

# No-reference image quality assessment for confocal endoscopy images with perceptual local descriptor

Xiangjiang Dong<sup>1</sup>,<sup>a</sup> Ling Fu,<sup>a,b</sup> and Qian Liu<sup>1</sup><sup>b,\*</sup>

<sup>a</sup>Huazhong University of Science and Technology, Wuhan National Laboratory for Optoelectronics, Wuhan, China

<sup>b</sup>Hainan University, School of Biomedical Engineering, Key Laboratory of Biomedical Engineering of Hainan Province, Hainan, China

## Abstract

**Significance:** Confocal endoscopy images often suffer distortions, resulting in image quality degradation and information loss, increasing the difficulty of diagnosis and even leading to misdiagnosis. It is important to assess image quality and filter images with low diagnostic value before diagnosis.

**Aim:** We propose a no-reference image quality assessment (IQA) method for confocal endoscopy images based on Weber's law and local descriptors. The proposed method can detect the severity of image degradation by capturing the perceptual structure of an image.

**Approach:** We created a new dataset of 642 confocal endoscopy images to validate the performance of the proposed method. We then conducted extensive experiments to compare the accuracy and speed of the proposed method with other state-of-the-art IQA methods.

**Results:** Experimental results demonstrate that the proposed method achieved an SROCC of 0.85 and outperformed other IQA methods.

**Conclusions:** Given its high consistency in subjective quality assessment, the proposed method can screen high-quality images in practical applications and contribute to diagnosis.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.27.5.056503](https://doi.org/10.1117/1.JBO.27.5.056503)]

**Keywords:** confocal endoscopy; image quality assessment; local binary pattern; differential excitation; human visual system.

Paper 210350GRR received Nov. 11, 2021; accepted for publication Apr. 29, 2022; published online May 18, 2022.

## 1 Introduction

Confocal endoscopy employs laser scanning confocal imaging technology to achieve real-time observation of mucosal cells and subcellular structures with micron-scale resolution to accurately locate lesions.<sup>1,2</sup> Probe-based confocal endoscopes are commonly used. They transmit laser and fluorescence images via a fiber bundle,<sup>3</sup> which has flexibility and accessibility for *in vivo* clinical imaging. The miniature objective system is the core component of confocal endoscopy for high resolution to directly visualize cells and is often assembled from multiple optical elements to correct aberrations.<sup>4</sup> It improves biopsy accuracy and contributes to the early diagnosis of cancer in various clinical fields such as human brain tumors<sup>5</sup> and gastrointestinal cancer.<sup>6</sup>

However, distortions, such as blur, noise, and decreased contrast, are common in confocal endoscopy imaging. Blur is the most common type of distortion in confocal endoscopy, which is caused by defocus, probe fiber core cross coupling,<sup>7</sup> and motion caused by the difference in movement speed between the investigated anatomical structures and the physician.<sup>8</sup> Owing to the small field-of-view of the confocal endoscopy,<sup>9</sup> to obtain a comprehensive view, a typical

---

\*Address all correspondence to Qian Liu, [qliu@hainanu.edu.cn](mailto:qliu@hainanu.edu.cn)

endoscopy examination produces thousands of images, most of which are not useful for diagnostic purposes<sup>5</sup> owing to image degradation and information loss caused by the above distortions. Manual removal of nondiagnostic and low-quality images is time-consuming and labor-intensive. Therefore, it is desirable to automatically screen high-quality images accurately and efficiently. An image enhancement method is observed to be beneficial in the automatic diagnosis of confocal endoscopy images,<sup>10</sup> and its development and evaluation also require the involvement of image quality assessment (IQA). Furthermore, the imaging performance evaluation of the confocal endoscopy also requires the participation of IQA; for example, Wang et al. analyzed the image histogram distribution to assess image contrast to validate the proposed confocal microendoscope.<sup>11</sup> Therefore, it is essential to develop an IQA method because it can benefit clinical applications of confocal endoscopy.

IQA is mainly divided into full reference (FR) and no reference (NR) methods. FR-IQA requires an ideal undistorted image when evaluating the quality, which is difficult to obtain in practical applications. NR-IQA is gaining attention because it does not require the use of reference images. The feature extraction and prediction model form the general NR-IQA framework. Common features describe poetry of images, including natural scene statistics (NSS) feature,<sup>12,13</sup> gradient feature,<sup>14</sup> frequency domain feature,<sup>15</sup> curvelet domain feature,<sup>16</sup> and discrete cosine transform (DCT) domain feature.<sup>17</sup> In addition, there are methods based on analysis of the perceptual process of image reception by the human eye, that is the human vision system (HVS), such as free energy theory<sup>18</sup> and phase congruency.<sup>19</sup> Owing to the powerful ability of image description, local descriptors in IQA have aroused extensive attention, such as local binary pattern (LBP),<sup>20,21</sup> from accelerated segment test (FAST),<sup>22</sup> speeded-up robust features (SURF),<sup>23</sup> Weber local descriptor (WLD),<sup>24</sup> and they have remarkable performance in multiply-distorted images.

Currently, NR-IQA has promising performance for images with a single distortion, such as blur, noise, and JPEG compression, while it is unsatisfactory when it comes to authentic and multiply distorted images<sup>25</sup> owing to joint distortion interactions. Deep learning techniques for IQA have been studied; however, the size of the dataset limits the network structure.<sup>26</sup> Bianco et al.<sup>27</sup> proposed the DeepBIQ method that uses a fine-tuned convolutional neural network (CNN) to extract features and feeds it to support vector regression (SVR) to predict the image quality. Liu et al.<sup>28</sup> proposed RankIQA, first pretrained the network on a large-scale self-build image pair database for an quality comparison of image pair task and fine-tuned the network to achieve promising performance. Ma et al.<sup>29</sup> proposed MEON, which first pretrained the network by an image distortion classification task on synthetic distortion images and then fine-tuned the network to achieve end-to-end image quality prediction. Zhu et al.<sup>30</sup> proposed MetalQA method, first adopted the meta-learning strategy to learn the prior knowledge from different NR-IQA tasks and fine-tuned the network to address the small sample problem. In general, deep learning requires a great number of training images, and the scale of the image quality database severely limits the performance of deep networks. Establishing larger datasets and designing a new method to reduce the requirement for training images is a direction that needs further exploration for deep IQA.

Medical images are multiply distorted because of variable imaging conditions; meanwhile, the content of medical images is distinct from that of natural images, mainstreaming IQA may decline in performance.<sup>31</sup> It is necessary to develop IQA of medical images. The study of medical IQA has increased because of the development of deep learning techniques, and it mainly focuses on ultrasound imaging, MRI, and OCT images. Zhang et al.<sup>32</sup> proposed DCNN-IQA-14 and ResNet-IQA based on traditional networks to predict the quality of ultrasound images. To overcome overfitting, a transfer-learning strategy was employed. Liu et al.<sup>33</sup> proposed a non-local residual neural network to assess slice-wise MRI image quality and applied random forest for the volumewise quality grade. Semisupervised learning and iterative self-training strategies were used for a few quality-annotated images. Wang et al.<sup>34</sup> analyzed the performance of four classic deep networks for assessing the quality of retinal OCT images by transfer learning, and the ResNet-50 network achieves highest performance. Medical images are more difficult to acquire and label than natural images, making the deep learning IQA method difficult to develop, transfer learning strategies based on traditional networks are widely adopted, and the potential of deep learning has yet to be fully explored.

The analysis of confocal image quality has made some progress. Aubreville et al.<sup>8</sup> proposed an improved Inception v3 network to detect motion artifacts in confocal endoscopy images. Kamen et al.<sup>35</sup> screened high quality and information-rich images by calculating image entropy before classifying images. Izadyyazdanabadi et al.<sup>36</sup> proposed a binary classification network to classify diagnostic and nondiagnostic images and applied fine-tuning and ensemble modeling techniques to improve performance and achieve high accuracy.<sup>37</sup> Despite this, none of these methods can quantify image quality, which limits the applications for screening and evaluating image enhancement algorithms. The signal-to-noise ratio is determined by a confocal endoscopy IQA<sup>38</sup> that can quantify image quality. However, regions of interest need to be manually selected before calculation, which does not apply to practical applications.

To meet the needs of practical applications, in this paper, we propose a new NR confocal laser endoscopy IQA (CEIQA) method, which combines local descriptors and Weber's law. First, we used a differential excitation (DE) map to describe the local variation information, calculated the LBP map of the image to describe structure information, and then computed the joint distribution histogram of DE and LBP as the first feature set. Second, for better perception of the ability to describe image information, we improved local ternary pattern (LTP) by changing its threshold function by referring to Weber's law and computed the histogram and entropy of improved LTP, which was used to measure the distribution of local variation patterns. Finally, SVR was applied to map the perceptual features to the quality score. Our main contributions are summarized as follows:

1. An NR-IQA method using local descriptors for confocal endoscopy images is proposed. The relationship between the local descriptors and image quality is detailed and analyzed.
2. A new dataset containing 642 confocal endoscopy images with corresponding subjective mean opinion score (MOS) from eight experienced researchers is established.
3. An extensive evaluation was conducted for the proposed method and other state-of-the-art NR-IQA methods. The experimental results show that CEIQA significantly outperforms other state-of-the-art methods.

The remainder of this paper proceeds as follows. In Sec. 2, we present the feature-extraction method for the proposed IQA. In Sec. 3, we present the details and results of the performance experiments. In Sec. 4, we discuss the limitations of the study and future work. Finally, in Sec. 5, we conclude the main contributions and provide potential application prospects for the proposed IQA.

## 2 Materials and Methods

### 2.1 Weber's Law and Differential Excitation

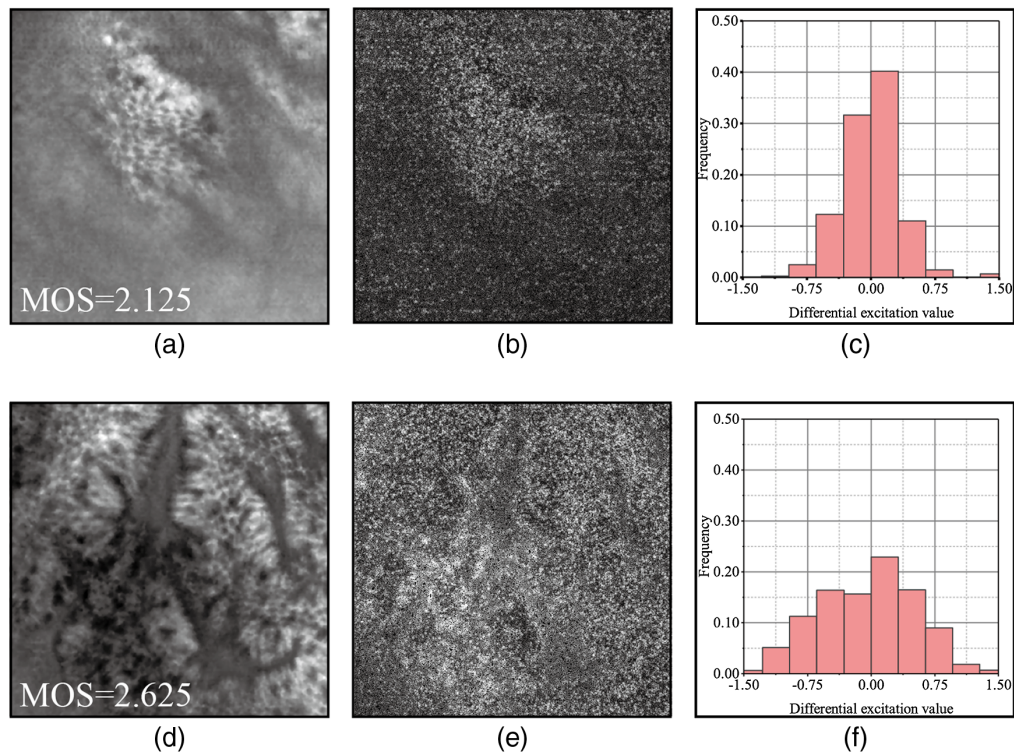
The perceptual process is sensitive to relative variations in pixel intensity in images when recording to the HVS. Weber's law can be used to describe this mechanism,<sup>39</sup> which is expressed as follows:

$$\frac{\Delta I_{\text{th}}}{I} = k, \quad (1)$$

where  $\Delta I_{\text{th}}$  represents the perceptual threshold,  $I$  represents the initial stimulus intensity, and  $k$  is a constant. Based on Weber's law, Chen et al.<sup>40</sup> proposed the WLD to extract local variation information in the image. One of the components is DE, which is calculated as follows:

$$I_{\text{DE}} = \arctan\left(\frac{\Delta I}{I}\right) = \arctan\left(\frac{\sum_{i=1}^{p-1} x_i - x_c}{x_c}\right), \quad (2)$$

where  $I$  denotes the original image,  $\Delta I$  characterizes the image local variation,  $I_{\text{DE}}$  is the DE map,  $x_c$  is the central pixel,  $x_i$  is the neighborhood pixels,  $p$  is the number of neighborhood pixels, and arctan function is used to prevent computation instability. After the above calculation, the range of  $I_{\text{DE}}$  becomes  $[-\pi/2, \pi/2]$ . Compared to Weber's law, DE regards the sum of the



**Fig. 1** DE of confocal endoscopy images. (a) and (d) Image with MOS of 2.125 and 2.625, respectively. (b) and (e) Corresponding DE map. DE map value is turned to  $0 - \pi/2$  by taking the absolute value and is scaled to 0 to 255. (c) and (f) Frequency histogram of DE map.

difference between the neighbors and the center as the change in the image. In this study, we adopted DE to describe the local variation in confocal endoscopy images. Figures 1(a) and 1(d) show two confocal endoscopy images with different MOSs and DE maps, respectively.

Figures 1(b) and 1(e) show that DE highlights the variation region in the image; thus, the distribution of the DE values shown in Figs. 1(c) and 1(f) represents the distribution of levels of local variation in the image. A low variation region means “the expected” according to the free energy theory,<sup>18</sup> which carries less information than a high-variation region. Figures 1(c) and 1(f) demonstrate that the DE value of the low-quality image is more often located around the zero point, whereas that of the high-quality image is more evenly distributed. In conclusion, the distributional characteristics of DE can be used as perceptual features that indicate image quality.

Nonetheless, Eq. (2) shows that during the accumulation phase, positive and negative variations will counteract each other, resulting in image variation information loss; meanwhile, the pattern and direction of local variation are ignored. Therefore, further analysis based on a DE map is required.

## 2.2 Improved LBP by Differential Excitation

The LBP<sup>41</sup> is a local descriptor with remarkable performance and has attracted considerable attention in IQA studies<sup>20,21</sup> owing to its ability to describe the local structure of an image. By comparing the interpixel relationships between the central pixel and its neighbors, LBP divides pixels into different patterns. The general form of LBP is rotation-invariant uniform LBP(LBP<sup>riu2</sup>),<sup>41</sup> defined as follows:

$$\text{LBP}_{P,R}^{\text{riu2}} = \begin{cases} \sum_{i=0}^{P-1} S(g_i - g_c), & \text{u(LBP}_{P,R}) \leq 2 \\ P + 1, & \text{else} \end{cases}, \quad (3)$$

where  $g_c$  and  $g_i$  are the central and circular neighborhood pixels, respectively,  $R$  is the radius of circular neighborhood,  $P$  is the number of  $g_i$ , and  $S(\cdot)$  is the thresholding function defined as follows:

$$S(g_i - g_c) = \begin{cases} 1, & g_i - g_c \geq 0 \\ 0, & g_i - g_c < 0 \end{cases}, \quad (4)$$

where  $u(\cdot)$  defined in Eq. (5) is a bitwise transition function that recognizes a uniform pattern whose number of bitwise transitions is less than two. Uniform patterns contribute more to the description of the image structure than basic patterns.  $\text{LBP}^{\text{riu}2}$  has  $P + 1$  uniform patterns and one other pattern, coding from 0 to  $P + 1$ :

$$u(\text{LBP}_{P,R}) = \|S(g_{P-1} - g_c) - S(g_0 - g_c)\| + \sum_{i=0}^{P-1} \|S(g_i - g_c) - S(g_{i-1} - g_c)\|. \quad (5)$$

The image structure contains information, and quality degradation will affect it, resulting in pattern shifts in the LBP.<sup>20</sup> Therefore, LBP can be used in the representation of image quality. The obtained LBP map is defined as follows:

$$I_{\text{LBP}} = \text{LBP}_{P,R}^{\text{riu}2}(I). \quad (6)$$

DE only calculates the intensity information and ignores local variation information of pattern and direction; meanwhile, LBP contains interpixel relationships without pixel intensity information. Therefore, it is helpful to calculate the joint distribution histogram of DE and LBP to compensate for both the shortage in describing the structure and preserving the local variation intensity and pattern of pixels. The joint distribution histogram was calculated as follows:

$$H(m, n) = P(I_{\text{LBP}} = m \cap I_{\text{DE}} = n), \quad (7)$$

where  $H(m, n)$  is a two-dimension joint histogram,  $P(\cdot)$  indicates the frequency function.  $m \in \{0, \dots, M\}$ , where  $M = P + 1$  is the number of uniform patterns of  $I_{\text{LBP}}$  and  $n \in \{1, \dots, N\}$ , where  $N$  is the number of bins in the histogram of the  $I_{\text{DE}}$ . After the joint histogram  $H(m, n)$  is obtained, it is an  $M \times N$  dimension feature to characterize the local variation and structure of the image.

### 2.3 Improved LTP by Weber's Law

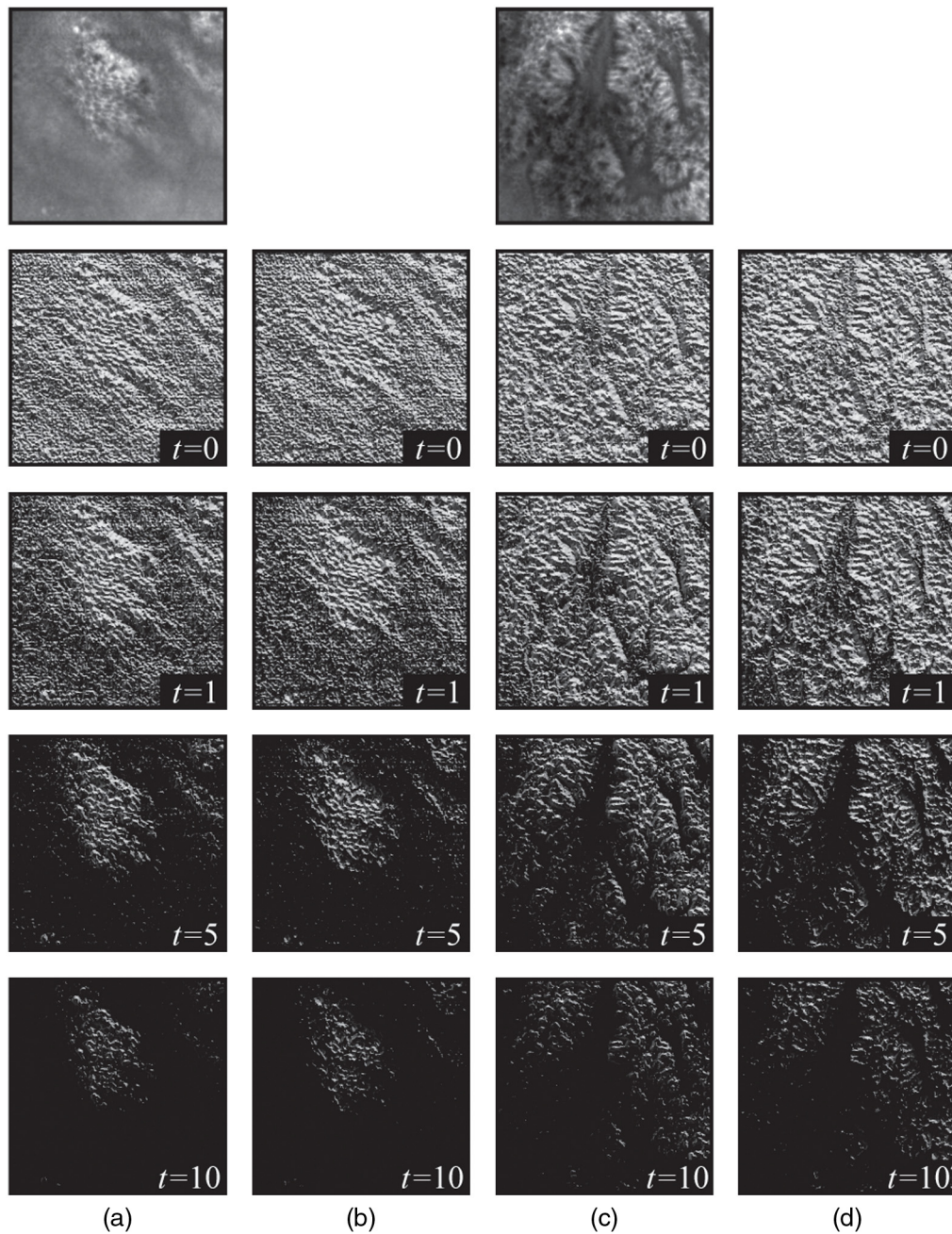
The LTP<sup>42</sup> is the generalization of LBP by changing the threshold function to obtain more detailed information of the local interpixel relationship. LTP was calculated as follows:

$$\text{LTP} = \sum_{i=0}^{P-1} 3^i \cdot T(g_i - g_c), \quad (8)$$

$$T(g_i - g_c) = \begin{cases} 1, & g_i - g_c > t \\ 0, & |g_i - g_c| \leq t \\ -1, & g_i - g_c < -t \end{cases}, \quad (9)$$

$$\text{LTP}_{\text{up/low}} = \sum_{i=0}^{P-1} 2^i C_{\text{up/low}}. \quad (10)$$

The obtained LTP map contains  $3^8$  patterns when  $P = 8$  because LTP codes  $C = T(g_i - g_c)$  can be  $-1, 0$ , and  $1$ . For lower computational complexity, the original LTP map is converted to an up-pattern and low-pattern map. The up-pattern map is obtained by turning the LTP code  $C$  from  $-1$  to  $0$  in Eq. (10), and a low-pattern map is obtained by turning the LTP code  $C$  from  $1$  to  $0$  and  $-1$  to  $1$ . Therefore, the up-pattern and low-pattern maps both contain  $2^8$  patterns with values ranging from  $0$  to  $255$  as LBP. Figure 2 shows the LTP map of the two confocal endoscopy images with different MOSs. Note that because of the threshold in the threshold function, LTP emphasizes the high variation region and underestimates the low variation region, which is similar to DE, while the choice of threshold value affects the screening strength of local variations. However, Fig. 2



**Fig. 2** LTP of confocal endoscopy images at different thresholds, the thresholds of LTP are 0, 1, 5, and 10 from top to bottom. (a) and (c) Up pattern. (b) and (d) Low pattern.

shows that the LTP map retains an image structure like LBP; thus, LTP can describe image quality degradation. Considering that LTP captures the local variation before extracting the structure, we apply Weber's law to the threshold function of LTP as follows:

$$T_{WB}(g_i - g_c) = \begin{cases} 1, & \frac{g_i - g_c}{g_c} > t \\ 0, & \frac{|g_i - g_c|}{g_c} \leq t \\ -1, & \frac{g_i - g_c}{g_c} < -t \end{cases} \quad (11)$$

where  $T_{WB}(g_i - g_c)$  consults the form of Weber's law regarding  $g_i - g_c$  as  $\Delta I$ ,  $g_c$  as  $I$ , and  $t$  as threshold  $k$ . To prevent the instability of the results, the overall image  $I$  is added by one. After

introducing Weber's law in the threshold function, the judgment of local variation is under the HVS.

Threshold  $t$  in the threshold function of LTP determines the ability to capture an image variation and further affects the image description. Therefore, the threshold  $t$  plays a key role in the performance of WB-LTP. Referring the form of Weber's law, the threshold  $t$  is calculated as follows:

$$t = \frac{\tan(\text{mean}(I_{\text{DE}}))}{256}, \quad (12)$$

where  $I_{\text{DE}}$  is the DE map in Eq. (2),  $\text{mean}(I_{\text{DE}})$  denotes the average variation intensity of image  $I$ . Function  $\tan()$  and factor  $1/256$  are used to transfer the range of  $t$  from  $[-\pi/2, \pi/2]$  to the range of  $\frac{\Delta I}{I}$  for corresponding to Eq. (11). The threshold  $t$  indicates that the region above or below average variation intensity of the image will be regarded as the high variation or low variation region, respectively.

After LTP improvement, we conducted further analysis to enrich the information expression of images. Considering that the up and low channels denote different directions of patterns, we referred to the calculation of the gradient magnitude<sup>14</sup> and computed the magnitude channel  $I_{\text{MAG}}$  as follows:

$$I_{\text{MAG}} = \sqrt{I_{\text{UP}}^2 + I_{\text{LOW}}^2}. \quad (13)$$

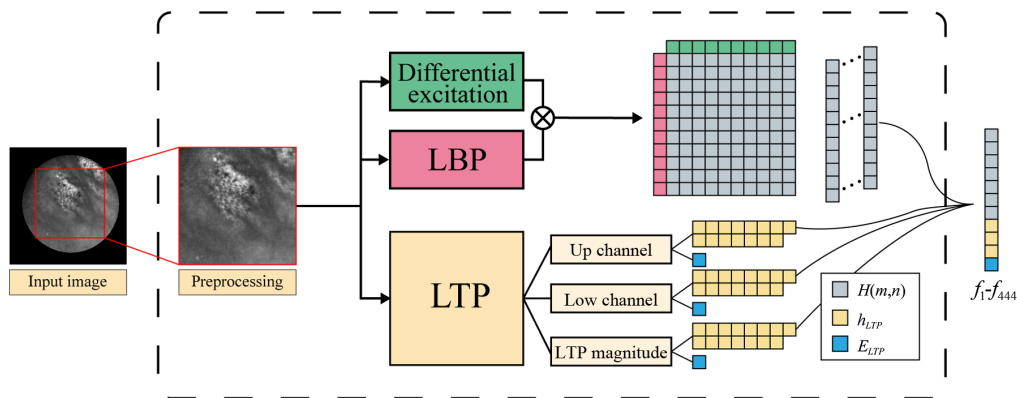
Because LTP reveals local structure distribution, it is useful to calculate the entropy of LTP to characterize image information as follows:

$$E = - \sum_{i=0}^{255} p_i \log_2 p_i, \quad (14)$$

where  $p_i$  denotes the frequency of grayscale value  $i$  in the image. We calculated  $E_{\text{up}}$ ,  $E_{\text{low}}$ , and  $E_{\text{MAG}}$ . Finally, we obtained the WB-LTP feature  $f_{\text{WB-LTP}} = \{h_{\text{up}}, h_{\text{low}}, h_{\text{MAG}}, E_{\text{up}}, E_{\text{low}}, E_{\text{MAG}}\}$ , where  $h_{\text{up}}$ ,  $h_{\text{low}}$ , and  $h_{\text{MAG}}$  are the histograms of  $I_{\text{up}}$ ,  $I_{\text{low}}$ , and  $I_{\text{MAG}}$ , respectively.

## 2.4 Feature Extraction and Quality Model

Because the field-of-view of confocal endoscopy is circular, acquired images appear as circular effective regions surrounded by black regions, and the latter interferes with the image description. Therefore, before feature extraction, the image should be preprocessed by adopting the square inscribed in the valid circle region, as shown in Fig. 3. After preprocessing, the DE-LBP of the image was calculated using  $p = 8$  in DE,  $R = 1$ , and  $P = 8$  in LBP and



**Fig. 3** Flowchart of feature extraction.  $H(m, n)$  denotes DE-LBP feature,  $h_{\text{LTP}}$  and  $E_{\text{LTP}}$  denote WB-LTP feature.

$M = N = 10$  to obtain 100-dimensional features. Then, computing the WB-LTP of the image with 15 bins in the histogram and acquire 48 dimension features.

To obtain multiscale information of the image, the image is downsampled twice beside the origin scale. Features are extracted in three scales; thus, the features have 444 dimensions. In WB-LTP, the structure information differs from the three scales of images, leading to different thresholds. Therefore, the thresholds of different scales  $t_s$  was calculated as  $t_s = t/2^s$ , where  $1/2^s$  is the scale factor,  $s \in \{0, 1, 2\}$  is the number of image downsampling.

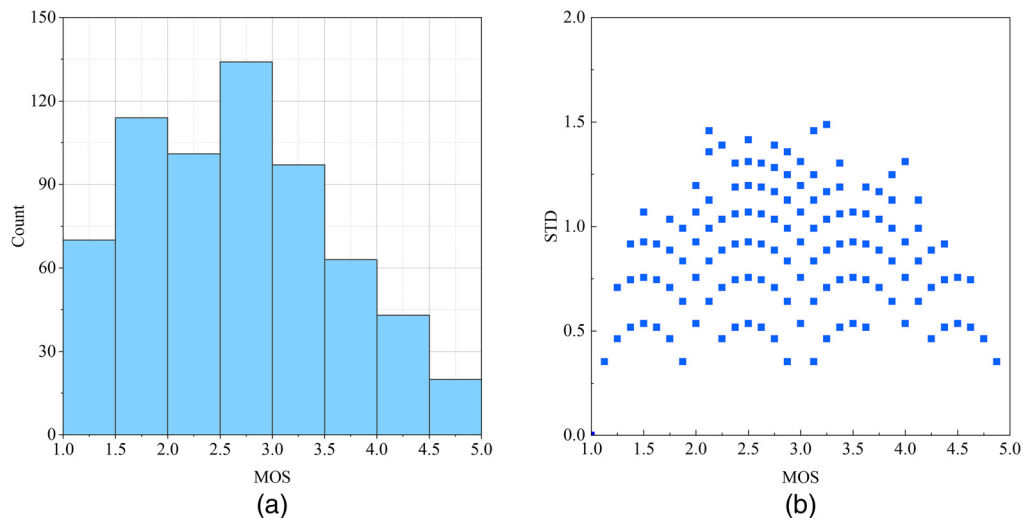
After obtaining the image features, SVR<sup>43</sup> with a radial basis function kernel was adopted to build a quality prediction model.

## 2.5 Confocal Endoscopy Image Database

To compare the performance of IQA methods, we established a database of confocal endoscopy images. The imaging experiment was conducted using a confocal endoscope designed by Wang et al.<sup>1</sup> The confocal endoscope has a field of view of  $300 \times 300 \mu\text{m}$  and a resolution of  $4.4 \mu\text{m}$ , and the image was obtained with  $1024 \times 1024$  pixels at a frame rate of 4 to 16 fps. Imaging experiments were conducted with *ex vivo* imaging of colonic and gastric tissues of female specific pathogen free and Sprague Dawley rats weighing  $\sim 150$  g. Imaging experiments obtained 656 images with blur, contrast distortion, and motion artifacts. All imaging experiments were approved by the animal experiment guidelines of the Animal Experimentation Ethics Committee of Huazhong University of Science and Technology (HUST, Wuhan, China).

To obtain meaningful MOS, eight researchers with extensive experience in instrument operation and image processing of confocal endoscopy rated the quality of images. Subjective quality assessment experiments apply single-stimulus (SS) methods. Every observer watches one confocal image 10 s on a computer monitor every time and gives a quality index ranging from one to five, where one denotes the lowest quality and five denotes the highest quality. To avoid observer exhaustion, a session lasted for half an hour and the observers watched 180 images. After a session, the observers rested for 8 min. Thus, there were four sessions in subjective quality assessment experiments.

After obtaining the eight rates of every image, the standard deviation (STD) of every image quality score was calculated. The image with STD higher than 1.5 was discarded because the result cannot reflect objective image quality. Eight images were discarded in this process, and 642 images remained. Finally, the MOS of the images was computed by averaging the scores of the researchers, and the distribution of the image quality scores is shown in Fig. 4(a). The relationship between STD and MOS is shown in Fig. 4(b), which we can see that the consistency of



**Fig. 4** (a) MOS distribution of the established confocal endoscopy image database. (b) The relationship between MOS and corresponding STD of images in the confocal image database.



observers' opinions is higher when faced with relatively high-quality and low-quality images compared with images of medium quality.

### 3 Results

#### 3.1 Experimental Protocol

We compared the proposed method with 11 state-of-the-art NR-IQAs with publicly available source codes. The methods used for comparison were the NSS feature-based IQA methods including BRISQUE<sup>12</sup> and SINQ<sup>44</sup> in the spatial domain, NBIQA<sup>13</sup> in the spatial and DCT domains, and CurvletQA<sup>16</sup> in the curvelet domain. There are local descriptor-based IQA methods including GWH-GLBP<sup>20</sup> using a gradient magnitude-weighted LBP feature, ORACLE<sup>45</sup> using the FAST algorithm, and the fast RetinaKeypoint descriptor (FREAK), NOREQI<sup>23</sup> using the SURF algorithm, and RATER<sup>46</sup> using FAST. In addition, there are image spatial and spectral entropy feature-based method SSEQ,<sup>15</sup> free-energy feature-based NFERM,<sup>18</sup> and gradient magnitude feature-based GM-LOG.<sup>14</sup>

All the above methods follow the process of feature extraction, SVR training, and prediction. Before feature extraction, all IQA models apply the same preprocessing method. For IQA employing color space information, grayscale values were used as inputs to the three color channels.

In the performance experiment, a confocal endoscopy image dataset was applied. First, 80% of the dataset were randomly chosen to train the SVR model, and the rest were used to test the performance. During the training phase of all IQA methods, SVR parameters were optimized using a grid search to achieve the best performance for fair comparison.

In the performance evaluation, Spearman rank order correlation coefficient (SROCC), Pearson's linear correlation coefficient (PLCC), and root mean square error (RMSE) were used to characterize the monotonicity and accuracy of prediction. SROCC and PLCC values closer to 1 and RMSE closer to 0 indicates better performance. Before the PLCC and RMSE are calculated, the nonlinear logistic regression shown in Eq. (15) is required,<sup>47</sup> where  $x$  is the predicted score,  $f(x)$  is the fitting score, and  $\beta_{1-5}$  is the regression parameter:

$$f(x) = \beta_1 \left( \frac{1}{2} - \frac{1}{\exp(\beta_2(x - \beta_3))} \right) + \beta_4 x + \beta_5. \quad (15)$$

The random 80% to 20% train-test is repeated 1000 times, and the median of performance criteria are reported. For a fair comparison, all methods use the same training and testing sets in each repeat.

#### 3.2 Performance Comparison

Table 1 shows the performance of NR-IQA, and the best method is shown in bold. As shown in Table 1, CEIQA outperforms the other IQAs in all criteria, followed by NOREQI.

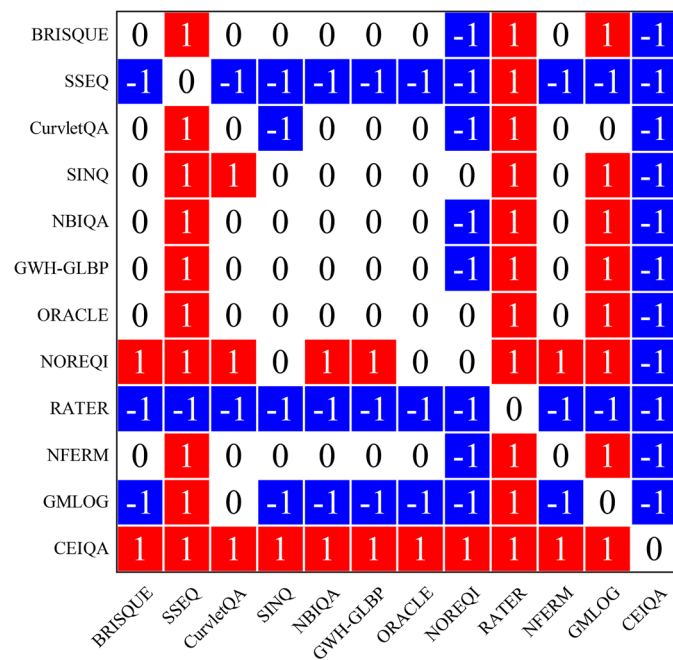
To further verify whether the performance difference is significant, we conducted the corrected resampled paired Student's  $t$ -test<sup>48</sup> between different methods of SROCC values of 1000 train-test repeats. The results are shown in Fig. 5, where symbols "1," "-1," and "0" mean that the method in the row is statistically (with 95% confidence) better, worse, or similar to the method in the column, respectively. Figure 5 shows that CEIQA is significantly superior to all other NR-IQA methods on the confocal endoscopy image dataset, followed by NOREQI.

To further analyze the relationship between the algorithm's performance and the consistency between the predicted quality scores and MOS, the scatter plots of MOS against the predicted quality scores of the test dataset from the IQA methods in a train-test repeat are shown in Fig. 6. The corresponding SROCC values were labeled in the plots. For a clear view, we only show the results of BRISQUE, SSEQ, CurvletQA, NOREQI, GWHGLBP, ORACLE, GMLOG, and the proposed CEIQA method.

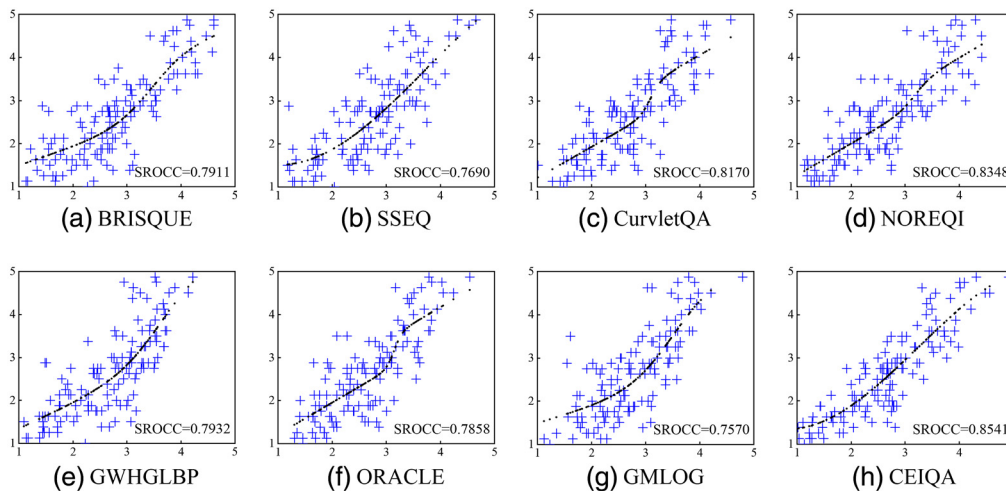
As shown in Fig. 6, the CEIQA with the highest SROCC shows the promising performance with most densely and evenly distribution of points around the fitting curve, which indicates the great consistency between MOS and the predicted quality scores. From Fig. 6, we can also see

**Table 1** Performance comparison of NR-IQA methods. The best IQA methods are highlighted in boldface.

IQA	SROCC	STD	PLCC	STD	RMSE	STD	Time (s)
BRISQUE <sup>12</sup>	0.8002	0.0346	0.8194	0.0301	0.5263	0.0343	0.0454
SSEQ <sup>15</sup>	0.7324	0.0409	0.7510	0.0370	0.6042	0.0372	0.6583
CurvletQA <sup>16</sup>	0.7837	0.0361	0.7970	0.0319	0.5541	0.0344	1.3147
SINQ <sup>44</sup>	0.8094	0.0320	0.8249	0.0270	0.5175	0.0322	1.8766
NBIQA <sup>13</sup>	0.8017	0.0328	0.8175	0.0281	0.5292	0.0326	9.5719
GWH-GLBP <sup>20</sup>	0.8007	0.0305	0.8047	0.0280	0.5438	0.0335	0.0645
ORACLE <sup>45</sup>	0.8054	0.0307	0.8185	0.0273	0.5248	0.0310	0.3492
NOREQI <sup>23</sup>	0.8259	0.0286	0.8370	0.0250	0.5017	0.0304	0.3310
RATER <sup>46</sup>	0.6463	0.0476	0.6778	0.0429	0.6733	0.0350	11.8290
NFERM <sup>18</sup>	0.7948	0.0355	0.8137	0.0356	0.5309	0.0412	30.7871
GM-LOG <sup>14</sup>	0.7677	0.0367	0.7885	0.0327	0.5628	0.0355	<b>0.0441</b>
DE	0.7687	0.0379	0.7881	0.0347	0.5652	0.0374	0.0082
LBP	0.8286	0.0279	0.8454	0.0238	0.4916	0.0301	0.0696
DE-LBP	0.8401	0.0287	0.8552	0.0248	0.4771	0.0329	0.0799
LTP	0.8383	0.0282	0.8522	0.0235	0.4803	0.0292	0.0408
WB-LTP	0.8434	0.0253	0.8541	0.0217	0.4769	0.0288	0.0405
CEIQA	<b>0.8543</b>	<b>0.0251</b>	<b>0.8648</b>	<b>0.0215</b>	<b>0.4615</b>	<b>0.0283</b>	0.1166



**Fig. 5** Results of statistically significant experiments using corrected resampled paired Student's *t*-test. Symbols "1," "-1," and "0" mean that the method in the row is statistically (with 95% confidence) better, worse, or similar than the method in the column, respectively.



**Fig. 6** Scatter plots of MOS against the predicted MOS from the different IQA methods. The  $x$  axis denotes the predicted score of IQA methods, and the  $y$  axis denotes the MOS. The SROCC is reported in figures. The dashed line is the fitting curve calculated by Eq. (15).

that the higher SROCC indicates better consistency between MOS and the quality scores predicted by the method.

The robustness of the algorithm is an important factor in the performance. To evaluate the robustness of IQA, we calculated the STD of the three criteria across 1000 repeats, which are shown in Table 1. The lower the STD, the more stable the algorithm performs for varying images. According to Table 1, CEIQA is the most stable among the IQA methods.

The rightmost column of Table 1 shows the time taken by different algorithms to extract the confocal endoscopy image features, which accounts for most of the total algorithm runtime. CEIQA has a promising speed and runs faster than NOREQI with the second-best performance. Experiments were performed on MATLAB R2017a with an Intel i7-8750HQ CPU at 2.20 GHz.

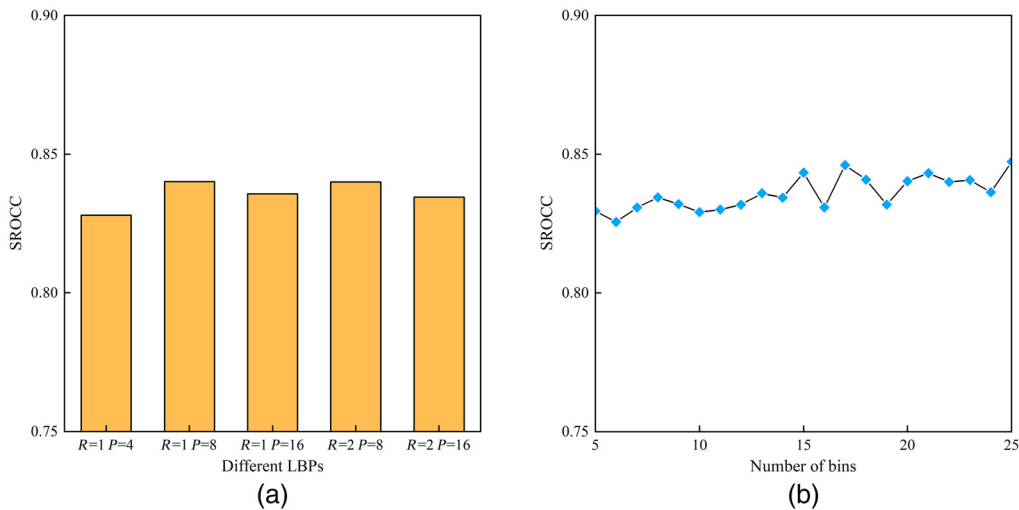
The experimental results demonstrated that CEIQA with a combination of perceptual laws and local descriptors has promising performance. Furthermore, to verify the enhancement of the proposed method using perceptual laws and the performance of the IQA components, we compared the performance of LBP and LTP before and after being improved using perceptual laws using the same 1000 train-test procedure. The results are presented in Table 1.

According to the results, the performance of LBP and LTP is remarkable, which demonstrates that the local descriptor is suitable for the quality prediction of confocal endoscopy images with multiple distortions. The same conclusion can be drawn from the fact that the NOREQI with the second-best performance also uses the local descriptors SURF, as shown in Table 1. Note that LTP uses the threshold calculation method proposed by Freitas et al.<sup>49</sup> Furthermore, DE-LBP, which combines LBP and DE, significantly outperforms origin LBP, while WB-LTP also performs better than origin LTP owing to Weber's law engagement. Introducing perceptual laws enhances the capability to describe image information and assess image quality.

### 3.3 Analysis of Parameters

There are two types of parameters in the proposed CEIQA method. The first is the neighborhood radius  $R$  and the number of neighborhood pixels  $P$  in DE-LBP. The second is the number of bins in the WB-LTP histogram. To verify if the performance of CEIQA is sensitive to the variations of parameters, we performed two experiments with a variety of parameters. First, we compared the performance of DE-LBP for different values of  $R$  and  $P$ . The 1000 train-test process, as in Sec. 3.1, is conducted, and the median of the SROCC of the entire loop is shown in Fig. 7(a). Then, the performance experiment of WB-LTP for different numbers of bins was conducted, and the result is shown in Fig. 7(b).

As shown in Fig. 7, DE-LBP's performance is stable across different settings of LBP and WB-LTP's performance is stable in different bins number of the histogram in a moderately large



**Fig. 7** Performance of the proposed methods of different parameters. (a) SROCC of DE-LBP in different radius  $R$  and number of neighboring pixels  $P$ . (b) SROCC of different numbers of bins in WB-LTP histogram.

range. In conclusion, the CEIQA, that consists of DE-LBP and WB-LTP, is robust to parameter variations and has good generalization capability.

## 4 Discussion

It is meaningful to analyze the relationship between the algorithm's performance in predicting MOS and screening images. In clinical practice, image screening was employed by setting the quality threshold first, and then the images with a quality score lower than the quality threshold are labeled as "the low-quality image" which should be screened out and the images with a quality score higher than or equal to the quality threshold are labeled as "the high-quality image" that need to be kept. Therefore, the high consistency of the predicted quality scores and MOS indicates the high accuracy of image filtering. As shown in Table 1 and Fig 6, the proposed CEIQA method with great consistency between the predicted quality scores and MOS has the potential for practical application.

This study has some limitations. First, the confocal endoscopy images were obtained from a single confocal endoscopy instrument and two tissue types. The type of distortion is limited. This also limits the ability of the proposed CEIQA method to effectively characterize image quality. The lack of image morphology can cause the overfitting of the IQA. For example, IQA will provide higher quality scores to images with the tissue of high local variability and ignore distortion during imaging. The algorithm also must be improved to ignore image content and focus on image distortion. Second, the objectivity of subjective IQA of image MOS must be further improved, which in turn will affect the performance evaluation and application potential of the algorithm.

Future work will focus on obtaining confocal endoscopy images of more tissues, imaging conditions, and more detailed and clear types of distortion, as well as conducting more comprehensive subjective IQA experiments to obtain more meaningful MOS. With additional data, CEIQA is expected to show better performance and can be universally applied. Other directions of research could involve using the IQA method to evaluate and improve confocal endoscopy image enhancement and deconvolution algorithms or using the IQA method to select high-quality images in clinical practice and analyzing the effectiveness of the method.

## 5 Conclusion

In this study, we proposed a new NR-IQA method named CEIQA based on Weber's law and a local descriptor. The image structural information is measured using the local descriptor, which

is then improved by Weber's law to extract perceptual features. The method is compared with 11 state-of-the-art NR-IQA methods on the introduced dataset of confocal endoscopy images. The dataset contains 642 images with authentic distortion and the corresponding MOS assessed by eight experimenters. As shown in the experimental results, CEIQA is significantly superior to other NR methods in terms of accuracy and robustness, which demonstrates that CEIQA has great potential for practical application and contributes to clinical diagnosis.

## Disclosures

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 81971692).

## Code, Data, and Materials Availability

The confocal endoscopy image dataset can be obtained from the authors upon reasonable request. The CEIQA source code and demo code is available in the Figshare repository via the link <https://figshare.com/articles/software/CEIQAcode/16987909>.

## References

1. J. Wang et al., "A confocal endoscope for cellular imaging," *Engineering* **1**(3), 351–360 (2015).
2. H. Li et al., "Advanced endoscopic methods in gastrointestinal diseases: a systematic review," *Quantum. Imaging Med. Surg.* **9**(5), 905–920 (2019).
3. Y. Wang et al., "Four-plate piezoelectric actuator driving a large-diameter special optical fiber for nonlinear optical microendoscopy," *Opt. Express* **24**(17), 19949 (2016).
4. L. Yang et al., "Five-lens, easy-to-implement miniature objective for a fluorescence confocal microendoscope," *Opt. Express* **24**(1), 473 (2016).
5. N. L. Martirosyan et al., "Prospective evaluation of the utility of intraoperative confocal laser endomicroscopy in patients with brain neoplasms using fluorescein sodium: experience with 74 cases," *Neurosurg. Focus* **40**(3), E11 (2016).
6. Z. Li et al., "New classification of gastric pit patterns and vessel architecture using probe-based confocal laser endomicroscopy," *J. Clin. Gastroenterol.* **50**(1), 23–32 (2016).
7. A. K. Eldaly et al., "Deconvolution and restoration of optical endomicroscopy images," *IEEE Trans. Comput. Imaging* **4**(2), 194–205 (2018).
8. M. Aubreville et al., "Deep learning-based detection of motion artifacts in probe-based confocal laser endomicroscopy images," *Int. J. Comput. Assist. Radiol. Surg.* **14**(1), 31–42 (2019).
9. H. Li et al., "500  $\mu\text{m}$  field-of-view probe-based confocal microendoscope for large-area visualization in the gastrointestinal tract," *Photonics Res.* **9**(9), 1829 (2021).
10. N. Ghatwary et al., "Automatic grade classification of Barretts esophagus through feature enhancement," *Proc. SPIE* **10134**, 1013433 (2017).
11. J. Wang et al., "Near-infrared probe-based confocal microendoscope for deep-tissue imaging," *Biomed. Opt. Express* **9**(10), 5011 (2018).
12. A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.* **21**(12), 4695–4708 (2012).
13. F.-Z. Ou, Y.-G. Wang, and G. Zhu, "A novel blind image quality assessment method based on refined natural scene statistics," in *IEEE Int. Conf. Image Process.*, Taipei, Taiwan, pp. 1004–1008 (2019).
14. W. Xue et al., "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Trans. Image Process.* **23**(11), 4850–4862 (2014).

15. L. Liu et al., "No-reference image quality assessment based on spatial and spectral entropies," *Signal Process. Image Commun.* **29**(8), 856–863 (2014).
16. L. Liu et al., "No-reference image quality assessment in curvelet domain," *Signal Process. Image Commun.* **29**(4), 494–505 (2014).
17. M. A. Saad, A. C. Bovik, and C. Charrier, "DCT statistics model-based blind image quality assessment," in *18th IEEE Int. Conf. Image Process.*, Brussels, Belgium, pp. 3093–3096 (2011).
18. K. Gu et al., "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia* **17**(1), 50–63 (2015).
19. C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Trans. Neural Network* **22**(5), 793–799 (2011).
20. J. Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Process. Lett.* **23**(4), 541–545 (2016).
21. Q. Li et al., "Blind image quality assessment using statistical structural and luminance features," *IEEE Trans. Multimedia* **18**(12), 2457–2469 (2016).
22. M. Oszust, "Local feature descriptor and derivative filters for blind image quality assessment," *IEEE Signal Process. Lett.* **26**(2), 322–326 (2019).
23. M. Oszust, "No-reference image quality assessment using image statistics and robust feature descriptors," *IEEE Signal Process. Lett.* **24**(11), 1656–1660 (2017).
24. Y. Chen et al., "No-reference image quality assessment based on differential excitation," *Acta Autom. Sin.* **46**(8), 1727–1737 (2020).
25. D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.* **25**(1), 372–387 (2016).
26. X. Yang, F. Li, and H. Liu, "A survey of DNN methods for blind image quality assessment," *IEEE Access* **7**, 123788–123806 (2019).
27. S. Bianco et al., "On the use of deep learning for blind image quality assessment," *Signal Image Video Process.* **12**(2), 355–362 (2018).
28. X. Liu, J. Van De Weijer, and A. D. Bagdanov, "RankIQA: learning from rankings for no-reference image quality assessment," in *IEEE Int. Conf. Comput. Vision*, pp. 1040–1049 Venice (2017).
29. K. Ma et al., "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.* **27**(3), 1202–1213 (2018).
30. H. Zhu et al., "MetaIQA: deep meta-learning for no-reference image quality assessment," in *IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, Seattle, pp. 14131–14140 (2020).
31. S. Zhang et al., "CNN-based medical ultrasound image quality assessment," *Complexity* **2021**, 9938367 (2021).
32. S. Liu et al., "Real-time quality assessment of pediatric MRI via semi-supervised deep non-local residual neural networks," *IEEE Trans. Image Process.* **29**, 7697–7706 (2020).
33. J. Wang et al., "Deep learning for quality assessment of retinal OCT images," *Biomed. Opt. Express* **10**(12), 6057 (2019).
34. L. S. Chow and R. Paramesran, "Review of medical image quality assessment," *Biomed. Signal Process. Control* **27**, 145–154 (2016).
35. A. Kamen et al., "Automatic tissue differentiation based on confocal endomicroscopic images for intraoperative guidance in neurosurgery," *Biomed. Res. Int.* **2016**, 6183218 (2016).
36. M. Izadyyazdanabadi et al., "Improving utility of brain tumor confocal laser endomicroscopy: objective value assessment and diagnostic frame detection with convolutional neural networks," Orlando, Florida (2017).
37. M. Izadyyazdanabadi et al., "Convolutional neural networks: ensemble modeling, fine-tuning and unsupervised semantic localization for neurosurgical CLE images," *J. Vis. Commun. Image Represent.* **54**, 10–20 (2018).
38. V. Becker et al., "Intravenous application of fluorescein for confocal laser scanning microscopy: evaluation of contrast dynamics and image quality with increasing injection-to-imaging time," *Gastrointest. Endosc.* **68**(2), 319–323 (2008).
39. A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, New Jersey (1989).

40. J. Chen et al., "WLD: a robust local image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(9), 1705–1720 (2010).
41. T. Ojala et al., "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987 (2002).
42. W.-H. Liao, "Region description using extended local ternary patterns," in *20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, pp. 1003–1006 (2010).
43. C. Chang and C. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**(3), 1–27 (2011).
44. L. Liu et al., "Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment," *Signal Process. Image Commun.* **58**, 287–299 (2017).
45. M. Oszust, "Optimized filtering with binary descriptor for blind image quality assessment," *IEEE Access* **6**, 42917–42929 (2018).
46. M. Oszust, "No-reference image quality assessment with local gradient orientations," *Symmetry-Basel*. **11**(1), 95 (2019).
47. VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment—Phase II," 2003, <http://www.vqeg.org/>.
48. C. Nadeau and Y. Bengio, "Inference for the generalization error," *Mach. Learn.* **52**, 239–281 (2003).
49. P. G. Freitas, W. Y. L. Akamine, and M. C. Q. Farias, "No-reference image quality assessment based on statistics of local ternary pattern," in *Eighth Int. Conf. Quality Multimedia Experience*, Lisbon, Portugal, pp. 1–6 (2016).

**Xiangjiang Dong** is a master student at Huazhong University of Science and Technology. He received his BS degree in physics from Huazhong University of Science and Technology in 2016. His current research interests include medical image processing and image quality assessment.

**Ling Fu** received her PhD in Swinburne University of Technology in 2007. She is currently a doctoral supervisor at Huazhong University of Science and Technology. Her research fields include biomedical optical imaging and optical microscopy.

**Qian Liu** received his PhD from Huazhong University of Science and Technology in 2005. He is currently a doctoral supervisor at Hainan University. His research fields include medical imaging and medical devices.