Blind quality assessment of screen content images via edge histogram descriptor and statistical moments.

FARHADI TOLIE, H., FARAJI, M.R. and QI, X.

2024

Blind Quality Assessment of Screen Content Images via Edge Histogram Descriptor and Statistical Moments

Hamidreza Farhadi Tolie[a, *], Mohammad Reza Faraji[a], and Xiaojun Qi[b]

[a]Department of Computer Science and Information Technology, Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan 45137-66731, Iran
[b]Department of Computer Science, Utah State University, Logan, UT 84322-4205 USA

Abstract

With the growth in utilizing desktop sharing and remote control applications in recent years for many purposes like online education and remote working, quality assessment (QA) of screen images has become a hot topic. It could be used to enhance the user's quality experience. Currently, most screen image QA methods require a reference image, and the existing blind/no-reference methods do not consider both the image's content and chrominance degradations. This paper proposes a novel blind quality assessment method for screen content images (SCIs) through block-based content representation, which extracts content- and chromatic-based features on local, semi-global, and global scales. Our proposed edge histogram descriptor- and statistical moments-based (EHDSM) method divides the image into 16 blocks and then describes each block using its local edge and semi-global chrominance features. It also takes the global chrominance features into account to investigate how the image's color information is changed in the presence of chrominance distortions. Local features are extracted using edge histogram descriptor, while the semi-global and global features are measured by computing the statistical moments. Next, the quality assessment is achieved by training a support vector regression (SVR) model. Extensive experiments on three commonly-used SCI datasets have verified the superiority of our proposed EHDSM method compared with the state-of-the-art blind screen content image quality assessment methods.

Keywords— image quality assessment (IQA), screen content image (SCI), edge histogram descriptor (EHD), image content descriptor, chrominance features

* Corresponding author
h.farhadi@iasbs.ac.ir (H. F. Tolie); m.faraji@iasbs.ac.ir (M. R. Faraji); xiaojun.qi@usu.edu (X. Qi)

1 Introduction

The popularity of screen content images (SCIs) has risen by their utilization in applications such as online gaming, mobile Web browsing, screen sharing, and remote controls [1]. SCIs are basically a combination of natural scene images (NSIs) and computer-generated graphics, shapes, and texts. Thus, compared with NSIs, they usually have sharp edges, high contrast, and limited colors in specific regions [2, 3]. However, similar to NSIs, they lose their quality during the processing, compression, or transmission. Therefore, existing natural scene image quality assessment methods cannot efficiently evaluate the quality of SCIs, and their quality needs to be assessed using specifically designed SCI quality assessment (SCIQA) methods [4].

On the other hand, the original or reference image is unavailable in most cases, such as the remote screen sharing and transmission applications. As a result, no-reference or blind image quality assessment (BIQA) methods are the only choices. Several BIQA methods are designed to assess the perceptual quality of NSIs. Representative methods include natural image quality evaluator (NIQE) [5], integrated local NIQE (ILNIQE) [6], blind/referenceless image spatial quality evaluator (BRISQUE) [7], and gradient magnitude and the Laplacian of Gaussian (GM-LoG) [8]. Specifically, NIQE [5] uses natural scene statistics (NSS) to derive statistical features from the image. ILNIQE [6] improves NIQE by designing and training a multivariate Gaussian model using five types of NSS features. BRISQUE [7] uses the locally normalized luminance coefficients of an image to calculate its NSS. GM-LoG [8] proposes the joint statistics of two features including gradient magnitude and Laplacian of Gaussian to calculate the NSS of an image. However, these models cannot effectively evaluate the quality of the SCIs [9, 10] due to their distinct characteristics of computer-generated graphics, shapes, and texts. Yang *et al.* [11] also demonstrate that the intensity distribution of the SCIs and NSIs are remarkably different. Hence, researchers have recently proposed several specifically designed blind SCIQA (BSCIQA) methods to assess the perceptual quality of SCIs. These methods can generally be classified into three categories, namely, feature extraction-based, codebook-based, and neural network-based methods.

The first category includes methods that extract features capable of capturing the characteristics of SCIs. Representa-

tive blind methods include the blind quality measure for SCIs (BQMS) [2], screen image quality evaluator (SIQE) [3], no-reference luminance and texture-based method (NRLT) [12], BSCIQA by orientation selectivity mechanism (OSM) [13], the hybrid region features fusion method (HRFF) [14], the perceptual quality measure by spatial continuity (PQSC) [9], and quality assessment of SCIs via Fisher Vector Coding (FVC) [15]. BQMS extracts screen content-based features via free energy theory [16] and structural degradation model. SIQE extracts four types of features, including image complexity, screen content statistics, global brightness quality, and sharpness of details. NRLT exploits statistical luminance features in the form of histograms and also extracts statistical texture features by employing the local binary patterns (LBP) descriptor [17]. OSM employs the orientation selectivity mechanism [18] to extract the orientation information to describe the distorted SCIs. It also uses structure features as complementary information to further describe SCIs. HRFF first segments SCIs into sharp edge patches (SEPes) and non-SEPes, then extracts features such as entropy, contrast, and sharpness loss. It finally combines these features with global features extracted by the BRISQUE [7] method. PQSC extracts statistical chromatic and texture features to represent the chromatic continuity and degree of texture variation in SCIs. FVC employs the fisher vector coding technique to represent SCIs using offline-generated Gaussian mixture models. To assess the SCI's quality, these methods usually train a machine learning model such as SVR [2, 9, 15] using the subjective scores, also known as mean opinion scores (MOSs) which are the mean value of subjects' ratings, as the ground truth. Also, sometimes difference mean opinion scores (DMOSs) are used as the ground truth which are the difference in quality between images [11, 19, 20, 21].

The second category includes methods that employ dictionary learning algorithms to create and learn a codebook or dictionary, which will be used to predict the quality of SCIs. Representative blind methods include content-specific codebooks (CSC) [22] and macro-micro modeling of tensor domain dictionary (MMMTDD) [23]. The CSC method learns a codebook by training the K-singular value decomposition (K-SVD) dictionary learning algorithm [24] over small pictorial and textual patches. It then uses the sparse representation to effectively encode the patches via learned codebooks. Finally, it employs a pooling scheme to aggregate these patch-based codebooks to describe distorted SCIs. MMMTDD employs tensor decomposition to learn a dictionary with the principal components. It then uses a macro-micro model to automatically generate micro and macro features in the dictionary space. Micro features describe the particularity of the statistical distribution of sparse codes and macro features describe the relationship between the statistical distribution and quality degradation of SCIs. Both CSC and MMMTDD methods use SVR with the radial basis function (RBF) kernel to produce the quality score for a distorted SCI image.

The third category includes methods that employ neural networks and deep learning to assess the quality of SCIs. Representative blind methods include pseudo-natural input convolutional neural network (PICNN) [25], blind SCI quality assessment using stacked auto-encoders (BSCIQA-SAE) [26], and quadratically optimized model based on the deep convolutional neural network (QODCNN) [27]. PICNN employs a naturalization module to make SCIs more similar to natural images to extract deeper information. BSCIQA-SAE trains two SAEs and two regressors on hand-crafted features extracted from pictorial and textual regions to predict the quality of pictorial and textual areas separately. The two predicted scores are fused to produce the final perceptual quality score. QODCNN pre-trains a CNN model on some SCI patches and fine-tunes the pre-trained model with selected distorted image patches. An adaptive pooling scheme is then used to combine the quality scores from patches to measure the quality of an image.

Earlier feature extraction-based BSCIQA methods [2, 3] generally assume that the features follow certain distributions and utilize parametric models to extract quality-aware features [9]. Parametric models-based IQA methods assume that the extracted features for representing the image content follow a certain mathematical model or distribution [9], which may lead to information loss and inconsistency with human judgments. Furthermore, the majority of BSCIQA methods do not consider the chrominance information of an image, which has proven to be effective in both the blind and full-reference (FR) SCIQA or other areas of image quality assessment [9, 28, 29] since the human visual system (HVS) is highly sensitive to such information [30]. The codebook-based methods usually evaluate the quality of SCIs by automatically extracting and aggregating features from SCIs. However, the feature aggregation approach (e.g., percentage-based pooling [22] and log-normal distribution-based local pooling [23]) is still the bottleneck of these methods and has not been effectively resolved yet. Neural network-based methods highly depend on the availability of comprehensive SCI datasets. A majority of existing neural network-based SCIQA methods [25]-[27] focus on proposing practical training approaches to tackle the issues of the lack of large SCI datasets. However, these practical training approaches are not entirely successful.

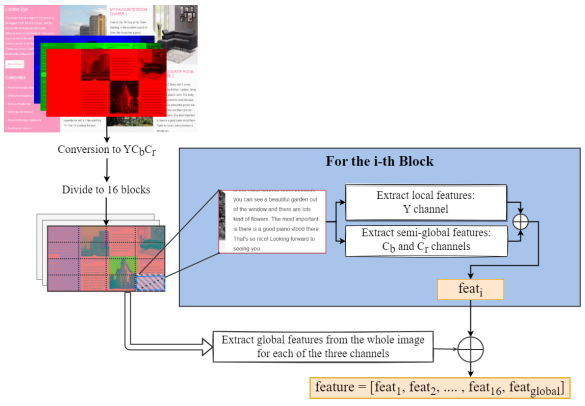

Figure 1: Framework of the proposed EHDSM method. $\oplus$ indicates the concatenation operation. $feat_i(i = 1, 2, .., 16)$ is the concatenated local and semi-global feature vector of the i-th block, which are further concatenated with the global features $feat_{global}$ to form the final feature vector.

Considering the HVS-compatible statistical details of an image [31] and the importance of zero-crossing at multi-

ple scales [32], we propose an edge histogram descriptor- and statistical moments-based (EHDSM) method to address the aforementioned limitations associated with feature extraction-based BSCIQA methods. EHDSM incorporates the histogram of edges and chrominance information at three scales (i.e., local, semi-global, and global scales) to extract quality-aware features without using a parametric model. Specifically, it converts a distorted SCI to the YC_bC_r color space and divides it into non-overlapping 4×4 blocks, as shown in Figure 1. The YC_bC_r color space is more compatible with SCI's characteristics and therefore yields better assessment results [28, 10]. It is also suggested by ITU-R BT.601 for video broadcasting (i.e., visuals displayed on screens). Next, it extracts features at local and semi-global scales from each block and at the global scale from the whole image. Particularly, local features are derived using the edge histogram descriptor (EHD) by computing the frequency of five types of edges. Semi-global and global scale features are chrominance information computed using first- and second-order statistical moments (i.e., mean and standard deviation).

The proposed EHDSM method makes the following contributions: 1) Introducing a blind block-based image quality assessment method to extract local content-based and semi-global and global chrominance features to effectively capture the effects of different distortion types and levels of SCIs; 2) Describing the content variation (i.e., pictorial and textual content) within an image with the HVS-compatible edge histogram descriptor on a local scale; 3) Extracting chromatic information using the statistical moments based on the content variation of an image (i.e., semi-global features) and its global statistics; 4) Achieving superior accuracy and computational run-time on three common SCI datasets compared with four NSI BIQA methods, seven state-of-the-art BSCIQA methods, and eight variant methods of the EHDSM.

The rest of this paper is organized as follows: Section 2 presents the proposed EHDSM blind SCIQA method. Section 3 compares the performance of the proposed method with state-of-the-art BSCIQA methods on three publicly available SCI datasets. Section 4 draws the conclusion.

2 Proposed Method

Edges and edge histograms are important and powerful features that have been commonly used for quality assessment for a large variety of images ranging from NSIs to SCIs [33]. Due to the high correlation of edge features with most structural distortions, a number of IQA techniques employ edges to represent the image content toward capturing the effect of distortions [34, 35]. For example, FR SCIQA methods such as gradient similarity score (GSS) [36], edge similarity (ESIM) [37], and multi-scale difference of Gaussian (MDOG) [38] use edge width, contrast, strength, and direction to measure the image quality. To improve efficiency and generalizability, normalization techniques can be further employed to make edge histograms invariant to translation, scale, and rotation and therefore enrich edge features to represent an image and capture its quality [39, 40].

As mentioned earlier, many studies employ edges and histograms to represent images. However, the integration of content descriptor-based edge histograms with the chrominance features have not been fully explored in BSCIQA and there is room for improvement. Unlike histogram or moments-based methods, which often treat images as a set of independent pixels, the proposed method takes into account the spatial arrangement of pixels by calculating the distribution of the proposed five edge types through the image blocks. Furthermore, by computing statistical moments across each channel of the entire image as well as its constituent blocks, our model gains the capability to comprehensively measure the overall changes in both content and chrominance information throughout the image. Our proposed blind screen content quality assessment method utilizes edge histograms (i.e., EHD) to capture edge variations in image blocks instead of the whole image to effectively describe the content of each block and the content variations among blocks. To the best of our knowledge, current BSCIQA methods have rarely considered the chrominance information loss when constructing feature extraction-based metrics for quality measurement. Therefore, the performance of the state-of-the-art content-based IQA methods drops when color distortions are introduced. For instance, their performance is very poor in presence of the color saturation change (CSC) distortion (refer to Table 4). Our EHDSM blind SCIQA method proposes an integrated hierarchical content and color feature extraction scheme to capture the content's clarity and variation (i.e., textual and pictorial contents) alongside the chrominance richness to effectively describe SCIs distorted under different distortion types and levels.

Observing SCIs often contain a variety of content including natural scenes, shapes, text, and graphics, we uniquely extract local EHD features [41] from the Y channel in the YC_bC_r color space to capture changes in brightness, which usually result in edges within various content. These local EHD features represent the frequency and the directionality of the brightness changes in an image and distinguish how the image content is distributed. Furthermore, the HVS is highly sensitive to the chrominance information of an image [28]. Therefore, we use the first and second statistical moments from both C_b and C_r channels to extract chrominance features at semi-global and global scales to measure the image's quality. These two kinds of features complement each other and are fused in the proposed method to provide highly discriminative power to assess the quality of the image. The SVR with the RBF kernel is trained to map the extracted quality-aware features to their corresponding subjective quality scores. In the following subsections, we explain how to extract features at different scales in more detail.

2.1 Local features: EHD

Local features have been used in many IQA studies to represent the frequency and directionality of the brightness changes in an image and distinguish how the image content is distributed. However, our work focuses on describing the pictorial and textual content of an image via five types of edge patterns. In other words, we employ edge histograms to describe the content of each block and the content variations among blocks to more accurately represent SCIs. To this end, we convert an SCI image to the YC_bC_r color space and divide the Y channel of the distorted SCI into nonover-

1	-1	1	1	√2	0	0	√2	2	-2
1	-1	-1	-1	0	-√2	-√2	0	-2	2
(a) e_v (vertical)		(b) e_h (horizontal)		(c) $e_{d\text{-}45}$ (diagonal-45)		(d) $e_{d\text{-}135}$ (diagonal-135)		(e) e_{nd} (non-diagonal)	

Figure 2: Filter coefficients corresponding to 5 edge types.

lapping 4×4 blocks (i.e., 4 blocks in the row direction and 4 blocks in the column direction) to ensure 16 blocks are produced regardless of the image size. This is done to get a fixed-size feature vector for each image. For each block, we then extract non-overlapping 2×2 image patches (i.e., local scale). For each patch, we check the presence of five edge types via the following equation:

$$max\{m_v, m_h, m_{d45}, m_{d135}, m_{nd}\} > T_{edge} \quad (1)$$

where T_{edge} is a threshold parameter (empirically set to 16) and m_v, m_h, m_{d45}, m_{d135}, and m_{nd} are the edge magnitudes of vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional edges, respectively. Eq. 1 indicates that the patch contains significant edge information when the maximum value of the five computed edge magnitudes is greater than the threshold. Otherwise, the patch contains little edge information. The edge magnitude m_{dir} ($dir = \{v, h, d45, d135, nd\}$) corresponding to one edge direction is computed by applying the respective edge filter shown in Figure 2 on a given image patch as follows:

$$m_{dir} = |\sum_{i=1}^{2}\sum_{j=1}^{2} P(i,j) \times e_{dir}(i,j)| \quad (2)$$

where P is the 2×2 image patch and e_{dir} is the edge filter corresponding to one of the five directions *(dir =*$\{$*v, h, d45, d135, nd*$\}$*)* shown in Figure 2.

For each block at the local scale, we construct two 5-bin histograms, where each bin contains information regarding each of the five edge types. One 5-bin histogram EHD_A contains the count of each edge type and another 5-bin histogram EHD_B contains the magnitudes of each edge type. When any patch within a block contains significant edge or non-edge information measured by the predetermined threshold parameter T_{edge}, the corresponding histogram bin of the dominant edge type in EHD_A and EHD_B increases by 1 and by the magnitude of the dominant edge, respectively. The total number of all patches in a block and the sum of the dominant edge magnitude within a block are respectively used to normalize EHD_A and EHD_B to achieve more invariance and robustness. The i-th image block is described by a 10-bin local EHD feature $feat_{i_{local}}$ as follows:

$$feat_{i_{local}} = [EHD_{i_A}, EHD_{i_B}] \quad (3)$$

where EHD_{i_A} and EHD_{i_B} are the normalized EHD_A and EHD_B of the $i-$th image block, respectively.

Figure 3 presents two samples of SCI blocks with different content types and their corresponding normalized histograms showing distributions of five kinds of edge types at vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional orientations in EHD_A and EHD_B. It shows that non-directional edge types are the common edge types for both SCI blocks. However, the block with textual

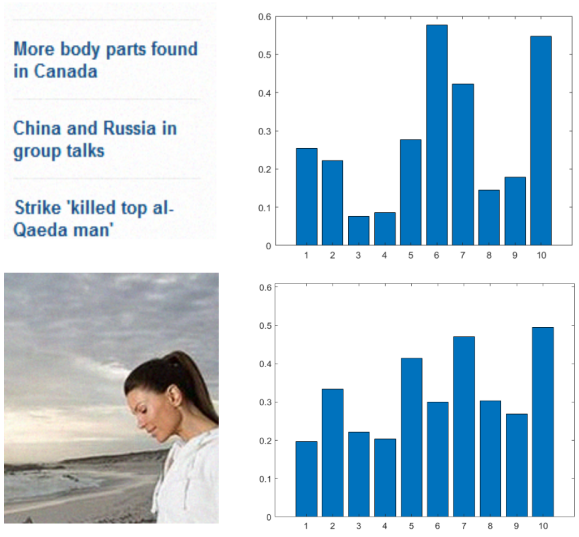

Figure 3: SCI blocks with different content types and their histograms by concatenating two normalized 5-bin histograms EHD_A and EHD_B . Top row: a sample SCI block with only textual content and its histogram. Bottom row: a sample SCI block with only pictorial content and its histogram.

content clearly has more vertical and horizontal edge types (i.e., bin numbers 1 and 2 in both normalized EHD_A and EHD_B or bin numbers 1, 2, 6, and 7 in the local 10-bin EHD, a concatenated histogram) than other directional edge types due to its sharp edges of the textual content in both vertical and horizontal directions. The block with pictorial content tends to have more horizontal edge types (i.e., bin number 2 in both normalized EHD_A and EHD_B or bin numbers 2 and 7 in the local 10-bin EHD) than the remaining directional edge types (i.e., vertical, 45-degree diagonal, and 135-degree diagonal). They also tend to be equally distributed due to the sharp edges of pictorial content in all orientations. These two examples demonstrate that the local 10-bin EHD of an SCI block captures the distribution of edge orientations and therefore captures the differences among various content types within a block. Since the objects within an image mainly occur at larger scales (i.e., image blocks), the edge histogram of a block differs according to the object content and different blocks capture various content types of different objects. As a result, the local EHD features capture the content within a block better and lead to a meaningful description of objects at different locations and different scales.

2.2 Semi-global and global chrominance features: statistical moments

The formulated local information, as explained in the previous section, has been proven to be an efficient way to represent the content of an image. In addition, local statistics are proper features to capture distortions since they are usually altered with the change of distortions, specifically the luminance change [42]. However, local statistics cannot effectively represent the whole image since some distortions affect the chrominance channels and some distortions de-

grade the entire image. For instance, color saturation change, which occurs when sharing a screen between different devices [28], alters the color instead of the structural information of an image.

To capture distortions at different scales, we propose to extract semi-global and global chrominance features to quantify and characterize the chrominance information of a distorted SCI. Instead of extracting the statistical information only from an entire image, we extract and combine the statistical information from both the non-overlapping image blocks and the entire image. This enables us to capture the color degradation in different regions of an image. Inspired by the study in [43], we compute the first- and second-order statistical moments (i.e., mean and standard deviation) to describe the semi-global chrominance features of each block and the global chrominance features of an image. Specifically, we use the statistical moments of each non-overlapping 4×4 block from both C_b and C_r channels to represent the chrominance features of an image on the semi-global scale. For the i-th block ($i = 1, ..., 16$), the mean μ_i and standard deviation σ_i of each channel are computed by:

$$\mu_i = \frac{1}{N_i} \sum_{(x,y)} BL_i(x, y) \tag{4}$$

$$\sigma_i = \sqrt{\frac{\sum_{(x,y)} (BL_i(x, y) - \mu_i)^2}{N_i}} \tag{5}$$

where BL_i is the i-th block of an SCI in the C_b or C_r channel, (x, y) is the coordinate location for the pixels in the i-th block, and N_i is the total number of pixels in the i-th block. Each image block is described by a 4-bin semi-global feature as follows:

$$feat_{i_{semi-global}} = [\mu_{i_{C_b}}, \mu_{i_{C_r}}, \sigma_{i_{C_b}}, \sigma_{i_{C_r}}] \tag{6}$$

where $\mu_{i_{C_b}}$ and $\mu_{i_{C_r}}$ are respectively the mean value of the i-th image block in the C_b and C_r channels, and $\sigma_{i_{C_b}}$ and $\sigma_{i_{C_r}}$ are respectively the standard deviation of the i-th image block in the C_b and C_r channels. Similarly, we use the same statistical moments of the whole image from all three channels (i.e., Y , C_b , and C_r) to represent the chrominance features of an image at the global scale as follows:

$$feat_{global} = [\mu_Y, \mu_{C_b}, \mu_{C_r}, \sigma_Y, \sigma_{C_b}, \sigma_{C_r}] \tag{7}$$

where μ_Y, μ_{C_b}, and μ_{C_r} are respectively the mean value of an image in the Y, C_b, and C_r channels, and σ_Y, σ_{C_b}, and σ_{C_r} are respectively the standard deviation of an image in the Y, C_b, and C_r channels.

It should be noted that both semi-global and global features are normalized by dividing their mean and standard deviation statistics by the maximum intensity value (i.e., 255).

The final feature vector of a distorted SCI is obtained by concatenating the 10-bin local and 4-bin semi-global feature vectors of all 16 blocks with the 6-bin global feature vector as follows:

$$feature = [feat_1, feat_2, ..., feat_{16}, feat_{global}] \tag{8}$$

where $feat_i$ is the concatenated local and semi-global feature vector $[feat_{i_{local}}, feat_{i_{semi-global}}]$ of the i-th image block.

As a result, we obtain a 230-dimensional feature vector to describe a distorted SCI from its corresponding 16 blocks, where 160 values are for local features (i.e., 80 (16×5) values are for normalized EHD_A and 80 (16×5) values are for normalized EHD_B), 64 (16×4) values are for semi-global features, and 6 values are for global features from the entire image.

To enhance the edge patterns with small magnitudes and ensure features with high and low values would contribute equally without any biases, we use the square root operation to normalize the final 230-dimensional feature vector [44] by:

$$F = [\sqrt{f_1}, \sqrt{f_2}, ..., \sqrt{f_i}, ..., \sqrt{f_{230}}] \tag{9}$$

where f_i is a value in the final 230-dimensional feature vector.

2.3 Quality regression

After feature extraction, similar to [12, 9, 45], we use SVR provided by the LIBSVM package [46] to map the extracted 230-dimensional quality-aware features F to their corresponding subjective quality scores. We choose SVR in our method since it automatically learns to differentiate the effect of various distortions and gives us the flexibility to define acceptable errors and find an appropriate hyperplane to fit the data. We use the RBF with *gamma = 1, cost = 128*, and *epsilon = 1* since these are optimal values used in NRLT [12] and PQSC [9]. To ensure a fair evaluation, we randomly divide each dataset into training and testing subsets 1000 times and use 80% of the data for training SVR and the rest for testing. The median of the 1000 results is reported as the overall performance.

3 Experimental Results

We evaluate the IQA performance of the proposed EHDSM method on three publicly available SCI datasets and compare its performance with seven state-of-the-art BSCIQA methods using three common evaluation metrics. Our extensive experiments demonstrate the superiority of the proposed EHDSM method in terms of accuracy, monotonicity, consistency, and computational run-time.

3.1 The testing datasets

We evaluate the proposed EHDSM method by conducting experiments on three publicly available SCI datasets: screen content image quality assessment database (SIQAD) [11], screen content image database (SCID) [47], and quality assessment of compressed SCI (QACS) [48]. SIQAD contains 20 reference SCIs and 980 distorted images degraded by seven types of distortions, including Gaussian noise (GN), Gaussian blur (GB), motion blur (MB), contrast change (CC), JPEG, JPEG 2000 (J2K), and layer segmentation-based coding (LSC), each of which includes seven different levels. SCID contains 40 reference SCIs with a resolution of 1280×720 and 1800 distorted images degraded by nine types of distortions, including GN, GB, MB, CC, JPEG, J2K, color saturation change (CSC), high efficiency video coding

Table 1: Comparison of the performance of the proposed EHDSM using different block sizes on SIQAD, SCID, and QACS datasets.

Dataset	Criteria	Size: 2×2	Size: 4×4	Size: 6×6	Size: 8×8
SIQAD	PLCC	0.9151	**0.9157**	0.8925	0.8576
	SRCC	0.9034	**0.9073**	0.8869	0.8521
	RMSE	5.7579	**5.75261**	6.4426	7.3066
SCID	PLCC	0.9192	**0.9371**	0.9101	0.8700
	SRCC	0.9162	**0.9367**	0.9095	0.8646
	RMSE	5.5689	**4.9468**	5.8627	6.9867
QACS	PLCC	**0.9461**	0.9442	0.9296	0.9168
	SRCC	**0.9393**	0.9376	0.9205	0.9091
	RMSE	**0.7139**	0.7268	0.8140	0.8804

(HEVC) and its extension for screen content coding (SCC) indicated as HEVC-SCC, and color quantization with dithering (CQD), each of which includes five different levels. The QACS dataset contains 24 pristine reference SCIs of a resolution of 2560×144, 1920×1080, and 1280×720, covering wide application scenarios and 492 distorted images degraded by two types of distortions, including HEVC and SCC.

3.2 Evaluation metrics and parameter setting

We use three common metrics [49] including Pearson's linear correlation coefficient (PLCC), Spearman's rank-order correlation coefficient (SRCC), and root mean squared error (RMSE), to measure the prediction accuracy, monotonicity, and consistency [50, 10] of the eight compared BSCIQA methods, respectively. PLCC represents the linear correlation between the objective and subjective scores. SRCC represents the rank-order correlation between the objective and subjective scores. RMSE measures the deviation between the objective and subjective scores. A higher PLCC value (close to 1), a higher SRCC value (close to 1), and a smaller RMSE value (close to 0) indicate better performance (i.e., high degree of consistency between the objective quality evaluation algorithm and subjective ratings). Following the same procedures used in [9, 10] and [49], we apply a non-linear logistic regression with five parameters to remove the non-linearity of objective quality predictions before calculating the above three evaluation metrics as follows:

$$Z_i = \kappa_1 \{ \frac{1}{2} - \frac{1}{1 + \exp\left[\kappa_2 (s_i - \kappa_3)\right]} \} + \kappa_4 s_i + \kappa_5 \quad (10)$$

where s_i is the perceived quality score of the i-th distorted SCI computed by an IQA model, Z_i is the corresponding mapped predicted objective score, and κ_1, κ_2, κ_3, κ_4, and κ_5 are the five parameters that are fitted during the curve fitting process.

The proposed EHDSM method has two main parameters that need to be properly set to extract the final normalized quality-aware features. The first and most important parameter is the block size, which allows the proposed method to extract a fixed-size feature vector (e.g., 230 values when the block size is 4×4) from an image of any resolution. The second parameter is the threshold value required to categorize a patch as edge or non-edge content. In the following, we investigate the sensitivity of the proposed method to the parameter setting by varying one parameter and fixing the other one each time.

We evaluate the performance of the proposed EHDSM method with different block sizes to empirically set the optimal block size with respect to the obtained accuracy and computational run-time. Since a larger block size leads to more blocks in row and column directions, it results in a larger feature vector and longer training and testing time of the SVR model. For example, for a block size of 8×8, we obtain a 902-dimensional feature vector to describe a distorted SCI from its corresponding 64 blocks, where 640 values are for local features (i.e., 320 (64×5) values are for normalized EHD_A and 320 (64×5) values are for normalized EHD_B), 256 (64×4) values are for semi-global features, and 6 values are for global features from the entire image. Its feature length is almost 4 times of the length of the final normalized quality-aware feature obtained from 4×4 blocks. So, we limit the largest block size to be 8×8 and evaluate the performance of the proposed EHDSM method with four different block sizes of 2×2, 4×4, 6×6, and 8×8. Table 1 lists the median PLCC, SRCC, and RMSE values obtained by evaluating the SVR model 1000 times on the SIQAD, SCID, and QACS datasets using the proposed EHDSM method with each of four block sizes and the threshold value of 16. It demonstrates that the two smallest block sizes (i.e., 2×2 and 4×4) achieve the best comparable results in terms of three metrics on all three datasets. Specifically, the block size of 4×4 leads to the best performance on SIQAD and SCID datasets, and the block size of 2×2 leads to the best performance on the QACS dataset. We empirically choose the block size of 4×4 in the proposed method due to its superior performance on SIQAD and SCID datasets. Since SCID contains the largest number of reference and distorted images in a mid-resolution (i.e., 1280×720) among the three datasets, it makes the inference of the trained machine learning model more reliable and generalizable and supports the compatibility and capability of the block size of 4×4 to represent SCIs in IQA tasks.

In addition, according to Table 1, while the block size

4×4 yields better results across the three datasets, the performance of the proposed method drops on all datasets when the block size increases from 4 to 6 and increases from 6 to 8. This is mainly because increasing the block size significantly increases the number of blocks, which subsequently leads to the image being represented by a vast variety of contents. For example, when the block size is equal to 4, an image would be represented using 16 blocks with different content variations. When increasing the block size to 6 or 8, an image would be represented using 36 (6×6) or 64 (8×8) blocks, respectively. Moreover, it also increases the dimension of local features (36×10 and 64×10 for block sizes of 6 and 8, respectively) and semi-global features (36×4 and 64×4 for block sizes of 6 and 8, respectively). Therefore, considering a limited number of images in SCI datasets, learning an accurate mapping from the extracted high-dimensional features to the subjective scores is quite challenging and it decreases the performance of the proposed method.

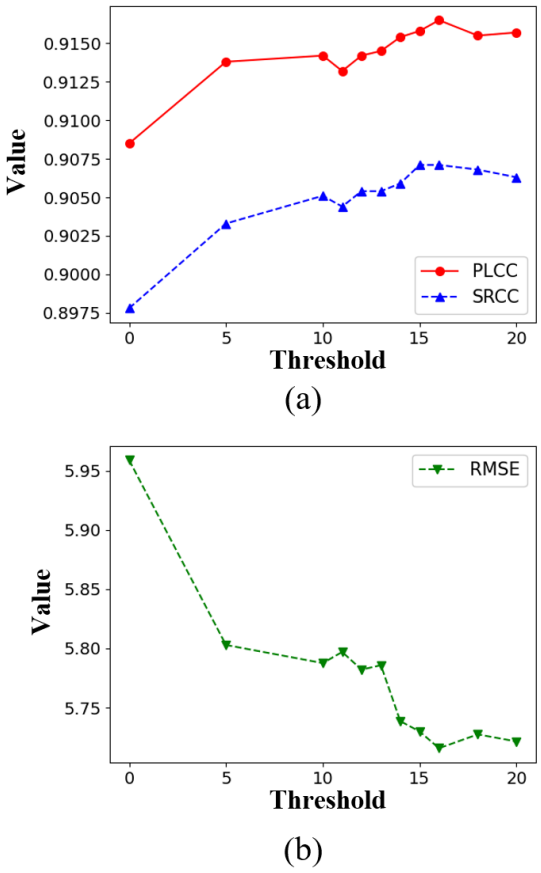

Figure 4: Plots of different metrics obtained by the proposed EHDSM method using different threshold values on the SIQAD dataset. (a) PLCC and SRCC values; (b) RMSE values.

We evaluate the performance of the proposed EHDSM method with different threshold values to empirically set the optimal threshold with respect to the obtained accuracy. Patches with useful and effective content information tend to have high PLCC and SRCC and low RMSE values. However, patches with noisy information tend to be incorrectly recognized as useful edges, which may lead to inaccuracy in IQA results. As a result, a relatively small threshold is preferred to not only select useful and effective content information but also eliminate the noise impacts. Since it is crucial to capture and balance the effects of different content types and noise levels, we experiment with 11 thresholds (i.e., 0, 5, 10, 11, 12, 13, 14, 15, 16, 18, and 20) and select an optimal smallest threshold that leads to the highest SRCC and PLCC and lowest RMSE values with smallest impact from noises. Figure 4. (a) and (b) plot median PLCC, SRCC, and RMSE values obtained by evaluating the SVR model 1000 times on the SIQAD dataset using the proposed EHDSM method with each of 11 thresholds and the block size of 4×4. It shows that EHDSM achieves improving IQA performance when the threshold increases. However, the highest IQA performance in terms of the PLCC, SRCC, and RMSE metrics is achieved when the threshold is 16. Further increasing the threshold leads to decreasing IQA performance. Based on the results shown in Figure 4, we empirically choose a threshold of 16 in the proposed method since it is the smallest threshold that captures and balances the effects of different content types and noise levels. It is noteworthy that there is not a huge performance gap between different threshold values. This indicates that the proposed method is less sensitive to this parameter and can accurately capture the edge distribution of both content and noise in the testing stage.

3.3 Performance comparison

Table 2 compares the proposed EHDSM method with four NSI BIQA methods, including NIQE [5], ILNIQE [6], BRISQUE [7], and GM-LOG [8] and seven state-of-the-art BSCIQA methods including BQMS [2], SIQE [3], NRLT [12], OSM [13], HRFF [14], PQSC [9], and FVC [15] on three publicly available datasets SIQAD, SCID, and QACS in terms of three metrics PLCC, SRCC, and RMSE. The source codes of these methods, except for OSM, HRFF, and FVC, are obtained from their author's websites. To make a fair comparison, we choose the same sets of training and testing images and perform the same testing procedure for all compared methods that have source codes. Performance results of OSM, HRFF, and FVC are copied from their respective published papers. The best value of each of the three evaluation metrics is highlighted in bold. Table 2 clearly shows that the proposed EHDSM method yields the highest overall PLCC and SRCC values and the lowest overall RMSE values on all three datasets. It also significantly outperforms the other methods on the SCID dataset that contains chrominance distortions (e.g., color saturation change and color quantization with dithering) since it is the only method that extracts semi-global and global chrominance features to quantify and characterize the chrominance information of the distorted SCI. In addition, the top four BSCIQA methods, namely, EHDSM, PQSC, NRLT, and FVC, outperform all four NSI BIQA methods on three SCI datasets. All BSCIQA methods outperform two NSI BIQA methods, namely, NIQE and ILNIQE, on three SCI datasets.

Table 3 lists the scores of three metrics PLCC, SRCC, and RMSE when employing eight variants of the proposed EHDSM method, namely, EHD_A only variant 1, EHD_B only variant 2, local (EHD_A+EHD_B) only variant 3, semi-

Table 2: Comparison of PLCC, SRCC, and RMSE obtained by applying four BIQA methods, seven state-of-the-art BSCIQA, and the proposed EHDSM method on each of three publicly available datasets (i.e., SIQAD, SCID, and QACS).

Dataset	Criteria	NIQE	ILNIQE	BRISQUE	GM-LOG	BQMS	SIQE	NRLT	OSM	HRFF	PQSC	FVC	**EHDSM**
SIQAD	PLCC	0.3749	0.3854	0.8113	0.7608	0.8419	0.8225	0.9137	0.8306	0.852	0.9101	0.9014	**0.9157**
	SRCC	0.3568	0.3212	0.7749	0.7035	0.8348	0.8059	0.9024	0.8007	0.8320	0.8997	0.8915	**0.9073**
	RMSE	13.1520	13.2085	8.2565	9.2530	7.6760	8.1286	5.7879	7.9331	7.4150	5.8885	6.1684	**5.7261**
SCID	PLCC	0.3904	0.4079	0.7696	0.7883	0.7592	0.7208	0.8648	-	-	0.9142	0.8681	**0.9371**
	SRCC	0.3712	0.3546	0.7448	0.7619	0.7416	0.7150	0.8454	-	-	0.9111	0.8550	**0.9367**
	RMSE	12.9827	12.8425	9.0143	8.6754	9.2087	9.8060	7.0881	-	-	5.7450	7.0170	**4.9468**
QACS	PLCC	0.4240	0.2374	0.8421	0.9002	0.8622	0.8821	0.9004	0.7068	-	0.9354	0.9239	**0.9442**
	SRCC	0.3701	0.2603	0.8201	0.8903	0.8557	0.8708	0.8926	0.6804	-	0.9275	0.9198	**0.9376**
	RMSE	1.7091	1.8395	1.0959	0.9656	1.1126	1.0404	0.9594	1.5301	-	0.7804	0.8434	**0.7268**

Table 3: Comparison of PLCC, SRCC, and RMSE obtained by applying the proposed EHDSM method and its eight variants on each of three publicly available datasets (i.e., SIQAD, SCID, and QACS).

Dataset	Criteria	EHD_A	EHD_B	local	SG	global	local+SG	local+global	SG+global	**EHDSM**
SIQAD	PLCC	0.9068	0.8948	0.9152	0.5075	0.2440	0.9145	**0.9175**	0.5980	0.9157
	SRCC	0.8966	0.8778	0.9057	0.4756	0.2144	0.9055	**0.9076**	0.5705	0.9073
	RMSE	5.9912	6.3701	5.7397	12.2986	13.8178	5.7539	**5.6804**	11.4196	5.7261
SCID	PLCC	0.8475	0.8625	0.8721	0.5135	0.2955	0.9338	0.9265	0.5750	**0.9371**
	SRCC	0.8189	0.8342	0.8462	0.4764	0.2463	0.9330	0.9251	0.5505	**0.9367**
	RMSE	7.4939	7.1490	6.9110	12.1198	13.5104	5.0587	5.3350	11.5781	**4.9468**
QACS	PLCC	**0.9469**	0.935	0.9430	0.8282	0.3460	0.9433	0.9433	0.8395	0.9442
	SRCC	**0.9400**	0.9269	0.9355	0.8136	0.2813	0.9373	0.9370	0.8246	0.9376
	RMSE	**0.7113**	0.7818	0.7332	1.2338	2.0668	0.7281	0.7335	1.2005	0.7268

global (SG) only variant 4, global only variant 5, local+semi-global (local+SG) only variant 6, local+global only variant 7, and semi-global+global (SG+global) only variant 8, on three publicly available datasets SIQAD, SCID, and QACS. Table 3 shows that content descriptor-based edge histograms capturing edge variations are the most important features in IQA since both variant 1 (i.e., EHD_A only) and variant 2 (i.e., EHD_B only) achieve high performance for three metrics on three datasets. Variant 3 (i.e., Local (EHD_A+EHD_B) only) outperforms variant 1 and variant 2 on SIQAD and SCID datasets and outperforms variant 2 on the QACS dataset. In other words, both EHD_A and EHD_B features are necessary and contribute to the superiority of the proposed method. On the other hand, variant 4 (i.e., SG only), variant 5 (i.e., global only), and variant 8 (i.e., SG+global only) do not achieve high performance in terms of three metrics on three datasets since they only capture chrominance distortions. However, local+SG only variant 6 (combining local features with SG features) and local+global only variant 7 (combining local features with global features) tend to improve the variant 3 (i.e., local only) on all three datasets, especially they lead to significant improvement over variant 3 on the SCID since both SG and global features capture chrominance distortions in images from the SCID dataset. These experimental results indicate the influence of each feature is significant and all the local, semi-global, and global features are needed to correctly evaluate the quality of a distorted SCI. The results in Table 3 also demonstrate that the best performance on the SIQAD, SCID, and QACS datasets is achieved by variant 7, the proposed EHDSM method that combines local, semi-global, and global features, and variant 1, respectively. Since the SIQAD dataset does not contain any chrominance distortions, combining the semi-global chrominance features with variant 7 slightly drops the performance. Moreover, variant 1 leads to the best performance on the QACS dataset since this dataset only includes compression-related distortions, which degrades the structural information of an image rather than its statistical moments. Similarly, fusing the semi-global and global features with the local features slightly decreases the performance. However, the performance improvement is significant on the SCID dataset when fusing all features, which allows the model to assess the quality of degraded images under chrominance distortions. As a result, we propose to utilize all three feature variations to represent the SCIs.

3.4 Performance comparison of distortion types

We conduct a set of performance experiments on SCID, which contains more SCIs and a wide range of distortions, to comprehensively analyze the effectiveness of the proposed EHDSM method in terms of PLCC, SRCC, and RMSE under distortions. Table 4 presents the PLCC, SRCC, and RMSE values obtained by the proposed EHDSM and five state-of-

Table 4: PLCC, SRCC, and RMSE results of the proposed method and five peer methods on distorted SCIs in SCID.

Criteria	Distortion	BQMS	SIQE	NRLT	PQSC	FVC	EHDSM
PLCC	GN	0.8395	0.6989	**0.9753**	0.9611	0.959	**0.9691**
	GB	0.7464	0.8336	0.9264	**0.9325**	**0.967**	0.9222
	MB	0.7963	0.9325	0.9290	**0.9541**	0.939	**0.9504**
	CC	0.6733	0.4215	0.7766	0.7605	**0.936**	**0.8651**
	JPEG	0.8679	0.6932	0.9366	**0.9425**	0.93	**0.9574**
	J2K	0.7820	0.8235	0.9392	**0.9506**	0.927	**0.9564**
	CSC	0.3099	0.2938	0.2695	**0.7791**	-	**0.9492**
	HEVC-SCC	0.6434	0.6431	0.7857	**0.9012**	-	**0.9220**
	CQD	0.6512	0.6949	0.8956	0.8512	**0.941**	**0.9182**
SRCC	GN	0.8216	0.6781	**0.9626**	0.9451	0.953	**0.9539**
	GB	0.6944	0.8120	0.9161	**0.9253**	**0.953**	0.9142
	MB	0.7633	0.9116	0.9145	**0.9386**	0.933	**0.9323**
	CC	0.4821	0.3145	0.5969	0.6296	**0.930**	**0.7648**
	JPEG	0.8386	0.6527	0.9255	**0.9248**	0.924	**0.9348**
	J2K	0.7392	0.7801	0.9045	0.9124	**0.923**	**0.9277**
	CSC	-0.1216	-0.1009	-0.0248	**0.7602**	-	**0.9403**
	HEVC-SCC	0.5283	0.5338	0.7068	**0.8740**	-	**0.9094**
	CQD	0.6185	0.6515	0.8407	0.8298	**0.936**	**0.8995**
RMSE	GN	6.6764	8.7800	**2.7457**	3.4321	3.516	**3.0325**
	GB	6.8995	5.7087	3.9077	**3.7459**	**3.011**	4.0108
	MB	6.4919	3.8802	3.9726	**3.2438**	4.137	**3.3416**
	CC	6.4058	7.9356	5.5531	5.7286	**4.253**	**4.4064**
	JPEG	7.3384	10.6357	5.1646	4.9591	**3.86**	**4.2690**
	J2K	9.6903	8.9020	5.4034	4.8695	**3.752**	**4.5990**
	CSC	9.2109	9.2254	9.2347	**6.0070**	-	**3.0655**
	HEVC-SCC	10.4661	10.3733	8.4252	**5.9319**	-	**5.3055**
	CQD	9.5168	8.9806	5.5814	6.6143	**3.893**	**5.0076**

the-art peer methods on nine distortions on the SCIs in SCID, respectively. Four of the five compared methods, namely, BQMS, SIQE, NRLT, and PQSC, provide their source codes online so we can run the same experiments with the same setups to obtain their results under distortions. One of the five compared methods (i.e., FVC) provides the PLCC, SRCC, and RMSE values on seven distortions and does not provide their results on CSC and HEVC-SCC distortions. So we use "-" to reflect these missing results in Table 4. We highlight the best- and second-best performance values in red and blue, respectively. The results in Table 4 clearly show that the proposed EHDSM method more precisely assesses the quality of SCIs under various distortions due to its utilization of combined content and chrominance-based features. Specifically, it overally performs the best under JPEG, J2K, CSC, and HEVC-SCC distortions and performs the second-best under GN, MB, CC, and CQD distortions. Compared with the five BSCIQA methods, EHDSM has superior performance on SCID for eight of nine distortion types (i.e., best performance on four distortions and second-best results on four distortions). It ranks the 4^{th} on SCID for GB distortion mainly because the blurring operation directly affects the edges of an image. Since the proposed EHDSM method uses the local edge descriptor to represent an image's content, it is challenging to distinguish the degradation on the image's content resulted from GB distortion. However, we still achieve a comparable result on the GB distortion due to the fusion of multiple complementary features. In summary, our experimental results clearly demonstrate the generalizability and stability of the proposed method regarding various distortions.

3.5 Analysis of feature sets

Inspired by PQSC [9], we employ the t-distributed stochastic neighbor embedding (t-SNE) method [51] to validate the effectiveness of the proposed feature space (i.e., local, semi-global, and global features) in assessing quality degradations resulting from different distortions. t-SNE is an algorithm used for dimension reduction. It models each high-dimensional sample using a two-dimensional or three-dimensional point in a way that similar samples are mapped to similar points close to each other and dissimilar samples are mapped to distant points. As a result, it is capable of preserving the global information of the actual feature space [51] and visualizing the samples using their corresponding

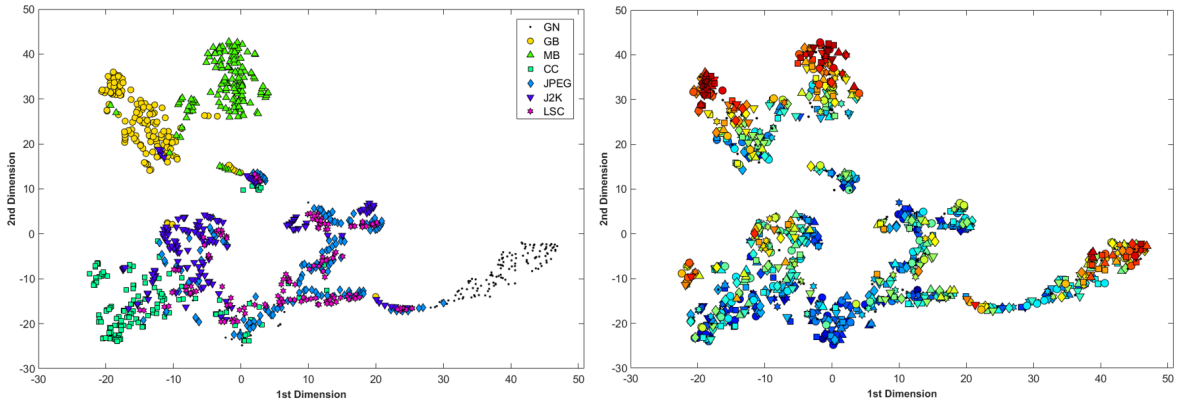

Figure 5: t-SNE scatter plots of the proposed feature representation on SIQAD: data points are colored by distortion types (left plot) and DMOS scores (right plot).

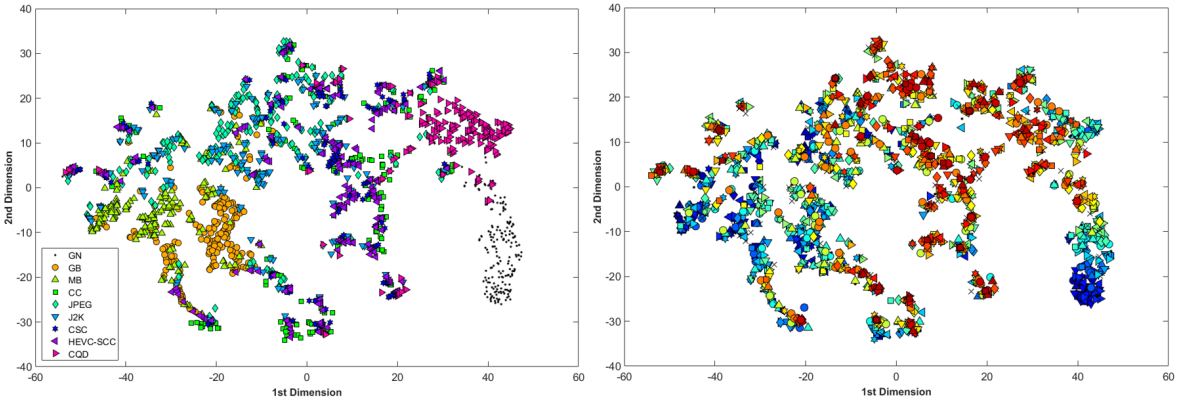

Figure 6: t-SNE scatter plots of the proposed feature representation on SCID: data points are colored by distortion types (left plot) and MOS scores (right plot).

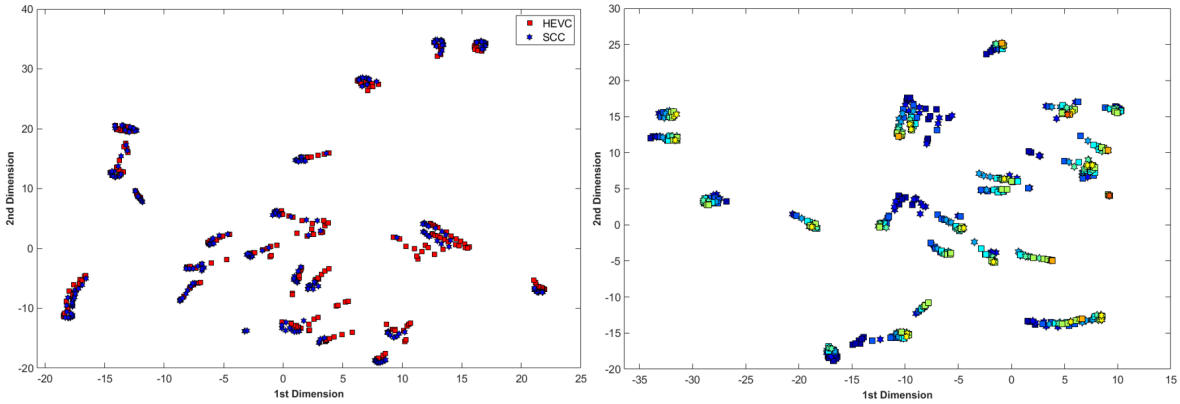

Figure 7: t-SNE scatter plots of the proposed feature representation on QACS: data points are colored by distortion types (left plot) and MOS scores (right plot).

feature maps in the reduced dimension. Here, we employ the t-SNE method to visually analyze the effectiveness of the proposed features in two dimensions since it is impossible to visualize the high-dimensional features of size 230 extracted by the proposed EHDSM method.

Figures 5, 6, and 7, show two scatter plots on SIQAD, SCID, and QACS, respectively. The one on the leftside is the t-SNE scatter plot of the extracted quality-aware features of the proposed EHDSM method in two dimensions colored by different distortion types. The one on the rightside is the same t-SNE scatter plot shown in leftside, in which data points are colored by their respective MOSs or DMOSs. In all six scatter plots, each data point represents an SCI, and different colors represent distortion types or distortion levels. The scatter plots on the leftside of Figures 5 through 7 show that the proposed features are able to cluster the distortions with similar artifacts. For example, the left scatter plot in Figure 5 shows that SCIs from SIQAD, which are distorted by GN, GB, and MB, are grouped around their corresponding clusters and are almost separated from other clusters. The left scatter plot in Figure 6 shows that the SCIs from SCID, which are distorted by GN and GB, are grouped around their corresponding clusters and are almost separated from other clusters. Since GN and CQD distortions add some noise to the reference images, the quality-aware features of their distorted SCIs are well separated from others and are grouped quite close together. Since the remaining distortions (i.e., CC, JPEG, J2K, CSC, HEVC-SCC, and LSC) introduce the same artifacts (e.g., compression and blurring effects), their distorted SCIs tend to be clustered together in the proposed feature space. The scatterplots on the rightside of Figures 5 through 7 show that distorted SCI images of almost the same MOS or DMOS values are grouped together. For example, red or blue points represent SCI images whose subjective scores are close to each other. Overall, the efficiency of the proposed feature sets is clearly demonstrated in the distinction of both the distortion types and levels.

3.6 Cross-dataset evaluation

Cross-dataset evaluation is commonly used to validate the generalizability of the proposed method. Considering the number of images and common distortion types in publicly available datasets, we select the SIQAD and SCID datasets as training and testing datasets. Specifically, we train the proposed and compared peer methods on one dataset and test their performance over the other dataset. This process is done with six types of distortion (GN, FB, MB, CC, JPEG, and J2K) that are common in both datasets.

Table 5 compares experimental results of the cross-dataset evaluation of five BSCIQA methods. The column of training with SIQAD (Table 5. (a)) shows the testing results of each compared method on SCID when it is trained over six selected distortions of SCIs in SIQAD. The column of training with SCID (Table 5. (b)) shows the testing results of each compared method on SIQAD when it is trained over six selected distortions of SCIs in SCID. The best and second-best results are highlighted in red and blue, respectively. Previous studies [15, 23, 52] report cross-dataset validation of their proposed methods over the common distortions in two datasets. Here, we report the cross-dataset validation results of the proposed EHDSM and four state-of-the-art BSCIQA methods to not only validate the generalizability of each compared method but also compare their generalizability.

Based on the experimental results shown in Table 5, we observe the following: 1) All compared methods achieve better cross-dataset validation results when they are trained over six selected distortions of SCIs in SCID. The vast size and complexity of SCID may make the machine learning model more generalizable and stable, which leads to better cross-dataset performance on SIQAD. 2) The cross-dataset performance of almost all BSCIQA methods is worse than their in-dataset performance, mainly due to the diverse contents of the SCIs in training and testing datasets. 3) The NRLT achieves the best cross-dataset performance, and the proposed EHDSM method achieves the second-best performance. The block-based feature extraction strategy of the proposed method may make the model more dependent on imagery contents compared with NRLT. However, this could be tackled by utilizing diverse image contents in the learning process and increasing the training samples.

3.7 Influence of training sizes

To evaluate the influence of training sizes on the IQA performance, we use five training sizes (i.e., 40%, 50%, 60%, 70%, and 80% of SCIs in SIQAD) to train the corresponding machine learning model of the proposed EHDSM method and four state-of-the-art BSCIQA methods and test the performance of these compared methods on the remaining SCIs in SIQAD. Figure 8 presents the SRCC values of the five compared methods by training their respective machine learning models on 392, 490, 588, 686, and 784 SCIs in SIQAD. It is clear that the performance of all methods improves when the training size increases. The proposed EHDSM method, NRLT, and PQSC achieve the top three SRCC values under each of the five trainings. They also achieve satisfactory SRCC values of above 0.85 even with a low training size of 392 images. Both EHDSM and NRLT become steady with growth in training sizes. For example, SRCC value of NRLT improves 0.79%, and the SRCC value of EHDSM improves 1.31% when using 588 images (60% training size) and 784 (80% training size) images to train the respective machine learning model. The two properties of the top three methods, namely, satisfactory testing performance with a small training set and the steadiness with growth in training size, are attractive in practice due to the lack of SCI training images.

3.8 Computational time

To evaluate the computational time of the BSCIQA methods, we apply each method to all distorted SCI images in a dataset to calculate its average run-time (i.e., divide the total run-time by the total number of images in the dataset). Figure 9 compares the average run-time of the proposed EHDSM method and four state-of-the-art BSCIQA methods on three datasets, SIQAD, SCID, and QACS. It clearly shows that our proposed EHDSM method has the fastest average run-time, which is a significantly low computational run-time of less than 2.5 seconds. Specifically, its run-time is less than one-third of the run-time of the second-fastest method (i.e., NRLT) on all three datasets. It is also interesting to note that the average run-time is not correlated with the feature-length.

Table 5: Cross-dataset evaluation results of the proposed and four BSCIQA methods for SCIs of six types of distortions in SIQAD and SCID.

	(a) Training with SIQAD			(b) Training with SCID		
	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE
BQMS	0.3793	0.3730	12.3448	0.5726	0.5595	12.0440
SIQE	0.3741	0.3649	12.3735	0.3747	0.3850	13.6205
NRLT	**0.6226**	**0.6229**	**10.4411**	**0.7851**	**0.7742**	**9.0990**
PQSC	0.5469	0.5494	11.1700	0.6912	0.6466	10.6160
EHDSM	**0.5990**	**0.5841**	**10.6834**	**0.7061**	**0.6503**	**10.4023**

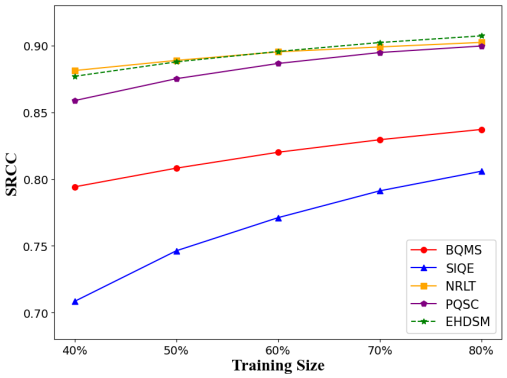

Figure 8: SRCC results of the proposed and four BSCIQA methods for SCIs in SIQAD using different training sizes.

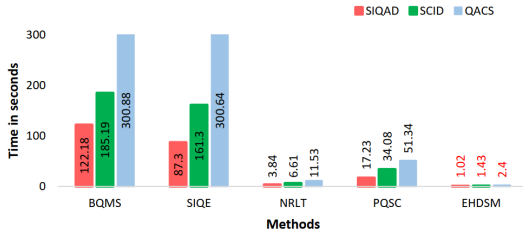

Figure 9: Comparison of average run-time of the proposed and four BSCIQA methods on three SCI datasets

For example, the length of the features for each SCI extracted by the proposed BQMS, SIQE, NRLT, PQSC, and our proposed EHDSM method is 13, 15, 270, 520, and 230, respectively. Even though BQMS and SIQE have the shortest feature length, their average run-times are the slowest due to their complicated computational cost to extract the features. In summary, the proposed EHDSM method is more effective in terms of prediction accuracy, monotonicity, and consistency and more efficient in terms of run-time compared to four recently proposed state-of-the-art BSCIQA methods.

4 Conclusion and Future Work

We propose a BSCIQA method that extracts edge and chrominance features in local, semi-global, and global scales using the EHDSM from a distorted SCI in the YC_bC_r color space. Our major contributions are as follows: 1) Extracting quality-aware features at three scales (i.e., local, semi-global, and global scales) to effectively capture the effects of different distortion types and levels of SCIs. 2) Employing EHDs at local scales to describe the content of an image to accurately represent the characteristics of SCIs. 3) Using statistical moments at both semi-global and global scales to effectively quantify the content and color degradations of a distorted SCI. Experimental results on three common datasets, including SIQAD, SCID, and QACS, demonstrate the superiority of the proposed EHDSM method over its eight variant methods, four NSI BIQA methods, and seven state-of-the-art BSCIQA methods in terms of three common metrics (i.e., PLCC, SRCC, and RMSE) under all distortions and under each specific distortion, visualization of features in a reduced dimension, cross-dataset evaluation, training sizes, and the average run-time. The novelty of our proposed EHDSM method includes employing EHD to capture the frequency and directionality of edges at local scales (i.e., in image patches within each image block) and employing first- and second-order statistical moments to capture chrominance information at semi-global scales (i.e., in image blocks) and at global scales (i.e., in the entire image). The integration of local, semi-global, and global features sufficiently perceives the quality of SCIs.

Currently, the extracted features are stacked together without any further analysis. Therefore, the performance of the proposed EHDSM method drops when assessing the quality of images that are less seen in the training stage. In the future, we would like to carry out research on investigating the pre- and post-assessment fusion strategies to distinguish the best strategy that yields the highest performance. Similar to our proposed approach, we intend to analyze the possible combination of features such as addition, multiplication, concatenation, weighted fusion, etc. before making the final quality evaluation (i.e. pre-assessment fusion). Moreover, we plan to study the effect of fusing the assessed qualities from the individual feature sets (i.e., post-assessment fusion).

Furthermore, we are interested in exploring more accurate color information descriptors. As screen content images consist of both pictorial and textural content, the impact of color degradation varies on each particular content. Thus,

we plan to devise a new color information descriptor based on the content of the image.

Acknowledgements- No funding was received for conducting this study. The authors have no relevant financial or non-financial interests to disclose.

Data availability- Data sharing is not applicable to this article as no new datasets were generated or analyzed during the current study.

Statements and Declarations

Conflict of interest- The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Wang, K. Gu, S. Ma, and W. Gao, "Joint chroma downsampling and upsampling for screen content image," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, pp. 1595–609, September 2016.

[2] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Learning a blind quality evaluation engine of screen content images," *Neurocomputing*, vol. 196, p. 140–149, July 2016.

[3] K. Gu, J. Zhou, J. F. Qiao, G. Zhai, W. Lin, and A. C. Bovik, "No-reference quality assessment of screen content pictures," *IEEE Transactions on Image Processing*, vol. 26, pp. 4005 – 4018, August 2017.

[4] K. G. et al., "Saliency-guided quality assessment of screen content images," *IEEE Transactions on Multimedia*, vol. 18, p. 1098–1110, June 2016.

[5] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, pp. 209–212, March 2013.

[6] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Transactions on Image Processing*, vol. 24, p. 2579–2591, August 2015.

[7] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, pp. 4695–4708, December 2012.

[8] 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 Transactions on Image Processing*, vol. 23, p. 4850–4862, November 2014.

[9] Y. Fang, R. Du, Y.Zuo, W. Wen, and L. Li, "Perceptual quality assessment for screen content images by spatial continuity," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, pp. 4050–4063, November 2020.

[10] H. F. Tolie and M. R. Faraji, "Screen content image quality assessment using distortion-based directional edge and gradient similarity maps," *Signal Processing: Image Communication*, vol. 101, February 2022.

[11] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Transactions on Image Processing*, vol. 24, p. 4408–4421, August 2015.

[12] Y. Fang, J. Yan, L. Li, J. Wu, and W. Lin, "No reference quality assessment for screen content images with both local and global feature representation," *IEEE Transactions on Image Processing*, vol. 27, pp. 1600–1610, April 2018.

[13] N. Lu and G. Li, "Blind quality assessment for screen content images by orientation selectivity mechanism," *Signal Processing*, vol. 145, pp. 225–232, 2018.

[14] L. Zheng, L. Shen, J. Chen, P. An, and J. Luo, "No-reference quality assessment for screen content images based on hybrid region features fusion," *IEEE Transactions on Multimedia*, vol. 21, pp. 2057–2070, August 2019.

[15] Y. Bai, Z. Zhu, C. Zhu, and Y. Wang, "Blind image quality assessment of screen content images via fisher vector coding," *IEEE Access (Early Access)*, January 2022.

[16] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in freeenergy principle," *IEEE Transactions on Image Processing*, vol. 21, pp. 41–52, January 2012.

[17] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *EEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, July 2002.

[18] J. Wu, W. Lin, G. Shi, Y. Zhang, W. Dong, and Z. Chen, "Visual orientation selectivity based structure description," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4602–4613, 2015.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, April 2004.

[20] H. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[21] L. C. H.R. Sheikh, Z.Wang and A. Bovik, "Live image quality assessment database release 2," *http://live.ece.utexas.edu/research/quality*, 2005.

[22] Y. Bai, M. Yu, Q. Jiang, G. Jiang, and Z. Zhu, "Learning content-specific codebooks for blind quality assessment of screen content images," *Signal Processing*, vol. 161, pp. 248–258, 2019.

[23] Y. Bai, Z. Zhu, G. Jiang, and H. Sun, "Blind quality assessment of screen content images via macro-micro modeling of tensor domain dictionary," *IEEE Transactions on Multimedia*, vol. 161, pp. 248–258, 2020.

[24] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4311–4322, November 2006.

[25] J. Chen, L. Shen, L. Zheng, and X. Jiang, "Naturalization module in neural networks for screen content image quality assessment," *IEEE Signal Processing Letters*, vol. 25, p. 1685–1689, November 2018.

[26] L. Zheng, L. Shen, J. Chen, P. An, and J. Luo, "No reference quality assessment for screen content images using stacked autoencoders in pictorial and textual regions," *IEEE Transactions on Cybernetics*, pp. 1–13, October 2020.

[27] X. Jiang, L. Shen, G. Feng, L. Yu, and P. An, "An optimized cnn-based quality assessment model for screen content image," *Signal Processing: Image Communication*, vol. 94, May 2021.

[28] R. Wang, H. Yang, Z. Pan, B. Huang, and G. Hou, "Screen content image quality assessment with edge features in gradient domain," *IEEE Access.*, vol. 7, pp. 4818–4831, 2019.

[29] Y. Fang, J. Yan, R. Du, Y. Zuo, W. Wen, Y. Zeng, and L. Li, "Blind quality assessment for tone-mapped images by analysis of gradient and chromatic statistics," *IEEE Transactions on Multimedia*, vol. 23, pp. 955 – 966, April 2020.

[30] J. V. D. Weijer and C. Schmid, "Coloring local feature extraction," *Lecture Notes in Computer Science*, pp. 334–348, 2006.

[31] H. E. Gerhard, F. A. Wichmann, and M. Bethge, "How sensitive is the human visual system to the local statistics of natural images?," *PLoS computational biology*, vol. 9, no. 1, p. e1002873, 2013.

[32] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980.

[33] W. T. Loh and D. B. L. Bong, "Quality assessment for natural and screen visual contents," *IEEE International Conference on Image Processing, Taipei, Taiwan*, p. 3025–3026, September 2019.

[34] m. G. Martini, c. T.E.R. Hewage, and B. Villarini, "Image quality assessment based on edge preservation," *Signal Processing: Image Communication*, vol. 27, no. 8, pp. 875–882, 2012. Special issue on: pervasive mobilemultimedia.

[35] D. Sadykova and A. P. James, "Quality assessment metrics for edge detection and edge-aware filtering: A tutorial review," in *Quality assessment metrics for edge detection and edge-aware filtering: A tutorial review*, pp. 2366–2369, 2017.

[36] Z. Ni, L. Ma, H. Zeng, C. Cai, and K. Ma, "Gradient direction for screen content image quality assessment," *IEEE Signal Processing Letters*, vol. 23, pp. 1394 – 1398, October 2016.

[37] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, and K.-K. Ma, "ESIM: Edge similarity for screen content image quality assessment," *IEEE Transactions on Image Processing*, vol. 26, pp. 4818–4831, October 2017.

[38] Y. Fu, H. Zeng, L. Ma, Z. Ni, J. Zhu, and K. Ma, "Screen content image quality assessment using multiscale difference of gaussian," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, pp. 2428 – 2432, September 2018.

[39] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7-1, pp. 11–32, 1991.

[40] A. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, vol. 29, no. 8, pp. 1233–1244, 1966.

[41] C. Won, D. Park, and S.-J. Park, "Efficient use of mpeg7 edge histogram descriptor," *Electronics and Telecommunications Research Institute (ETRI)*, vol. 24, pp. 23–30, February 2002.

[42] L. Kabbai, M. Abdellaoui, and A. Douik, "Image classification by combining local and global features," *The Visual Computer*, vol. 35, pp. 679–693, 2019.

[43] M. A. Stricker and M. Orengo, "Similarity of color images," *Proceedings of SPIE*, vol. 2420, pp. 381–392, March 1995.

[44] R. Kusumoto, X. Han, and Y.-W. Chen, "Hybrid aggregation of sparse coded descriptors for food recognition," in *2014 22nd International Conference on Pattern Recognition*, pp. 1490–1495, 2014.

[45] R. Cai and M. Fang, "Blind image quality assessment by simulating the visual cortex," *The Visual Computer*, pp. 1–18, 2022.

[46] C. C. Chang and C. J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

[47] Z. Ni, L. Ma, H. Zeng, Y. Fu, L. Xing, and K. Ma, "SCID: a database for screen content images quality assessment," *International Scientific Publications and Consulting Services*, pp. 774–779, 2016.

[48] S. Wang, K. Gu, X. Zhang, W. Lin, L. Zhang, S. Ma, and W. Gao, "Subjective and objective quality assessment of compressed screen content images," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 4, pp. 532–543, 2016.

[49] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," *http://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx*, August 2015.

[50] J. Ji, K. Xiang, and X. Wang, "Scvs: blind image quality assessment based on spatial correlation and visual saliency," *The Visual Computer*, pp. 1–16, 2022.

[51] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.

[52] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

Hamidreza Farhadi Tolie is currently a Ph.D. student at National Subsea Centre, Robert Gordon University, UK. He got his B.Eng degree in Information Technology, in 2018, and his M.Sc. degree in Computer Science, in 2021, both from the Institute for Advanced Studies in Basic Sciences (IASBS). His research interests include, image quality assessment, image enhancement, image processing, and computer vision.

Mohammad Reza Faraji received his B.S. degree in Applied Mathematics from Zanjan University in 2005, his M.S. degree in Systems Engineering from Amirkabir University of Technology in 2009, and his Ph.D. degree in Computer Science from Utah State University in 2015. Since 2017, he collaborates with the Department of Computer Science and Information Technology at Institute for Advanced Studies in Basic Sciences. His research interests include content-based image retrieval, image quality assessment, computer vision, and data science.

Xiaojun Qi received her B.S. degree in Computer Science from Donghua University in 1993, her M.S. degree in Pattern Recognition and Artificial Intelligence from the Chinese Academy of Sciences in 1996, Shenyang, and her Ph.D. degree in Computer Science from Louisiana State University in 2001. She is a Professor in the Department of Computer Science and the director of the Computer Vision Laboratory at Utah State University (USU). She has expertise in artificial intelligence, with the four central themes being image processing, machine learning, deep learning, and computer vision.