Perceptual Quality Assessment of Enhanced Colonoscopy Images: A Benchmark Dataset and an Objective Method

Guanghui Yue, *Member, IEEE*, Di Cheng, Tianwei Zhou, Jingwen Hou, Weide Liu, Long Xu, *Member, IEEE*, Tianfu Wang, and Jun Cheng, *Senior Member, IEEE*

***Abstract*—In colonoscopy, the captured images are usually with low-quality appearance, such as non-uniform illumination, low contrast, etc., due to the specialized imaging environment, which may provide poor visual feedback and bring challenges to subsequent disease analysis. Many low-light image enhancement (LIE) algorithms have recently proposed to improve the perceptual quality. However, how to fairly evaluate the quality of enhanced colonoscopy images (ECIs) generated by different LIE algorithms remains a rarely-mentioned and challenging problem. In this study, we carry out a pioneering investigation on perceptual quality assessment of ECIs. Firstly, considering the lack of specific datasets, we collect 300 low-light images with diverse contents during the real-world colonoscopy and conduct rigorous subjective studies to compare the performance of 8 popular LIE methods, resulting in a benchmark dataset (named ECIQAD) for ECIs. Secondly, in view of the distinctive distortion characteristics of ECIs, we propose an effective no-reference Enhanced Colonoscopy Image Quality (ECIQ) method to automatically evaluate the perceptual quality of ECIs via analysis of brightness, contrast, colorfulness, naturalness, and noise. Extensive experiments on ECIQAD demonstrate the superiority of our proposed ECIQ method over 14 mainstream no-reference image quality assessment methods.**

***Index Terms*—Colonoscopy images, image quality assessment, subjective assessment, no-reference, low-light image enhancement**

This work was supported in part by Shenzhen Science and Technology Program (No. RCBS20200714114920379), in part by Guangdong Basic and Applied Basic Research Foundation (Nos. 2021A1515011348, 2019A1515111205, 2019A1515110401), in part by National Natural Science Foundation of China (Nos. 62001302, 62103286), in part by Natural Science Foundation of Shenzhen (No. JCYJ20190808145011259), in part by Tencent "Rhinoceros Birds" - Scientific Research Foundation for Young Teachers of Shenzhen University, and in part by Social Science Youth Foundation of Ministry of Education of China (No. 21YJC630181). (Corresponding author: Tianwei Zhou.)

G. Yue, D. Cheng, and T. Wang are with the Marshall Laboratory of Biomedical Engineering, Shenzhen University, and also with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China (e-mail: yueguanghui@szu.edu.cn; chengdigogogo@outlook.com; tfwang@szu.edu.cn).

T. Zhou is with the College of Management, Shenzhen University, Shenzhen 518060, China (e-mail: tianwei@szu.edu.cn).

J. Hou is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore (e-mail: jingwen003@e.ntu.edu.sg.

L. Xu is with the Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China (e-mail: lxu@nao.cas.cn).

J. Cheng and W. Liu are with Institute for Infocomm Research, A*STAR, Singapore 138632 (e-mail: sam.j.cheng@gmail.com; weide001@e.ntu.edu.sg).

I. Introduction

RECENTLY, colorectal cancer has become one of the main killers threatening human life [1]. Early screening, diagnosis, and treatment of colorectal diseases via endoscopy (e.g., wireless capsule endoscopy and colonoscopy) are clinically considered as effective ways to prevent colorectal cancer [2]–[7]. In colonoscopy, the captured images are usually with low-quality appearance, such as non-uniform illumination, low contrast, biased color, etc., resulting from the specialized imaging process in the dark intestinal environment [8]. Generally, low-quality images impede the visual interpretation of colon surface and increase the misdiagnosis rate of subsequent disease analysis [9]. To improve the perceptual quality, many low-light image enhancement (LIE) algorithms have been proposed recently [10]–[14]. However, the LIE method's performance varies greatly when dealing with different images. Thus, effective quality assessment methods are necessary to select optimal enhanced colonoscopy images (ECIs) for better visual experience and disease analysis.

In the literature, image quality assessment (IQA) has been widely discussed in natural scene images (NSIs) [15]–[18]. It can be classified into subjective and objective methods. The former visually compares the image quality by a group of qualified observers, and its outcome accurately reflects the perceptual feelings. In the past years, many works have been conducted to subjectively compare the quality of NSIs with synthesized distortions [19] or authentical distortions [20]. However, the comparison of ECIs generated by different LIE algorithms has been rarely systematically discussed. To investigate the effectiveness of their proposed LIE methods, most works usually present a limited number of ECIs (e.g., dozens of cases) for subjective comparison [21], [22], thereby easily making the qualitative evaluation biased. To this end, it is of vital significance to construct large-scale ECI quality assessment datasets by rigorous subjective studies to fairly compare the performance of different LIE algorithms as well as to understand the distortion characteristics of ECIs, which provides guidance for the design of objective methods.

Since no perfect counterpart of ECIs is available during colonoscopy, the quality of ECIs should be objectively evaluated in a no-reference (NR) manner. Currently, many NR methods have been proposed for general-purpose and distortion-specific IQA tasks. Methods for the former task aim to effectively evaluate the quality of images with dif-

ferent distortions simultaneously. In this direction, extensive efforts have been contributed from the perspectives of natural scene statistics [23], dictionary construction [24], entropy computation [25], etc. Compared to general-purpose methods, distortion-specific methods exhibit more performance advantages by analyzing the intrinsical characteristics of a specific distortion. Recent years have seen the prosperity of specific-purpose methods, such as representative works in evaluating images with blur distortion [26], images with noise [27], and images with contrast change [28], [29]. Overall, these works demonstrate the remarkable success of NSI quality assessment. However, unfortunately, related works for ECIs has been rarely reported. To date, only a few attempts have been made to investigate the quality of compressed telesurgery videos [30] and synthetically distorted laparoscopic images [31]. Note that neither compression artifacts nor synthesized distortions can well represent the characteristics of real-world ECIs. Currently, one widely-used way for performance comparisons of LIE algorithms is directly adopting classical general-purpose or contrast-specific quality assessment methods designed for NSIs, without considering the suitability of these methods [9]. As observed in Section V, most IQA methods designed for NSIs cannot fairly compare the performance of LIE algorithms, indicating the urgent need of designing specific methods for ECIs.

In this study, we carry out a pioneering investigation on perceptual quality assessment of ECIs. Specifically, we firstly construct an Enhanced Colonoscopy Image Quality Assessment Dataset (ECIQAD) to fill the gaps in this field. To keep the dataset in line with clinical practice, we collect 300 low-light images with diverse contents during real-world colonoscopy and process them with 8 popular LIE algorithms, totally resulting in 2,400 images. For each image, its quality is reported in the form of mean opinion score (MOS) through rigorous subjective studies. Then, after carefully analyzing the distortion characteristics, we propose a new objective method, termed ECIQ, to automatically evaluate the quality of ECIs from the perspectives of characterizing brightness, contrast, colorfulness, naturalness, and noise. Finally, we investigate the effectiveness of our ECIQ as well as 14 mainstream NR IQA methods designed for NSIs on the constructed ECIQAD. Experimental results show the superiority of our method over competing methods. In summary, the main contributions of this paper are twofold:

- **ECI benchmark dataset with subjective scores**: we construct a real-world ECI quality assessment dataset, including 2,400 ECIs with rich intestinal scenes and subjective annotations. The constructed dataset can be utilized as a benchmark to develop and evaluate objective IQA methods of ECIs. To the best of our knowledge, this is the first attempt to systematically point out and build a platform for fair performance comparisons between different LIE algorithms.
- **Objective method for evaluating ECIs**: By summarizing the intrinsical distortion characteristics of ECIs, we propose an effective Enhanced Colonoscopy Image Quality (ECIQ) method that can objectively compare LIE algorithms in terms of the perceptual quality of their generated ECIs. Extensive experiments on the constructed ECIQAD demonstrate that our proposed ECIQ method achieves higher consistency with subjective behaviors than 14 off-the-shelf competing methods.

II. Related Works

In this section, first, we briefly review the datasets in quality assessment of endoscopic images. Then, we introduce some representative works of general-purpose NR IQA methods, followed by the review of contrast-specific NR IQA methods.

A. Datasets in Endoscopic IQA Tasks

In the literature, subjective evaluation is widely used for comparing the performance of LIE algorithms. For example, Li and Meng [22] compared their proposed LIE method with four methods via subjective studies on 400 wireless capsule endoscopy images. Regrettably, the authors did not give details of the scoring rule on evaluating enhanced images, and their testing dataset is not publicly available. Later, a more concise way is to conduct qualitative evaluation on a limited number of enhanced endoscopic images [9], [21]. Obviously, the drawn conclusion by such an approach may be biased and controversial because the presented images are usually subjectively selected. Lévêque *et al*. [30] conducted subjective studies to construct the quality dataset of telesurgery videos with synthesized compression artifacts and transmission errors. Khan *et al*. [31] treated the IQA problem as the classification of distortion type and level and reported a dataset with laparoscopic images processed by five synthesized distortions under four levels of severity, including defocus blur, motion blur, noise, uneven illumination, and smoke. Note that, the synthetically distorted images hold distinct artifacts compared to enhanced images. In addition, simplifying IQA into classification tasks cannot provide continuous scores to accurately compare different LIE algorithms. Thus, a public benchmark dataset with convinced subjective scores is highly necessary for fairly comparing LIE algorithms and summarizing distortion characteristics of enhanced endoscopic images.

B. General-Purpose NR IQA Methods

Attracted by the advantages in effectively evaluating images with multiple distortions simultaneously, scholars have made great efforts in designing general-purpose NR IQA methods recently. For instance, inspired by the fact that distortions could corrupt the regularity of natural statistics, Mittal *et al*. [23] estimated the quality of images by quantifying statistic features in the spatial domain. In contrast, Ghadiyaram and Bovik [32] mined richer statistical features from multiple domains to better characterize image distortions. In [33], a group of features were extracted from a corpus of high-quality images and fitted with a multivariate Gaussian model (named natural model). The quality of a query image was estimated by computing the distance between its multivariate Gaussian model and the natural model. Given that the amount of image's information would be changed with the addition of distortions, Liu *et*

al. [25] established an IQA model by modeling local spatial and spectral entropy. Min *et al*. [34] designed a novel IQA framework by calculating the similarity between the distorted image and its pseudo reference. Ye *et al*. [35] utilized raw-image-patches to extract local features and constructed a visual codebook for the IQA task by performing K-means clustering on local features. Jiang *et al*. [24] introduced a codebook-based IQA method by optimizing multistage discriminative dictionaries. Gu *et al*. [36] proposed a NR IQA method by extracting free-energy-based features, structural information, gradient magnitude, and naturalness-related features. Inspired by the frequency-dependent property of visual cortex, Zhou *et al*. [37] designed an IQA method for omnidirectional images via analysis of local-global naturalness and multifrequency information. Although very versatile, general-purpose methods are sometimes not very effective in evaluating images with a specific distortion.

C. Contrast-Specific NR IQA Methods

Contrast-specific methods are specifically designed for evaluating contrast-changed images by analyzing the image characteristics induced by contrast change. In the past years, several NR methods have been proposed for evaluating the quality of contrast-changed NSIs. For instance, Panetta *et al*. [38] proposed a contrast metric based on image statistics. Considering that enhancement algorithms would corrupt the intensity distribution, Fang *et al*. [39] built an IQA model by modeling the regularity of the mean, standard deviation, skewness features, kurtosis, and entropy. Likewise, Nafchi and Cheriet [28] measured Minkowski distance and entropy for evaluating the quality of contrast-changed images. Gu *et al*. [40] proposed a new IQA method based on global and local feature analysis. Later, they also designed an IQA method by learning a regression model with big-data samples [41]. Yue *et al*. [42] introduced a method for evaluating contrast-changed images by considering local maximum entropy and global intensity distribution change. Zhou *et al*. [43] designed a new method by analyzing structures from the aspects of spatial intensity, spatial distribution, and orientation of structures. Jiang *et al*. [29] evaluated the quality of enhanced images by extracting quality-aware features from the luminance and chrominance components. Overall, the literature has accumulated some remarkable quality assessment methods for contrast-changed NSIs. However, as illustrated in Section V, most existing methods for NSIs cannot be well competent for the quality assessment task of ECIs. Thus, it is crucial to design specialized IQA metrics for ECIs by comprehensively considering their distortion characteristics.

III. Dataset for ECI Quality Assessment

A. Image Collection and Preparation

In the field of IQA, an excellent dataset should to the greatest extent cover real-world scenes within a limited number of images. Since this study aims to investigate the quality assessment issue of ECIs, we especially pay more attention to low-light images in various scenes. With such a principle, we initially captured thousands of low-light images using Olympus 290 endoscope at the Shenzhen University General Hospital in 2020. To meet the clinical practice, we captured images from various stages of the entire colonoscopy process. The images are with the size of 1920×1080, and the patient's information is shown on the left. For privacy protection, we removed patient's information (e.g., name, gender, age, treatment time, etc.). In addition, to remove redundant black background in the original endoscopic images, we uniformly cropped them into 881×881 pixels. Then, in order to meet the clinical practice, we carefully selected 300 representative low-light images with various contents. Some examples are presented in Fig. 1(a) . As seen, these images hold sufficient diversity of contents, such as smooth/creased intestine, polyps, surgical bleeding/instrument, floating faeces, global/local tissue region, etc. As suggested by [44], we also compute spatial information (SI) and colorfulness (CF) indices to describe content variation of raw images. It is clear from Fig. 2 that, there are obvious differences in the content distribution across raw images.

With these raw images, our next step is to process them by LIE algorithms to generate ECIs. In this study, we select 8 representative LIE algorithms, including 4 popular traditional algorithms (namely, BPDHE [45], LIME [46], MF [47], and BIMEF [48]) and 4 deep learning-based algorithms (namely, Retinex [49], DSLR [50], Zero-DCE [51], and DeepLPF [52]). For fair implementations, we first use the official source code to obtain the model based on the experimental settings of each paper or directly download the pre-trained model released by the authors of each LIE algorithm. Then, the raw colonoscopy images are directly fed into each model to generate ECIs. As a result, we totally have 2,400 (= 300×8) ECIs from 300 raw images. Some examples are shown in Fig. 1(b), where we have two findings: 1) For each LIE algorithm, its performance in improving visibility varies with the image content change. In serious cases, the generated ECIs look unnatural. 2) ECIs generated by different LIE algorithms have diverse degrees of artifacts in brightness, contrast, colorfulness, noise, etc., which brings intractable challenges to quality assessment. Therefore, a comprehensive subjective study on all of the generated ECIs is crucial to help us better understand the performance of different LIE algorithms.

B. Subjective Tests

In this study, 18 observers (11 males and 7 females with normal or corrected-to-normal vision, aged from 21 to 26 years old) majored in biomedical engineering participated in subjective experiment. Before the experiment, all of these observers should basically master the key points of image screening in colonoscopy through careful training by two senior gastroenterologists. During the experiment, they were asked to sit in front of the screen. The experiment was conducted in a laboratory with normal indoor lighting akin to doctor's office environment. To coincide with clinical practice, we set a flexible viewing distance, i.e., one to three times of the image height. Moreover, the observers can keep a comfortable posture within the specified viewing zone.

The subjective evaluation consists of a training stage and a testing stage. In the training stage, the administrator first intro-

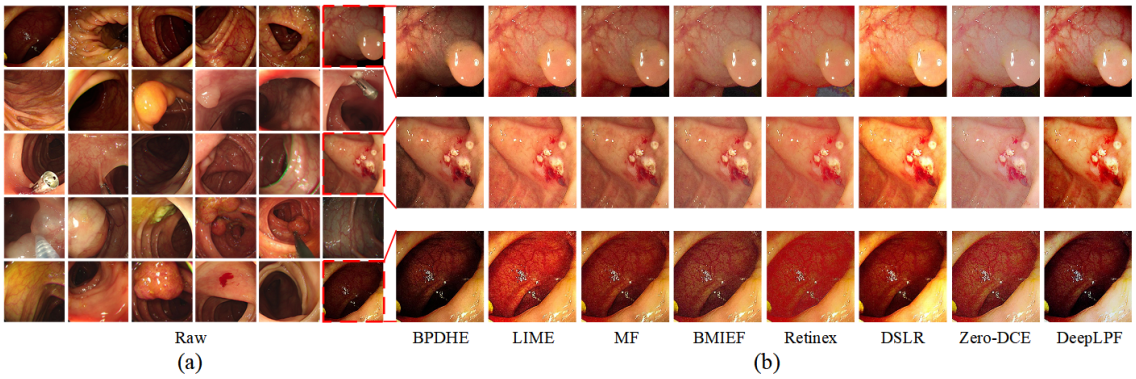


Fig. 1. Raw colonoscopy images and their enhanced results by different LIE algorithms.

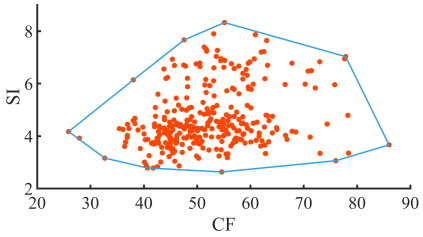


Fig. 2. Spatial information versus colorfulness plot.

duced experiment instructions and shows the entire experiment procedure. Then, 80 samples were provided to help observers familiarize the experiment on a scoring software. The testing stage could not be started until observers fully understand the experiment and get considerable rating accuracy on the prepared training samples. In the testing stage, observers should grade the quality of one randomly displayed ECI by a five-grade rating scale, as shown in Table I. We set the rating scale within [1, 9] to give more flexibility to uncertain cases. For example, one can give a score of 8 if he/she hesitates between 7 and 9. Here, the scoring rule not only considers the visual experience, but also pays attention to information integrity (i.e., the proportion of abnormal brightness), which reflects the image's utility. The rating score was automatically recorded by the scoring software. Once the scoring task of one image was completely, next ECI will be displayed on the screen immediately. The subjective test included 3 sessions conducted on three different days, and each session evaluates 800 ECIs. The observers were encouraged to stop to release the accumulated visual fatigue every 10 minutes during the test. All ECIs were randomly presented in their original resolution with no duplication on a 27-inch 1920×1080 Philips screen.

TABLE I
SCORING RULES FOR QUALITY ASSESSMENT OF ECIS.

Ratings	Descriptions
1	Severely annoying, informationless, lack of utility
3	Annoying, less information, limited utility
5	Slightly annoying, little impaired information and utility
7	Distortion perceptible, but not annoying, good utility
9	Distortion imperceptible, full of information, perfect utility

C. Subjective Data Processing

Generally, there always exists discrepancy among the subjective scoring data reported by different observers due to their own understanding about the task. Therefore, we should clean the subjective scoring data before confirming the final quality score of each ECI. Here we use the outlier rejection method recommended by ITU-R BT.500 [53] to remove abnormal data removed. As a result, two observers' data are removed. After that, the MOS value (i.e., M_p) of the p-th image can be calculated by

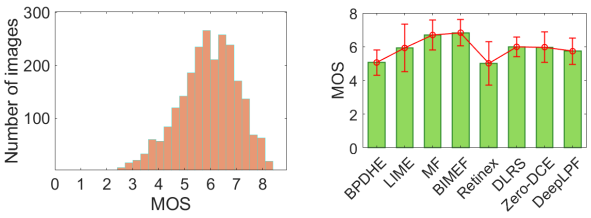
$$M_p = \frac{1}{Q}\sum_{q=1}^{Q} r_{p,q}, \tag{1}$$

where Q is the number of valid ratings after outlier rejection, $r_{p,q}$ denotes the rating score of the p-th image obtained from the q-th observer. Fig. 3(a) illustrates the distribution of MOS values of all ECIs in the proposed ECIQAD[1]. It is clear that our ECIQAD spans a wide range from severely annoying to distortion imperceptible. One phenomenon worth our attention is that there are very few images with the highest scores. A possible reason is that, apart from the low-light appearance caused by dark acquisition environment, the sensor noise is another distinctive element that affects the perceptual quality. Perfect ECIs are difficult to obtain because LIE algorithms can easily amplify the influence of noise when increasing the image contrast. Fig. 3(b) shows the MOSs of ECIs generated by each LIE algorithm, from which we can compare the performance of these algorithms in terms of mean and standard deviation (shown in the form of vertical error bars). As seen, BIMEF performs the best among 8 competing algorithms.

IV. PROPOSED ECIQ METHOD

As observed in Section III, artifacts from brightness, contrast, color, naturalness, and noise are the main factors affecting the perceptual quality of ECIs. Therefore, we first extract five kinds of features from training samples to characterize the distortions in the above five factors. Then, the extracted features are fused via the support vector regression (SVR) module to generate the IQA model. For a test ECI, its quality

[1]The dataset can only be used strictly for the non-profit academic purpose and will be available at https://github.com/Nano-cd/ECIQAD.



(a) (b)

Fig. 3. MOS distributions of ECIs in the proposed ECIQAD: (a) Overall distribution, (b) MOSs of ECIs generated by each LIE algorithm.

can be directly estimated by feeding its features into the well-trained IQA model. Fig. 4 shows the overall framework of our proposed ECIQ method, in which feature extraction is the key point. In what follows, we will detail the feature extraction process, followed by the IQA model generation.

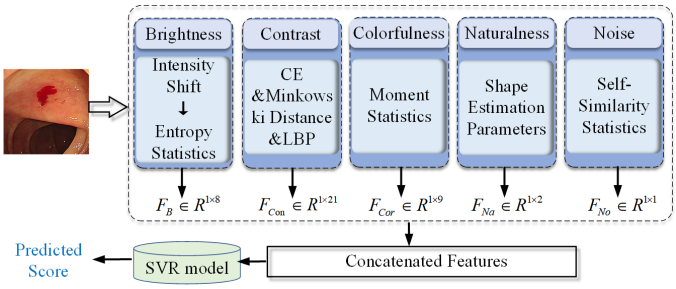


Fig. 4. Framework of the proposed ECIQ method. For a test image, we first extract features from five aspects of considerations, i.e., brightness, contrast, colorfulness, naturalness, and noise. Then, these features are concatenated to generate a 41-dimensional feature vector. Finally, the quality of the test image can be obtained by feeding its 41-dimensional feature vector into the pre-trained SVR model.

A. Brightness

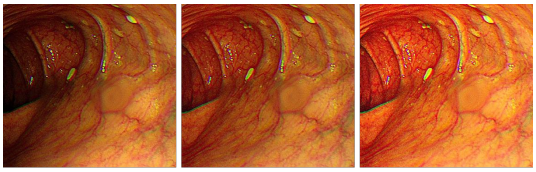
In colonoscopy images, there are usually dark regions from the far-off intestinal wall captured in low-light environment. One concern of existing LIE algorithms is to adjust the image brightness to help images display more details for improving visibility. Fig. 5 presents three representative ECIs with the same content, where the brightness is visually increasing from left to right. Fig. 5(a) is the result that only slightly enhances the dark regions to maintain high visibility for the remaining regions, while Fig. 5(c) shows the result that excessively enhances the whole regions. Both cases inevitably reduce the rendering effect of image details and accordingly result in low perceptual quality. In contrast, as illustrated in Fig. 5(b), the ECI with high quality holds more appropriate brightness, regardless of in dark regions or bright regions, for clear colon surface representation compared to other two ECIs. Motivated by these observations, we here measure the brightness from the perspective of detail computation after shifting the image intensity towards the minimum/maximum value with the observation that intensity shift will amplify the brightness difference between ECIs and bring noticeable detail change. Specifically, following [54], we first create a group of derived gray images G_i $(i = 1, 2, \cdots, 8)$:

$$G_i = G_o \cdot M_i, \tag{2}$$

where G_o is the ECI's gray image obtained by processing image I with MATLAB function "rgb2gray", M_i denotes the i-th multiplier. In this study, we empirically set $M = \{1/8, 1/6, 1/4, 1/2, 2, 4, 6, 8\}$ to balance efficiency and efficacy. With Eq. (2), the brightness of G_o will decrease or increase, and some values may exceed 255. Therefore, we directly assign those values exceeding 255 to 255 so as to restrain the image into the standard range of [0, 255]. Next, we compute the information entropy E_{M_i} of each derived images to characterize its brightness preservation ability as information entropy is widely adopted to measure image details in the IQA field:

$$E_{M_i} = -\sum_{j=1}^{255} p_{j,M_i} \cdot \log(p_{j,M_i}), \tag{3}$$

where p_{j,M_i} indicates the occurrence probability of pixels with value j in the i-th derived image. For later convenience, we denote these eight brightness-related features as $F_B = \{E_{M_1}, E_{M_2}, \cdots, E_{M_8}\}$.



(a) (b) (c)

Fig. 5. Illustration of representative ECIs: (a)-(c) are ECIs under different brightness. Zoom in for a better view.

B. Contrast

Contrast is an essential factor deciding the effect of image enhancement. Given that observers consider both global and local distortions when scoring image quality [55], we propose to measure the image contrast from both global and local aspects. Specifically, we adopt the Minkowski distance formulation [28] to characterize global contrast of ECIs with the observation that ECIs with the same content but different contrast are generally with diverse intensity distributions. Formally, the global contrast C_g can be computed by

$$C_g = \left(\frac{1}{K} \sum_{k=1}^{K} |I_k^p - \mathrm{Avg}(I^p)|^q \right)^{1/q}, \tag{4}$$

where K denotes the number of pixels in an ECI, $\mathrm{Avg}(\cdot)$ is the mean operation. I_k means the k-th pixel in the image I, and I^p stands for the image to the power of p. The inclusion of parameter p is inspired by the fact that the gamma transfer function is widely used for contrast enhancement. Different assignments of p will likely emphasize the severity of contrast distortion of I in different degrees. Similar to the assignments of M_i, we set p to $\{1/8, 1/6, 1/4, 1/2, 2, 4, 6, 8\}$. q controls the formulation of computing the data's deviation degree away from the center. Here, we set it to 4. Through Eq. (4), we

have eight global contrast-related features, simply marked as $F_{C_g} = \{C_{g,1}, C_{g,2}, \cdots, C_{g,8}\}$.

For local contrast measurement, we apply the contrast energy because it has particularly contrast-aware attributes by simulating the characteristics of human visual system (HVS) [56], [57]. Specifically, the contrast energy is computed by

$$CE_f = \frac{\alpha \cdot Z(I_f)}{Z(I_f) + \alpha \cdot \kappa} - \phi_f, \tag{5}$$

where $f \in \{gr, yb, rg\}$ denotes the individual color channels of the enhanced image I, where gr=0.299R+0.587G+0.114B, yb=0.5(R+G)$-B$, and rg=$R-G$. $Z(I_f) = ((I_f \otimes f_v)^2 + (I_f \otimes f_h)^2)^{1/2}$ is the filter response, in which $\otimes$ stands for the convolution operation, f_v and f_h are vertical and horizontal second-order derivatives of Gaussian functions, respectively. α is applied to compute the maximum value of $Z(I_f)$, and κ is a contrast gain. ϕ_f is used to constrain the noise in channel f with a threshold. In this study, we adopt the same settings as [57] regarding parameter assignments of f_v, f_h, κ, and ϕ_f. As a result, we get three local contrast-related features, simply marked as $F_{C_{la}} = \{CE_{gr}, CE_{yb}, CE_{rg}\}$.

Generally, LIE algorithms usually adjust the intensity distribution of local regions to enhance the image details. Motivated by this, we here further measure the local contrast via local intensity statistics. For this purpose, the famous local rotation invariant uniform local binary pattern (ULBP) operator [58] is applied on the gray image G_o. The ULBP operator computes the relationship between a considered pixel v_c and its adjacent pixels in a local region, i.e.,

$$\text{ULBP}^{riu2}_{P,R}(v_c) = \begin{cases} \sum_{i=0}^{P=1} s(v_i - v_c), & \text{if } \mathcal{U}(\text{ULBP}_{P,R}) \leq 2 \\ P+1, & else \end{cases} \tag{6}$$

where

$$\begin{aligned} \mathcal{U}(\text{ULBP}_{P,R}) = &\|s(v_{P-1} - v_c) - s(v_0 - v_c)\| \\ &+ \sum_{i=1}^{P-1} \|s(v_i - v_c) - s(v_{i-1} - v_c)\|. \end{aligned} \tag{7}$$

In Eq. (6), $s(\cdot)$ is the sign function to illustrate the relationship between two pixels, i.e., $s(v_i - v_c) = 1$ if $v_i \geq v_c$, otherwise, $s(v_i - v_c) = 0$. v_i is value of the central pixel's i-th neighbor in the gray image G_o. P is the number of neighbors considered, and R is the radius of the neighbor. In this study, we set P and R to 8 and 1, respectively. According to Eqs. (6)-(7), we have $P+2$ patterns, i.e., 1 non-uniform pattern and $P+1$ uniform patterns. Fig. 6 presents the ULBP statistical results of Figs. 5(a)-5(c). As seen, some bars (e.g., 2-4, 6-9) are monotonously decreasing or increasing with the change of image contrast. This indicates the feasibility of using ULBP-related features for local contrast measurement to some extent. For simplicity, we mark these ten local contrast-related features as $F_{C_{lb}} = \{UL_0, UL_1, \cdots, UL_9\}$. By combing F_{C_g}, $F_{C_{la}}$, and $F_{C_{lb}}$, we have twenty-one contrast-related features F_{Con}.

C. Colorfulness

Clinically, colonoscopy images are usually with low-light appearance due to the dark acquisition environment. To improve visibility, LIE algorithms automatically adjust the intensity distribution, which inevitably changes image color. In serious cases, LIE algorithms may negatively induce color biases. For instance, blood vessels are excessively rendered towards bright red in Fig. 5(c). Thus, we here measure the image colorfulness to characterize the color distortion of ECIs. Since the RGB color space is not suitable for color feature extraction due to high correlation between three color channels [59], we transform an ECI from RGB to opponent-color space and extract color-related features in a statistical way. For each color channel, we calculate the mean $\mathcal{M}$, standard deviation $\mathcal{D}$, and skewness $\mathcal{S}$ as the moment statistics [39]. Specifically, these three operators can be respectively formulated as

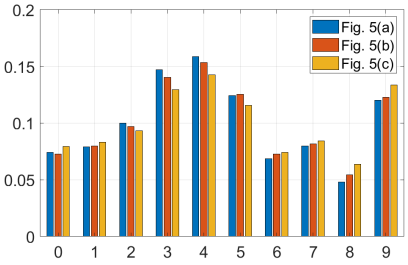


Fig. 6. Illustration of ULBP statistics of Figs. 5(a)-5(c).

$$\mathcal{M}_k = \text{Avg}(\mathcal{O}_k), k \in \{1, 2, 3\}, \tag{8}$$

$$\mathcal{D}_k = \sqrt{\text{Avg}((\mathcal{O}_k - \text{Avg}(\mathcal{O}_k))^2)}, \tag{9}$$

$$\mathcal{S}_k = \frac{\text{Avg}((\mathcal{O}_k - \text{Avg}(\mathcal{O}_k))^3)}{(\mathcal{D}_k)^3}, \tag{10}$$

where $\mathcal{O}_1 = (R-G)/\sqrt{2}$, $\mathcal{O}_2 = (R+G-2B)/\sqrt{6}$, and $\mathcal{O}_3 = (R+G+B)/\sqrt{3}$ denote the three color channels in the opponent-color space, respectively. Consequently, we can get three moment features from each channel for color distortion measurement and totally obtain nine color-related features, simply marked as $F_{Col} = \{\mathcal{M}_1, \mathcal{D}_1, \mathcal{S}_1, \mathcal{M}_2, \mathcal{D}_2, \mathcal{S}_2, \mathcal{M}_3, \mathcal{D}_3, \mathcal{S}_3\}$.

D. Naturalness

Many LIE algorithms aim to adjust image brightness or contrast of colonoscopy images for improving visibility. However, under- or over-emphasis on these aspects may result in unnatural looks, as shown in Figs. 5(a)-5(c). Thus, appropriate measurement of image naturalness may help improve the perceptual quality prediction accuracy of ECIs. In the literature, image naturalness, usually concluded from Mean Subtracted and Contrast Normalized (MSCN) coefficients, has been successfully applied in IQA tasks of NSIs [23] with the observation that distortions can corrupt the distribution of MSCN coefficients regularly with the change of distortion type and level. However, to the best of our knowledge, such an observation has not yet been investigated in the perceptual quality assessment of ECIs. Specifically, for the gray image of an ECI, its MSCN coefficients $\widehat{G}_o$ can be calculated by

$$\widehat{G}_o(i,j) = \frac{G_o(i,j) - \mu(G_o(i,j))}{\sigma(i,j) + C}, \tag{11}$$

where C is a small positive constant to avoid zero denominator. $\mu(G_o(i,j)) = \sum_{p=-\mathcal{P}}^{\mathcal{P}} \sum_{q=-\mathcal{Q}}^{\mathcal{Q}} \omega_{p,q} G_o(i,j)$ denotes the

local mean, and $\sigma(i,j) = (\sum_{p=-\mathcal{P}}^{\mathcal{P}} \sum_{q=-\mathcal{Q}}^{\mathcal{Q}} \omega_{p,q}(G_o(i,j) - \mu(G_o(i,j))^2)^{1/2}$ represents the standard deviation of the pixel in the (i,j) position. $\omega = \{\omega_{p,q} | p = -\mathcal{P}, \cdots, \mathcal{P}; q = -\mathcal{Q}, \cdots, \mathcal{Q}\}$ is the 2D circularly symmetric Gaussian weighted function. Here, we directly adopt the same settings as [23] regrading the Gaussian weighted function as this is not the main concern of our study. In Fig. 7, we illustrate the statistical distributions of MSCN coefficients for Figs. 5(a)-5(c), where we can observe that MSCN coefficients of ECIs exhibit Gaussian-like distributions. More concretely, the distribution shape changes with the appearance naturalness. This indicates the feasibility of modeling statistical distributions of MSCN coefficients as the naturalness index for the perceptual quality assessment of ECIs.

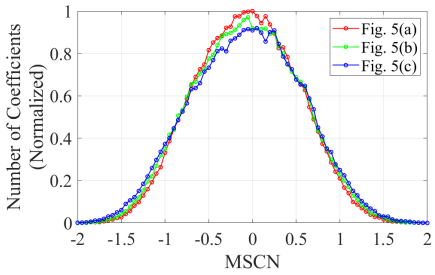


Fig. 7. Distributions of MSCN coefficients for Figs. 5(a)-5(c).

To quantify the MSCN coefficient distribution, we utilize the popular zero-mean generalized Gaussian distribution (GGD) model because it is suitable to capture a broader spectrum of distorted image statistics. Formally, the GGD model with zero mean is defined as:

$$f(x, \alpha, \sigma^2) = \frac{\tau}{2\beta\Gamma(1/\tau)} \exp\left(- \left(\frac{|x|}{\eta} \right)^{\tau} \right), \qquad (12)$$

where $\eta = \sigma\sqrt{\Gamma(1/\tau)/\Gamma(3/\tau)}$, and $\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt$ when $a > 0$. Parameters τ and σ^2 control the shape and variance of the distribution, respectively. We therefore collect them as naturalness-related features, marked as $F_{Na} = \{\tau, \sigma^2\}$.

E. Noise

The colonoscopy images acquired in a dark environment likely to suffer from noise distortion, leading to the degradation of perceptual quality. Therefore, it is necessary to evaluate the quality of ECIs from noise estimation. These years have reported many representative works to estimate the image noise level [60]. Most of them usually require specific priori hypotheses, e.g., marginal distribution, limiting their performance in real-world noise estimation. Since LIE algorithms may easily amplify the real-world noise caused by low-quality acquisition sensors, we in this study discard these classic methods of noise estimation and propose a simple but effective one. The key motivation behind our method lies in that the noise level can be directly estimated by comparing the noisy ECI with its high-quality counterpart. To be specific, let m denote an ECI with noise and n represent its clean version. The noise level F_{No} of p can be computed by

$$F_{No} = \Phi(m, n), \qquad (13)$$

where $\Phi(\cdot,\cdot)$ represents the similarity computation function or the full-reference IQA method, such as MAD [55] and SSIM [61]. In this study, we adopt the famous SSIM to achieve this. Now, one thing raising our attention is how to find the clean image of an ECI since we usually do not capture low- and high-quality image pairs in colonoscopy. To address such a problem, we use the Gaussian low-pass filter to process the ECI to generate its clean version.

F. IQA Model Generation

Overall, for an ECI, we have totally extracted 41 distortion-related features based on five aspects of considerations, i.e., brightness, contrast, colorfulness, naturalness, and noise. For readers' convenience, we summarize all the extracted features in Table II. These features can be concatenated into a 41-dimensional feature vector.The next concern is how to yield the quality score from this feature vector. Following previous IQA works [23], [37], we employ the SVR with the radial basis function kernel from the LIBSVM package [62] to learn the regression mapping from the feature space to the quality score due to its high simplicity and capability in processing high-dimensional data. Based on training samples, the parameters in SVR can be well-determined, thereby generating the regression model for quality assessment of ECIs. Please see [62] for more details about the parameter optimization. Once the regression model obtained, the quality of a query ECI can be predicted by feeding its 41 dimensional feature vector into the model. The preparation of training and testing samples will be introduced in experiment setup in Section V.

V. Experimental Results and Discussions

A. Experiment Setup

1) Train-test Splits: In the field of IQA, one dataset is usually randomly divided into two subsets to build and evaluate the IQA model. The train-test split process should be repeated many times, e.g., 100, 200, or 1000, to mitigate the performance biased caused by dataset division. In this study, we adopt the popular strategy of 80%-20% train-test split, i.e., 80% samples for training and the remaining 20% samples for testing, and repeat such a split 1000 times. The performance of an IQA method is reported in the form of the median value of each evaluation criterion across 1000 tests on the testing samples. Additionally, we also investigate the stability of our proposed method under different training percentages, from 20% to 90% in the step of 10%. To facilitate the subsequent comparisons between IQA methods, we adjusted the ECIs' spatial resolution to 375×375 in this study. Such an operation has little influence on the performance of each method.

2) Evaluation Criteria: We select four widely-used evaluation criteria, i.e., Spearman rank correlation coefficient (SRCC), Kendall's rank correlation coefficient (KRCC), Pearson linear correlation coefficient (PLCC), and Root mean-squared error (RMSE), to quantitatively evaluate and compare our proposed method against competing methods. Among

TABLE II
SUMMARY OF THE EXTRACTED FEATURES.

Feature Type	Feature Symbol	Feature Description	Feature ID	Computation
Brightness	$E_{M_1}\sim E_{M_8}$	Information entropy of intensity changing	$f_{01}\sim f_{08}$	(3)
Contrast	$C_{g,1}\sim C_{g,8}$	Minkowski distance formulation	$f_{09}\sim f_{16}$	(4)
	CE_{gr}, CE_{yb}, CE_{rg}	Contrast energy	$f_{17}\sim f_{19}$	(5)
	$UL_0\sim UL_9$	Local binary pattern	$f_{20}\sim f_{29}$	(6)$\sim$(7)
Colorfulness	$\mathcal{M}_1\sim\mathcal{M}_3$, $\mathcal{D}_1\sim\mathcal{D}_3$, $\mathcal{S}_1\sim\mathcal{S}_3$	Moment statistics from each color channel	$f_{30}\sim f_{38}$	(8)$\sim$(10)
Naturalness	τ, σ^2	Nature scene statistics	$f_{39}\sim f_{40}$	(11)$\sim$(12)
Noise	F_{Col}	Self-similarity	f_{41}	(13)

these evaluation criteria, SRCC and KRCC evaluate the prediction monotonicity. Formally, SRCC can be calculated by

$$\mathrm{SRCC} = 1 - \frac{6\sum_{i=1}^{\mathcal{N}} d_i^2}{\mathcal{N}(\mathcal{N}^2-1)}, \tag{14}$$

where d_i is the difference between the i-th image's ranks in subjective and objective evaluations, $\mathcal{N}$ is the number of images in the test set. KRCC is defined as:

$$\mathrm{KRCC} = \frac{2(\mathcal{N}_c - \mathcal{N}_d)}{\mathcal{N}(\mathcal{N}-1)}, \tag{15}$$

where $\mathcal{N}_c$ and $\mathcal{N}_d$ stand for the number of concordant and discordant pairs, respectively. PLCC and RMSE measure the prediction accuracy. Let x_i and s_i denote the predicted score and subjective score of the i-th image, PLCC is calculated by:

$$\mathrm{PLCC} = \frac{\sum_{i=1}^{\mathcal{N}}(x_i-\overline{x})\cdot(s_i-\overline{s})}{\sqrt{\sum_{i=1}^{\mathcal{N}}(x_i-\overline{x})^2}\cdot\sqrt{\sum_{i=1}^{\mathcal{N}}(s_i-\overline{s})^2}}, \tag{16}$$

where $\overline{x}$ and $\overline{s}$ denote the mean score of the predicted scores and subjective scores over the test set. RMSE is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}}(x_i-s_i)^2}. \tag{17}$$

Before computing PLCC and RMSE, a five-parameter nonlinear fitting function suggested by video quality experts group [63] is utilized to map the predicted scores to the subjective scale:

$$f(x) = \kappa_1\left[\frac{1}{2} - \frac{1}{e^{\kappa_2(x-\kappa_3)}+1}\right] + \kappa_4 x + \kappa_5, \tag{18}$$

where $\kappa_1 \sim \kappa_5$ are fitting parameters. They can be estimated by minimizing the sum of squared errors between the mapped objective score $f(x)$ and the corresponding subjective score. Generally, a superior objective IQA method is expected to show higher correlation between its predictions and subjective scores, behaving in the form of a value closer to 1 for SRCC, KRCC, and PLCC but closer to 0 for RMSE.

3) Compared Methods: Since very few IQA methods are specially designed for ECIs, we compare our ECIQ method with 14 well-known NR IQA methods designed for NSIs. These methods can be classified into three categories according to their applications. The first category includes five contrast-specific methods, namely, NR-CDIQA [39], NIQMC [40], MDM [28], BIQME [41], and NUIQ [29]. The second category contains seven general-purpose methods designed for quality assessment of synthesized distortions, including BRISQUE [23], GM-LOG [64], IL-NIQE [33], BPRI [34], BMPRI [65], NPQI [66], and CIQA [67]. The last category is composed of two methods proposed for evaluating authentically distorted images, including GWH-GLBP [68] and FRIQUEE [32]. During the model building and evaluation, all opinion-aware learning-based methods (except IL-NIQE, NIQMC, BPRI, and NPQI) adopt the 80%-20% train-test split with 1000 repetitions. As for opinion-unaware BIQA methods (IL-NIQE, NIQMC, BPRI, and NPQI), we directly test them on the same testing set as that used for opinion-aware methods. Note that, during the implementation of NUIQ, we replace its original SVMrank function by SVR function to match the regression task here. Considering that there exists domain shift between NSIs and ECIs, we discard the pre-train procedure of BIQME on the basis of NSIs with pseudo labels. Instead, we directly learn its IQA model on training samples of ECIQAD like our method.

B. Overall Performance Comparison and Analysis

Table III presents the experimental results of our method and 14 competing IQA methods on the constructed dataset. Through the quantitative comparison, we have below observations. First of all, all of the 14 competing methods obtain unsatisfactory results with the highest SRCC value of 0.7808, the highest KRCC value of 0.5970, and the highest PLCC value of 0.7881. Such results cannot meet the clinical needs since we usually acquire thousands of ECIs in colonoscopy if using the endoscope equipped with an LIE algorithm. It is worth noting that low predicting performance means that the objective predictions cannot correlate well with the subjective feelings and a large number of ECIs will be classified into the wrong quality level. Secondly, these opinion-unaware methods (i.e., IL-NIQE, NIQMC, BPRI, and NPQI) are generally inferior to these opinion-aware methods (i.e., NR-CDIQA, MDM, BIQME, NUIQ, BRISQUE, GM-LOG, BMPRI, CIQA, GWH-GLBP, and FRIQUEE) by a large margin. For instance, the average SRCC value of these four opinion-unaware methods is approximately 0.3796 lower than that of ten opinion-aware methods. Thirdly, learning-based methods in the first category achieve superior performance over those belonging to the second and third categories. The average SRCC values of learning-based methods in three categories are 0.6005, 0.5321, and 0.5557, respectively. Other evaluation criteria, e.g., PLCC and KRCC, have similar dynamics. Last but not the least, our proposed ECIQ method achieves superior performance over

14 competing methods. Much to our excitement, it is ahead of the runner up NUIQ with the excess values of 0.0577 and 0.0538 in terms of SRCC and PLCC, respectively.

TABLE III
PERFORMANCE COMPARISONS OF COMPETING METHODS ON THE CONSTRUCTED ECIQAD. ↑ (↓) MEANS THAT HIGHER (LOWER) VALUES INDICATE BETTER PERFORMANCE. THE BEST RESULT OF EACH EVALUATION CRITERION IS HIGHLIGHTED IN BOLD.

Methods	Evaluation Criteria				Time
	SRCC↑	KRCC↑	PLCC↑	RMSE↓	(Sec.)
NR-CDIQA [39]	0.4617	0.3252	0.4492	1.0156	0.0130
NIQMC [40]	0.2459	0.1656	0.2275	1.1070	0.9903
MDM [28]	0.3901	0.2692	0.4011	1.0424	**0.0036**
BIQME [41]	0.7695	0.5792	0.7681	0.7283	0.1659
NUIQ [29]	0.7808	0.5970	0.7881	0.7000	0.0986
BRISQUE [23]	0.4363	0.3034	0.4588	1.0111	0.0236
GM-LOG [64]	0.5235	0.3695	0.5368	0.9606	0.0178
IL-NIQE [33]	0.1184	0.0798	0.1259	1.1279	4.3640
BPRI [34]	0.1523	0.1023	0.1808	1.1182	0.5779
BMPRI [65]	0.4300	0.2948	0.4380	1.0237	0.2416
NPQI [66]	0.2216	0.1490	0.2385	1.1035	0.4205
CIQA [67]	0.7384	0.5493	0.7347	0.7724	0.1512
GWH-GLBP [68]	0.4486	0.3095	0.4461	1.0181	0.0258
FRIQUEE [32]	0.6627	0.4768	0.6562	0.8593	8.1482
ECIQ (Ours)	**0.8385**	**0.6541**	**0.8419**	**0.6132**	0.0682

Possible reasons for the above phenomena are as follows. First, as shown in Fig. 1, ECIs have complex distortions and show strong distinctive characteristics in different aspects, such as brightness, colorfulness, noise, etc. Most general-purpose IQA methods are mainly designed for quantifying common synthesized distortions (e.g., Gaussian blur, JPEG compression, etc.) that usually appear in NSIs. They have limited potentials in characterizing the complex distortions in ECIs. Second, opinion-unaware methods generally do not require any subjective scores for IQA model construction. Their design concepts, e.g., the determination of distortion types in BPRI, and the selection of high-quality pristine images in IL-NIQE, are highly related to designers' experience. As a result, the performance is easily decreased once the application changes, especially from NSIs to ECIs. In contrast, opinion-aware methods prefer to utilize machine learning tools to build the mapping relationship between features and subjective scores via supervised learning, thereby holding stronger adaptability than opinion-unaware methods towards application change. Third, as for learning-based methods in the first category, they are specifically designed for contrast-distorted NSIs by considering the characteristics of contrast distortion. Thus, it is acceptable that they are capable of obtaining better performance than methods in other two categories. Specifically, NUIQ, which considers the characteristics of distortions induced by classical LIE algorithms, instead of by simple mean shift operation, achieves the best among methods in the first category, indicating the correctness of evaluating contrast-changed images from multiple aspects. Finally, since the distortion of ECIs is complex, it is insufficient to quantify the distortion from only two or three aspects. In contrast, our proposed ECIQ method elaborately extracts many features from five aspects, i.e., brightness, contrast, colorfulness, naturalness, and noise, and can consequently better characterize the distortions of ECIs for superior performance over competing methods.

C. Individual Performance Comparison and Analysis

Since eight LIE algorithms participate in the material generation of ECIQAD, it is also interesting and meaningful to see the performance of our ECIQ method in evaluating ECIs generated by each individual LIE algorithm. For this purpose, we further conduct a series of complementary experiments in this section.

Table IV lists the results in terms of SRCC and PLCC, where three conclusions can be drawn. Here, we do not show KRCC and RMSE any more due to space limitation. First, for any of the 15 IQA methods, it cannot achieve the identical performance on ECIs generated by different LIE algorithms. Taking the SRCC index reported by NUIQ as an example, the difference between the maximum and minimum values is 0.3153 across eight LIE algorithms. This indicates that the selected LIE algorithms perform differently in enhancing colonoscopy images. One possible explanation is that these algorithms adjust the appearance of colonoscopy images from various perspectives, making the perceptual distortion artifacts of the generated ECIs different. Second, in the case of each LIE algorithm, these IQA methods have different performances. Specifically, our proposed ECIQ method outperforms other methods in the cases of LIME, MF, BIMEF, Retinex, DSLR, Zero-DCE, and DeepLPF, while is slightly inferior to NUIQ in the case of BPDHE. This demonstrates the high stability of our method across different cases and indicates its wider potential in real-world ECI comparison over competing methods. Third, almost all of the selected mainstream IQA methods cannot well handle the ECI quality assessment task from some LIE algorithms (e.g., BPDHE and DSLR), which throws out urgent demands of designing more advanced IQA methods in future works.

D. Running Time Comparison

In this section, we compare our proposed method with competing methods from the aspect of running time. The source code of each method is implemented on MATLAB R2019a on a high-performance PC, which is powered by a Windows 10 Pro 64-bit operation system and equipped with an i7-8700 CPU @3.20GHz. The *tic* and *toc* functions are used to record the total running time of each method consumed on processing 2,400 ECIs. We present the average running time (seconds per image) of each method in the last column of Table III. It is obvious that the proposed method has a moderate speed. Specifically, it requires 0.0682s in average to process an ECI with the size of 375×375 and is slightly slower than MDM, NR-CDIQA, GM-LOG, BRISQUE, and GWH-GLBP. Despite this, our method has better prediction performance than them, as demonstrated in Table III and Table IV. It is worth noting that an IQA method with higher prediction accuracy is highly desired in clinic as it is more conducive to selecting optimal ECIs, which may provide better visual experience and disease analysis for physicians.

TABLE IV
PERFORMANCE COMPARISONS ON ECIS GENERATED BY EACH LIE ALGORITHMS. ↑ (↓) MEANS THAT HIGHER (LOWER) VALUES INDICATE BETTER PERFORMANCE. THE BEST RESULT OF EACH EVALUATION CRITERION IS HIGHLIGHTED IN BOLD.

Methods	BPDHE [45]		LIME [46]		MF [47]		BIMEF [48]		Retinex [49]		DSLR [50]		Zero-DCE [51]		DeepLPF [52]	
	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑	SRCC↑	PLCC↑
NR-CDIQA [39]	0.2683	0.2757	0.3391	0.3552	0.3034	0.3319	0.4508	0.4413	0.4920	0.5015	0.0614	0.0714	0.3120	0.3332	0.6227	0.6184
NIQMC [40]	0.1817	0.2425	0.3156	0.3186	0.1946	0.2429	0.0887	0.1098	0.1449	0.1735	0.0288	0.1192	0.1039	0.1000	0.0180	0.0042
MDM [28]	0.2597	0.2848	0.5115	0.6594	0.1298	0.1846	0.0183	0.1183	0.3259	0.3700	0.3121	0.3394	0.1950	0.1469	0.6512	0.6294
BIQME [41]	0.4902	0.5062	0.7687	0.8179	0.6318	0.6130	0.6839	0.6876	0.6865	0.7065	0.4873	0.5137	0.7352	0.7341	0.7538	0.7551
NUIQ [29]	**0.5722**	**0.5725**	0.7775	0.8322	0.6746	0.6536	0.7153	0.7207	0.6786	0.7082	0.4622	0.4786	0.7207	0.7166	0.7057	0.6931
BRISQUE [23]	0.1833	0.1936	0.4313	0.3918	0.4371	0.4246	0.6143	0.6151	0.6283	0.6195	0.1773	0.1586	0.6709	0.6554	0.3949	0.3979
GM-LOG [64]	0.2448	0.2638	0.4330	0.4512	0.4332	0.4260	0.5550	0.5834	0.6509	0.6555	0.3042	0.3500	0.6114	0.6157	0.5812	0.5875
IL-NIQE [33]	0.0714	0.0858	0.1448	0.1506	0.1562	0.1637	0.2329	0.2561	0.1198	0.1183	0.1200	0.1464	0.1267	0.1625	0.1839	0.2031
BPRI [34]	0.1924	0.1993	0.0112	0.0317	0.1424	0.1182	0.2005	0.2366	0.1964	0.2320	0.0645	0.0833	0.2454	0.2799	0.1093	0.1312
BMPRI [65]	0.2889	0.3157	0.3187	0.3310	0.3764	0.3634	0.4887	0.5104	0.4798	0.4926	0.2670	0.3066	0.4579	0.4659	0.4907	0.4967
NPQI [66]	0.0079	0.0183	0.4658	0.4719	0.1040	0.1058	0.4160	0.4118	0.3966	0.4342	0.3293	0.3449	0.3532	0.3757	0.1991	0.2093
CIQA [67]	0.5624	0.5735	0.7153	0.7645	0.6558	0.6527	0.6185	0.6363	0.6205	0.6691	0.4272	0.4503	0.6701	0.6674	0.7481	0.7558
GWH-GLBP [68]	0.0816	0.1117	0.4838	0.4837	0.4102	0.4142	0.4100	0.4385	0.6020	0.6247	0.2615	0.2743	0.5555	0.5750	0.3447	0.3385
FRIQUEE [32]	0.2589	0.2910	0.7085	0.7198	0.5464	0.5434	0.5868	0.6018	0.6845	0.7243	0.3737	0.3881	0.6021	0.6016	0.4700	0.4672
ECIQ (Ours)	0.5392	0.5436	**0.8331**	**0.8698**	**0.7336**	**0.7420**	**0.7349**	**0.7312**	**0.8058**	**0.8390**	**0.5504**	**0.5730**	**0.7910**	**0.7907**	**0.7511**	**0.7538**

E. The Ablation Study

1) Effects of Training Percentage: Generally, an excellent IQA method should be stable under different ratios of training subset to testing subset. Here, we conduct more in-depth experiments to investigate the performance trend over training percentage variation. Specifically, similar to the main experiment, we randomly divide the constructed dataset ECIQAD into training and testing subsets with no content overlap 1,000 times, and respectively retrain and test the proposed ECIQ method on two subsets. Table V tabulates the experimental results in the form of median value of each evaluation criterion across 1,000 tests under different training percentages. As can be observed from the table, all of the SRCC, KRCC, and PLCC values increase while the RMSE value decreases with the increase of training percentage, which is in complete accord with the conclusions drawn by learning-based NR IQA methods in the literature. It is worth noting that, the performance of ECIQ is still comparable with that of most competing methods even when the training percentage is dropped to 30%. This demonstrates the excellent quality of our proposed ECIQ method in evaluating ECIs stably.

TABLE V
PERFORMANCE OF OUR PROPOSED ECIQ METHOD ON THE CONSTRUCTED ECIQAD UNDER DIFFERENT TRAIN-TEST SPLITS. ↑ (↓) MEANS THAT HIGHER (LOWER) VALUES INDICATE BETTER PERFORMANCE.

Train-Test Split	Evaluation Criteria			
	SRCC↑	KRCC↑	PLCC↑	RMSE↓
20% − 80%	0.7790	0.5910	0.7849	0.7066
30% − 70%	0.7997	0.6122	0.8054	0.6758
40% − 60%	0.8114	0.6249	0.8167	0.6574
50% − 50%	0.8207	0.6347	0.8259	0.6423
60% − 40%	0.8280	0.6423	0.8322	0.6317
70% − 30%	0.8335	0.6492	0.8388	0.6208
80% − 20%	0.8385	0.6541	0.8419	0.6132
90% − 10%	0.8411	0.6580	0.8476	0.6000

2) Effectiveness of Features Considered in Different Aspects: According to the descriptions in Section IV, the motivation of our proposed ECIQ method mainly lies in extracting a groups of features via five aspects of considerations, i.e., brightness, contrast, colorfulness, naturalness, and noise of ECIs. Although ECIQ has been proved effective in Section V, we are still unclear how these features in it contribute to the final performance. To answer this question, we here remove the features from each aforementioned aspect in turn and accordingly learn five new IQA models. All experiment settings are the same as the main experiment. The experimental results in the form of median value across 1000 tests are shown in Table VI, where we use $\mathcal{F}_1 \sim \mathcal{F}_5$ to denote features from the aforementioned aspects due to space limitation. The symbol "✓" ("✗") indicates that the feature in the column is included (not included). As seen, each of these features can be served as a supplement to other features towards the final performance shown in the last row. In other words, removing any of them from ECIQ will lead to performance drop, which confirms their positive role in obtaining the proposed high-performance ECIQ method. Specifically, the contrast-related features $\mathcal{F}_2$ seems to be the most important among the five aspects of features because its removal will bring the biggest performance drop compared to others and transform our ECIQ method into a mediocre one.

TABLE VI
ABLATION STUDY FOR INVESTIGATING THE CONTRIBUTIONS OF FEATURES CONSIDERED IN EACH ASPECT. ↑ (↓) MEANS THAT HIGHER (LOWER) VALUES INDICATE BETTER PERFORMANCE.

$\mathcal{F}_1$	$\mathcal{F}_2$	$\mathcal{F}_3$	$\mathcal{F}_4$	$\mathcal{F}_5$	Evaluation Criteria			
					SRCC↑	KRCC↑	PLCC↑	RMSE↓
✗	✓	✓	✓	✓	0.8349	0.6506	0.8391	0.6185
✓	✗	✓	✓	✓	0.7973	0.6087	0.7974	0.6864
✓	✓	✗	✓	✓	0.8197	0.6344	0.8213	0.6394
✓	✓	✓	✗	✓	0.8210	0.6364	0.8268	0.6381
✓	✓	✓	✓	✗	0.8378	0.6534	0.8408	0.6146
✓	✓	✓	✓	✓	**0.8385**	**0.6541**	**0.8419**	**0.6132**

3) Effects of the Parameter p in Eq. (4)*:* The parameter p is used to change the image contrast by means of the gamma transfer. Generally, the number of p affects the ability to characterize the contrast distortion and determines the number of the extracted contrast-related features. Here, we further investigate the performance of our method under three settings of p, i.e., $\{1/4, 1/2, 2, 4\}$, $\{1/6, 1/4, 1/2, 2, 4, 6\}$, and $\{1/8, 1/6, 1/4, 1/2, 2, 4, 6, 8\}$. Since the configuration of p follows the settings of M_i in Eq. (2), we here update the setting of M_i synchronously when selecting each of the above three settings. The experimental results are shown in Fig. 8, where characters on the x-axis denote the number of p. As

seen, the setting of p slightly affects the results. Through comparisons, we choose the third setting as it produces better results compared with other settings.

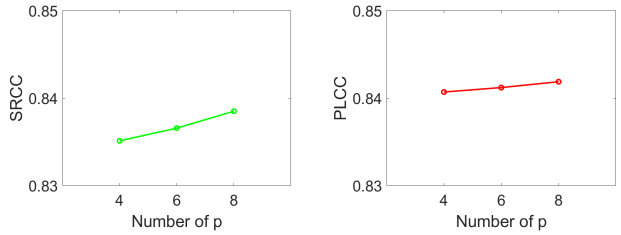


Fig. 8. Results of the proposed ECIQ method under different settings of p: (a) SRCC, (b) PLCC.

F. Limitations and Future Works

Although our method shows superiority over competing methods in the quality assessment task of ECIs, it still has the following limitations. On the one hand, the proposed ECIQ method only characterizes the distortions in ECIs from five aspects, i.e., brightness, contrast, colorfulness, naturalness, and noise, which is still insufficient. As shown in Table III and Table IV, there is still much room for performance improvement. One the other hand, the running time of the proposed ECIQ method is not satisfactory. Specifically, the proposed method spends 0.0682s on a 375×375 image and can only process approximately 15 ECIs per second, which is not suitable for the real-time system in clinic.

In order to break through the above limitations, future works can be carried out from the following directions. First, for achieving a higher prediction accuracy, one can extract more quality-aware features from ECIs by modeling the characteristics of HVS. Also, both low-level features and high-level semantic concepts can be extracted by using deep convolutional neural networks to improve the representation ability of distortions in ECIs. Second, for achieving a faster running speed, some redundant features should not be extracted. A possible future attempt for this is to analyze the importance of current extracted features in more detail, and find out which features contribute less to the performance improvement yet decrease computational efficiency.

VI. Conclusion

In this paper, we have carried out a pioneering investigation to subjectively and objectively evaluate perceptual quality assessment of ECIs. First, we construct a quality assessment dataset, named ECIQAD, by conducting rigorous subjective studies on 2,400 ECIs generated by 8 representative LIE algorithms. The constructed ECIQAD fills the gaps in this field to some extent and provides a reliable platform to fairly compare the performance of different IQA methods on ECIs. Second, we propose a new NR method, named ECIQ, to automatically evaluate the perceptual quality of ECIs. Specifically, after fully analyzing the distinctive characteristics of ECIs, we elaborately extract a groups of features in the consideration of brightness, contrast, colorfulness, naturalness, and noise, and apply the SVR tool to learn a quality prediction model. Experimental results on ECIQAD demonstrate that 14 mainstream NR IQA methods designed for NSIs are not well competent for the quality assessment task of ECIs, concluding the urgent requirements of designing specific quality assessment methods for ECIs. Our proposed ECIQ achieves a SRCC of 0.8385, a KRCC of 0.6541, a PLCC of 0.8419, and a RMSE of 0.6132. It is superior to these 14 competing methods and offers a reference for the follow-up research.

Acknowledgment

References

[1] R. L. Siegel, K. D. Miller, A. Goding Sauer, S. A. Fedewa, L. F. Butterly, J. C. Anderson, A. Cercek, R. A. Smith, and A. Jemal, "Colorectal cancer statistics, 2020," *Cancer J. Clin.*, vol. 70, no. 3, pp. 145–164, 2020.

[2] S.-L. Chen, H.-Y. Huang, and C.-H. Luo, "Time multiplexed vlsi architecture for real-time barrel distortion correction in video-endoscopic images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 11, pp. 1612–1621, 2011.

[3] T. H. Khan and K. A. Wahid, "Low power and low complexity compressor for video capsule endoscopy," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 10, pp. 1534–1546, 2011.

[4] J. Dong, Y. Cong, G. Sun, Y. Yang, X. Xu, and Z. Ding, "Weakly-supervised cross-domain adaptation for endoscopic lesions segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 5, pp. 2020–2033, 2021.

[5] G. Yue, S. Li, T. Zhou, M. Wang, J. Du, Q. Jiang, W. Gao, T. Wang, and J. Lv, "Adaptive context exploration network for polyp segmentation in colonoscopy images," *IEEE Trans. Emerg. Topics Comput. Intell.*, accepted, in press, DOI: 10.1109/TETCI.2022.3193677, 2022.

[6] G. Yue, S. Li, R. Cong, T. Zhou, B. Lei, and T. Wang, "Attention-guided pyramid context network for polyp segmentation in colonoscopy images," *IEEE Trans. Instrum. Meas.*, accepted, in press, DOI: 10.1109/TIM.2023.3244219, 2023.

[7] G. Yue, W. Han, B. Jiang, T. Zhou, R. Cong, and T. Wang, "Boundary constraint network with cross layer feature integration for polyp segmentation," *IEEE J. Biomed. Health Informat*, vol. 26, no. 8, pp. 4090–4099, 2022.

[8] J. Oh, S. Hwang, Y. Cao, W. Tavanapong, D. Liu, J. Wong, and P. C. De Groen, "Measuring objective quality of colonoscopy," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 9, pp. 2190–2196, 2008.

[9] B. Sdiri, M. Kaaniche, F. A. Cheikh, A. Beghdadi, and O. J. Elle, "Efficient enhancement of stereo endoscopic images based on joint wavelet decomposition and binocular combination," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 33–45, 2019.

[10] C. Winter, S. Rupp, M. Elter, C. Munzenmayer, H. Gerhauser, and T. Wittenberg, "Automatic adaptive enhancement for images obtained with fiberscopic endoscopes," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 10, pp. 2035–2046, 2006.

[11] C. Li, J. Guo, F. Porikli, and Y. Pang, "Lightennet: A convolutional neural network for weakly illuminated image enhancement," *Pattern Recognit. Lett.*, vol. 104, pp. 15–22, 2018.

[12] G. Fan, B. Fan, M. Gan, G. Chen, and C. P. Chen, "Multiscale low-light image enhancement network with illumination constraint," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7403–7417, 2022.

[13] J. Liang, Y. Xu, Y. Quan, B. Shi, and H. Ji, "Self-supervised low-light image enhancement using discrepant untrained network priors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7332–7345, 2022.

[14] Z. Zhao, B. Xiong, L. Wang, Q. Ou, L. Yu, and F. Kuang, "Retinexdip: A unified deep framework for low-light image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1076–1088, 2022.

[15] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image Represent.*, vol. 22, no. 4, pp. 297–312, 2011.

[16] G. Yue, C. Hou, K. Gu, N. Ling, and B. Li, "Analysis of structural characteristics for quality assessment of multiply distorted images," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2722–2732, 2018.

[17] G. Yue, C. Hou, K. Gu, T. Zhou, and G. Zhai, "Combining local and global measures for dibr-synthesized image quality evaluation," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 2075–2088, 2019.

[18] G. Yue, C. Hou, K. Gu, T. Zhou, and H. Liu, "No-reference quality evaluator of transparently encrypted images," *IEEE Trans. Multimedia*, vol. 21, no. 9, pp. 2184–2194, 2019.

[19] H. Sheikh, "Live image quality assessment database release 2," *http://live. ece. utexas. edu/research/quality*, 2005.

[20] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, 2015.

[21] M. Asif, L. Chen, H. Song, J. Yang, and A. F. Frangi, "An automatic framework for endoscopic image restoration and enhancement," *Appl. Intell.*, vol. 51, no. 4, pp. 1959–1971, 2021.

[22] B. Li and M. Q.-H. Meng, "Wireless capsule endoscopy images enhancement via adaptive contrast diffusion," *J. Vis. Commun. Image Represent.*, vol. 23, no. 1, pp. 222–228, 2012.

[23] 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, 2012.

[24] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2035–2048, 2018.

[25] 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.

[26] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, "No-reference image blur assessment based on discrete orthogonal moments," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 39–50, 2016.

[27] X. Huang, L. Chen, J. Tian, X. Zhang, and X. Fu, "Blind noisy image quality assessment using block homogeneity," *Comput. Elect. Eng.*, vol. 40, no. 3, pp. 796–807, 2014.

[28] H. Z. Nafchi and M. Cheriet, "Efficient no-reference quality assessment and classification model for contrast distorted images," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 518–523, 2018.

[29] Q. Jiang, Y. Gu, C. Li, R. Cong, and F. Shao, "Underwater image enhancement quality evaluation: Benchmark dataset and objective metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 5959–5974, 2022.

[30] L. Lévêque, W. Zhang, C. Cavaro-Ménard, P. Le Callet, and H. Liu, "Study of video quality assessment for telesurgery," *IEEE Access*, vol. 5, pp. 9990–9999, 2017.

[31] Z. A. Khan, A. Beghdadi, M. Kaaniche, and F. A. Cheikh, "Residual networks based distortion classification and ranking for laparoscopic image quality assessment," in *IEEE Int. Conf. Image Process.* IEEE, 2020, pp. 176–180.

[32] D. Ghadiyaram and A. C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *J. Vis.*, vol. 17, no. 1, pp. 32–32, 2017.

[33] 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, 2015.

[34] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, and C. W. Chen, "Blind quality assessment based on pseudo-reference image," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2049–2062, 2017.

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

[36] 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, 2015.

[37] W. Zhou, J. Xu, Q. Jiang, and Z. Chen, "No-reference quality assessment for 360-degree images by analysis of multifrequency information and local-global naturalness," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 1778–1791, 2021.

[38] K. Panetta, C. Gao, and S. Agaian, "No reference color image contrast and quality measures," *IEEE Trans. Consum. Electron.*, vol. 59, no. 3, pp. 643–651, 2013.

[39] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 838–842, 2014.

[40] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4559–4565, 2016.

[41] K. Gu, D. Tao, J.-F. 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, 2018.

[42] G. Yue, C. Hou, T. Zhou, and X. Zhang, "Effective and efficient blind quality evaluator for contrast distorted images," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2733–2741, 2019.

[43] Y. Zhou, L. Li, H. Zhu, H. Liu, S. Wang, and Y. Zhao, "No-reference quality assessment for contrast-distorted images based on multifaceted statistical representation of structure," *J. Vis. Commun. Image Represent.*, vol. 60, pp. 158–169, 2019.

[44] S. Winkler, "Analysis of public image and video databases for quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 616–625, 2012.

[45] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1752–1758, 2007.

[46] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, 2017.

[47] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Process.*, vol. 129, pp. 82–96, 2016.

[48] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," *arXiv preprint arXiv:1711.00591*, 2017.

[49] W. Y. J. L. Chen Wei, Wenjing Wang, "Deep retinex decomposition for low-light enhancement," in *Proc. British Mach. Vis. Conf.*, 2018.

[50] S. Lim and W. Kim, "Dslr: deep stacked laplacian restorer for low-light image enhancement," *IEEE Trans. Multimedia*, vol. 23, pp. 4272–4284, 2020.

[51] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1780–1789.

[52] S. Moran, P. Marza, S. McDonagh, S. Parisot, and G. Slabaugh, "Deeplpf: Deep local parametric filters for image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 12 826–12 835.

[53] "Methodology for the subjective assessment of the quality of television pictures," *document Recommendation BT.500-13, ITU-R*. [Online]. Available: https://www.itu.int/rec/R-REC-BT.500-13-201201-I/en

[54] K. Gu, S. Wang, G. Zhai, S. Ma, X. Yang, W. Lin, W. Zhang, and W. Gao, "Blind quality assessment of tone-mapped images via analysis of information, naturalness, and structure," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 432–443, 2016.

[55] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electr. Imag.*, vol. 19, no. 1, p. 011006, 2010.

[56] H. S. Scholte, S. Ghebreab, L. Waldorp, A. W. Smeulders, and V. A. Lamme, "Brain responses strongly correlate with weibull image statistics when processing natural images," *J. Vis.*, vol. 9, no. 4, pp. 29–29, 2009.

[57] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3888–3901, 2015.

[58] 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, 2002.

[59] J. Van de Weijer, T. Gevers, and A. D. Bagdanov, "Boosting color saliency in image feature detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 150–156, 2005.

[60] L. Dong, J. Zhou, and Y. Y. Tang, "Noise level estimation for natural images based on scale-invariant kurtosis and piecewise stationarity," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 1017–1030, 2016.

[61] 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, 2004.

[62] 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, 2011.

[63] V. Q. E. Group *et al.*, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II," in *2003, VQEG meeting*, 2003.

[64] 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, 2014.

[65] 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, 2018.

[66] Y. Liu, K. Gu, X. Li, and Y. Zhang, "Blind image quality assessment by natural scene statistics and perceptual characteristics," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 16, no. 3, pp. 1–91, 2020.
[67] H. Chen, X. Chai, F. Shao, X. Wang, Q. Jiang, M. Chao, and Y.-S. Ho, "Perceptual quality assessment of cartoon images," *IEEE Trans. Multimedia*, accepted, in press, DOI: 10.1109/TMM.2021.3121875, 2021.
[68] 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, 2016.

Guanghui Yue received the B.S. degree in communication engineering from Tianjin University in 2014, and the Ph.D. degree in information and communication engineering from Tianjin University, Tianjin, China, in 2019. He was a joint Ph.D. student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, from September 2017 to January 2019.

He is currently an Assistant Professor with the School of Biomedical Engineering, Health Science Center, Shenzhen University. His research interests include bioelectrical signal processing, multimedia quality assessment, 3D image visual discomfort prediction, pattern recognition, and medical image analysis.

Di Cheng is currently pursuing the master¡¯s degree from the School of Biomedical Engineering, Shengzhen University, Shengzhen, China.

His research interests include image quality assessment, image restoration, and deep learning.

Tianwei Zhou received the B.S. degree in automation from Tianjin University in 2014 and the Ph.D. degree in control science and engineering from Tianjin University, Tianjin, China, in 2019. She was a joint Ph.D. student with the Department of Electrical & Computer Engineering, National University of Singapore from August 2017 to August 2018.

She is currently an Assistant Professor with the College of Management, Shenzhen University. Her current research interests include image processing, synchronization, networked control systems, and event-triggered control.

Jingwen Hou received the bachelor's degree in E-commerce from Beijing University of Posts and Telecommunications, Beijing, China, and the master's degree in E-business from Carnegie Mellon University, Pittsburgh, U.S.

He is currently pursuing the Ph.D. degree with School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include multimedia quality assessment and computer vision.

Weide Liu obtained his B.S. degree and subsequently his Ph.D. degree from Nanyang Technological University, Singapore. He is presently a research scientist at the Institute for Infocomm Research, Agency for Science, Technology and Research in Singapore.

His research interests encompass computer vision, natural language processing, and machine learning.

Long Xu received his M.S. degree in applied mathematics from Xidian University, Xi'an, China, in 2002, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He was a Postdoc with the Department of Computer Science, City University of Hong Kong, the Department of Electronic Engineering, Chinese University of Hong Kong, from July Aug. 2009 to Dec. 2012. From Jan. 2013 to March 2014, he was a Postdoc with the School of Computer Engineering, Nanyang Technological University, Singapore. Currently,

He is with the State Key Laboratory of Space Weather, National Space Science Center, Chinese Academy of Sciences. His research interests include image/video processing, wavelet, machine learning, and computer vision. He was selected into the 100-Talents Plan, Chinese Academy of Sciences, 2014.

Tianfu Wang received the Ph.D. degree in biomedical engineering from Sichuan University, Chengdu, China, in 1997.

He is currently a Professor with Shenzhen University, Shenzhen, China. His current research interests include ultrasound image analysis, medical image processing, pattern recognition, and medical imaging.

Jun Cheng received the B. E. degree in electronic engineering and information science from the University of Science and Technology of China, and the Ph. D. degree from Nanyang Technological University, Singapore. He is now a senior research scientist in the Institute for Infocomm Research(I2R), Agency for Science,Technology and Research(A*STAR), working on AI for medical imaging, robust machine vision and perception. He is an Associate Editor for IEEE TIP and IEEE TMI.