, no. 11, pp. 4114–4128, Nov. 2020.



Hangwei Chen received the B.S. degree from Ningbo University, Ningbo, China, in 2020. He is currently working toward the M.S. degree with Ningbo University, Ningbo, China. His current research interest focuses on quality assessment.



Xiongli Chai received the B.S. and M.S. degrees from Ningbo University, Ningbo, China, in 2017 and 2020, respectively. He is currently working toward the Ph.D. degree in signal and information processing with Ningbo University, Ningbo, China. His current research interests include image/video processing and quality assessment.



Xiangchao Meng (Member, IEEE) received the B.S. degree in geographic information system from the Shandong University of Science and Technology, Qingdao, China, in 2012, and the Ph.D. degree in cartography and geography information system from Wuhan University, Wuhan, China, in 2017. He is currently a Lecturer with the Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, China. His research interests include variational methods and remote sensing image fusion.



Feng Shao (Member, IEEE) received the B.S. and Ph.D degrees from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively, all in electronic science and technology. He is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. From February 2012 to August 2012, he was a Visiting Fellow with the School of Computer Engineering, Nanyang Technological University, Singapore. He has authored or coauthored more than 100 technical articles in refereed journals and proceedings in the areas of 3D video coding, 3D quality assessment, and image perception, etc. He was the recipient of the ‘Excellent Young Scholar’ Award by NSF of China (NSFC) in 2016.



Yo-Sung Ho (Fellow, IEEE) received the B.S. and M.S. degrees in electronic engineering from Seoul National University, Seoul, South Korea, in 1981 and 1983, respectively, and the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 1990. He joined Electronics and Telecommunications Research Institute (ETRI), Daejon, South Korea, in 1983. From 1990 to 1993, he was with Philips Laboratories, Briarcliff Manor, NY, where he was involved in development of the Advanced Digital High-Definition Television

(AD-HDTV) system. In 1993, he rejoined the technical staff of ETRI and was involved in development of the Korean DBS digital television and high-definition television systems. Since 1995, he has been with the Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea, where he is currently a Professor of Information and Communications Department. His research interests include digital image and video coding, image analysis and image restoration, advanced video coding techniques, digital video and audio broadcasting, three-dimensional video processing, and content-based signal representation and processing.



Xuejin Wang received the B.S. and M.S. degrees from Fuzhou University, Fuzhou, China, in 2012 and 2015, respectively, and the Ph.D. degree from Ningbo University, Ningbo, China, in 2021. She is currently a Lecturer with the School of Computer Science and Mathematics, Fujian University of Technology, Fuzhou, China. Her current research interests include image/video processing and quality assessment.



Qiuping Jiang (Member, IEEE) received the Ph.D. degree from Ningbo University, Ningbo, China, in June 2018. He is currently an Associate Professor with the School of Information Science and Engineering, Ningbo University, Ningbo, China. From January 2017 to June 2018, he was a Visiting Student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include image processing, visual perception modeling, and computer vision. He was the recipient of the JVCI 2017 Best Paper Award Honorable Mention as the first author. He is a Reviewer for several prestigious journals and conferences, such as the IEEE T-NNLS, IEEE T-IP, IEEE T-CSVT, IEEE T-MM, IEEE T-SIPN, ICME, ICIP, etc.