# Blind Quality Assessment of Screen Content Images Via Macro-Micro Modeling of Tensor Domain Dictionary

Yongqiang Bai, Zhongjie Zhu, Gangyi Jiang, *Senior Member, IEEE*, and Huifang Sun, *Fellow, IEEE*

***Abstract*—Screen content images (SCIs) have been rapidly and widely applied in interactive multimedia applications. The problem of quality assessment for SCIs is an interesting research topic. Most of the existing methods use subjective and independent features in gray domain to predict the image quality, which cannot comprehensively characterize the image properties or lack unified mathematical explanation for SCIs. To address these problems, we propose a novel blind quality assessment method based on macro-micro modeling of tensor domain dictionary for SCIs in this article. In the proposed method, the tensor decomposition is explored first to avoid the loss of color information, and then a target dictionary is learned more effectively with the principal components. Furthermore, a macro-micro model is established to characterize the micro and macro features in the target dictionary space, which can provide a systematic mathematical interpretation for feature extraction. For the micro features, a log-normal pooling scheme is designed to enhance the effectiveness of feature aggregation by analyzing the particularity of the statistical distribution of sparse codes. Additionally, the statistical properties are mainly discussed and studied based on the Bernoulli law of large numbers, and then a reliable macro feature is generated to describe the relationship between the statistical distribution and quality degradation of SCIs. Experimental results determined by using three public SCI databases show that the proposed method can perform better than relevant existing methods in the prediction of the visual quality of SCIs, especially in terms of the generalization for distortion type and interpretability for feature generation.**

Manuscript received May 27, 2020; revised September 26, 2020 and November 9, 2020; accepted November 13, 2020. Date of publication November 20, 2020; date of current version November 18, 2021. This work was supported in part by the Natural Science Foundation of China under Grant 61671412, Grant 61871247, and Grant 61931022, in part by the Natural Science Foundation of Zhejiang Province under Grant LY19F010002 and Grant LY21F010014, in part by the Commonweal Projects of Zhejiang Province under Grant LGN20F010001, in part by the Natural Science Foundation of Ningbo, China under Grant 2018A610053 and Grant 202003N4323, in part by the General Scientific Research Project of Zhejiang Education Department under Grant Y201941122, in part by the Ningbo Municipal Projects for Leading and Top Talents under Grant NBLJ201801006, in part by the Fundamental Research Funds for Zhejiang Provincial Colleges and Universities, and in part by the School-level Research and Innovation Team of Zhejiang Wanli University. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Marco Carli. *(Corresponding author: Zhongjie Zhu; Gangyi Jiang.)*

Yongqiang Bai and Zhongjie Zhu are with the College of Information and Intelligence Engineering, Zhejiang Wanli University, Ningbo 315100, China (e-mail: byq-163@163.com; zhongjiezhu@hotmail.com).

Gangyi Jiang is with the Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: jianggangyi@nbu.edu.cn).

Huifang Sun is with Mitsubishi Electric Research Laboratories, Cambridge, MA 02139 USA (e-mail: hsun@merl.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TMM.2020.3039382.

Digital Object Identifier 10.1109/TMM.2020.3039382

***Index Terms*—Screen content image, image quality assessment, no-reference, macro-micro modeling, dictionary learning.**

## I. Introduction

SCREEN content images (SCIs) have been rapidly and widely applied in interactive multimedia applications such as social software, cloud computing, virtual screen sharing, online education and video games [1]–[5]. As composite images, SCIs not only contain natural scenes, document images and computer-generated graphics but also can be generated in a variety of ways. Figs. 1(a–c) show examples of a natural image and two typical SCIs, respectively. Clearly, there are some significant differences between these two SCIs. Compared with natural images, SCIs contain more thin lines, sharp edges, and little color variance due to the massive computer-generated graphics and combinations of image types. Additionally, in existing multimedia communication systems, various distortions, such as noise, compression, and blurriness, will inevitably be introduced, which will lead to image quality degradation of the SCIs [6]–[8]. Therefore, it is deemed necessary to design an efficient image quality assessment (IQA) method of SCIs for business promotion.

However, research results found that the preferred quality of SCIs was quite different from that of natural images. Hence, the existing IQA methods designed for natural images are inadequate for SCIs [9]. There are two main reasons for this. On the one hand, the intrinsic properties of SCIs are quite different from those of natural images, i.e., natural images have rich and complex distributions in terms of luminance and color, while SCIs generally contain fewer and simpler luminance/color variations and structures. On the other hand, the artificial components of SCIs, such as computer graphics and document images, are damaging to the natural scene statistics (NSS) approach [10], [11], which is quite mature for natural images and can be mapped to visual quality scores with no reference evaluation. Until now, finding particularly reliable statistical features, which can be used to characterize the intrinsic quality variations of SCIs, is still a difficult problem that must be solved effectively.

Blind IQA (BIQA), which aims to predict the quality of an image without accessing its pristine version, is an extremely vital and particularly challenging research area in perceptual image processing. Although challenging, some prior efforts have been devoted to BIQA of SCIs recently and can be


(a) (b) (c)

Fig. 1. Typical examples of natural images and SCI images. (a) Natural image, (b) Splicing SCI, and (c) Overlapping SCI.

divided into three categories: feature extraction-based methods [12]–[15], sparse representation-based methods [16]–[18] and neural network-based methods [19]–[21]. For these feature extraction-based methods, the key step is to manually extract quality-aware features by analyzing the characteristics of SCIs. Gu *et al.* extracted four types of descriptive features, i.e., picture complexity, screen content statistics, global brightness quality, and sharpness of details [12]. Lu *et al.* exploited several orientation features derived by using the orientation selectivity mechanism [13]. Fang *et al.* incorporated statistical luminance features and texture information with both local and global feature representations inspired by the human visual system (HVS) [14]. Zheng *et al.* combined entropy, contrast, and local phase coherence to achieve a fused quality representation of SCIs [15]. The features of these methods show obvious subjectivity, which cannot fully reflect the particularity of SCIs. Furthermore, sparse coding is explored to enhance the effectiveness and completeness of artificial features. Yang *et al.* considered texture information represented by using a local histogram of oriented gradient (HOG) features and predicted image quality via the sparse coding coefficients of HOG features [16]. Zhou *et al.* combined local and global features, which are extracted with different dictionaries [17]. Similarly, Wu *et al.* explored sparse representation to extract local structural feature and combined the luminance statistical feature and local binary pattern feature to predict image quality [18]. These methods adopted sparse coding to further optimize and integrate local features and global features, but because these features are still independent and not integrated in a theoretical system, the performance of these methods has not been improved significantly. In addition, neural network-based methods also achieve good performance by taking advantage of end-to-end characteristics. Chen *et al.* designed a naturalization module to make a neural network applicable to SCIIQA [19]. Jiang *et al.* employed data selection and adaptive pooling to improve the speed and performance of network fitting [20]. Yue *et al.* trained a network with inputs from an entire image to avoid the disadvantage of training with image patches [21]. Clearly, these methods are only applicable to the optimization of a network structure and lack interpretability. Additionally, overfitting occurs due to the limited sample sizes.

In summary, the previous relevant BIQA methods were dedicated to evaluate the quality of SCIs with various perceptual features from different perspectives and demonstrated moderate performance with the legacy benchmark databases. Nevertheless, these methods still have some problems that must be solved: (1) the above feature extraction processes are carried out in the grayscale domain, ignoring the color information, which is also important in visual perception, coding, and transmission for SCIs. Additionally, these features rely on prior knowledge about the functionalities of HVS and/or the mechanism of SCI quality degradation, so the performance exhibits a serious preference effect for partial distortion types. (2) The statistical characteristics of SCIs are still not described with a reliable mathematical model. Furthermore, although the properties of local connectivity and global discretization are considered in some feature extraction processes, the two sets of features are still independent of each other and lack accurate and unified mathematical interpretation.

To overcome the limitations of the previous methods, this paper proposes a novel blind quality assessment method for SCIs based on macro-micro modeling of tensor domain dictionary. To avoid the incompleteness of feature extraction, the principal component of the tensor decomposition (TD) is selected to construct the target dictionary for subsequent feature coding. Furthermore, a macro-micro model provides a systematic mathematical interpretation for feature extraction and effectively improves the prediction accuracy. In brief, compared with previous relevant work, the main contributions of this paper are summarized as follows:

1) Tensor decomposition is explored to effectively ensure the integrity of image information and, in particular, to reduce the complexity for subsequent feature coding. On the one hand, tensor decomposition, as a reversible process, can perfectly combine color and brightness information, limit information loss and optimize the structure of feature extraction. On the other hand, the selection of the principal component conforms to the characteristics of HVS and optimizes the reconstruction effect of quality-aware features with sparse representation.
2) Sparse coding technique is adopted to enhance the effect of feature extraction in the aspects of effectiveness and objectivity. In the target dictionary space, effective quality-aware features can be generated automatically by sparse coefficient of reconstruction, without depending on any prior knowledge about the functionalities of HVS and the mechanism of SCI quality degradation.
3) A macro-micro model is established to explicitly explain the causation of feature generation from unified mathematical principles. Considering the randomness of textual and pictorial regions in SCIs, dictionary atoms are adopted to reflect the details of texture for each region from a microscopic perspective, which can effectively solve the complexity and one-sidedness of artificial features. Additionally, a log-normal distribution-based local pooling scheme is designed to further reduce the information loss in micro feature aggregation. Furthermore, from a macroscopic perspective, an effective macro statistical feature for SCIs is generated by combining the Bernoulli law of large numbers and sparse coding. This statistical feature is related to the properties of the dictionary atoms mentioned above via unified mathematical theory, which ensures the integrity and interpretability of feature generation.

The rest of this paper is organized as follows. Section II describes the motivation and methodology of the proposed method

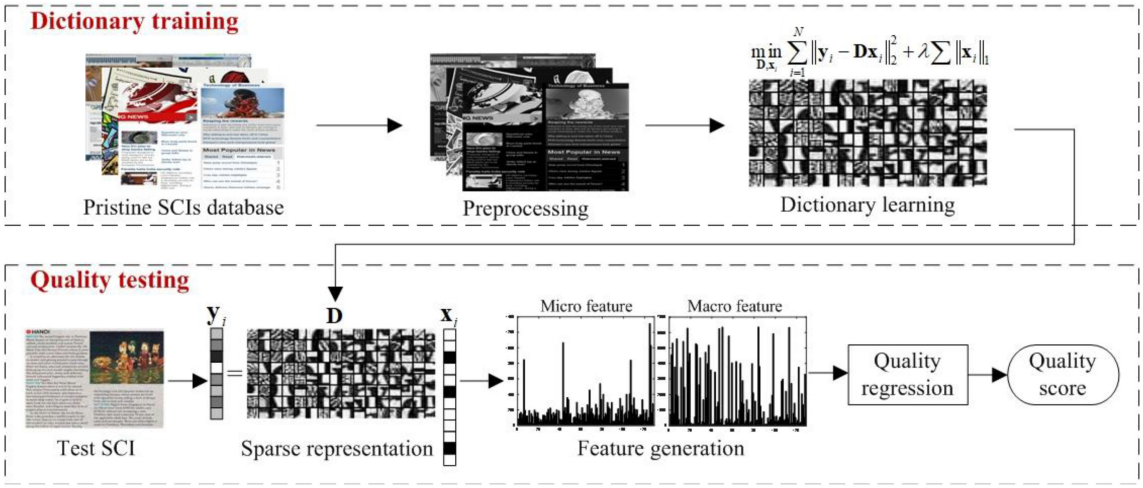



Fig. 2. Framework of the proposed BIQA method for SCIs.

in detail. Then, the relevant experimental results and a comparative analysis are presented in Section III. Finally, Section IV concludes this paper.

## II. Motivation and Methodology

Considering the randomness of SCI composition and the HVS characteristics, this paper presents a novel BIQA method for SCIs based on macro-micro modeling of tensor domain dictionary. The framework of the proposed method is illustrated in Fig. 2. Clearly, this method is based on the well-known sparse representation framework, in which the overall feature vectors are the combination of micro and macro features. Overall, the proposed method involves two stages: dictionary training and quality testing. For the former, the feature dictionary is constructed in advance based on a set of principal components after Tucker tensor decomposition. For the quality prediction, the micro and macro features of the test SCI are generated with the macro-micro model and the learned dictionary, and then support vector regression (SVR) is implemented to predict the overall quality score.

### A. Preprocessing

As mentioned above, the color information has not been taken into account in the previous research of BIQA for SCIs. Hence, TD is explored to tackle this gap. A tensor, as the high-order extension of a matrix, can seamlessly capture the correlation between the spatial and spectral domains simultaneously, with obvious advantages in the extraction of texture details. Therefore, it is an appropriate strategy for solving this primary problem.

Tucker tensor decomposition is a form of higher-order principal component analysis. It decomposes a tensor $\chi \in \mathbb{R}^{I_1 \times I_2 \times \cdots I_N}$ into a core tensor $\varsigma \in \mathbb{R}^{J_1 \times J_2 \times \cdots J_N}$ multiplied (or transformed) by a group of matrices $\mathbf{Y}^{(n)} \in \mathbb{R}^{I_N \times J_N}$ ($1 \leq n \leq N$) along each mode [22].

$$\chi \approx \varsigma \times_1 \mathbf{Y}^{(1)} \times_2 \mathbf{Y}^{(2)} \cdots \times_N \mathbf{Y}^{(N)} \tag{1}$$

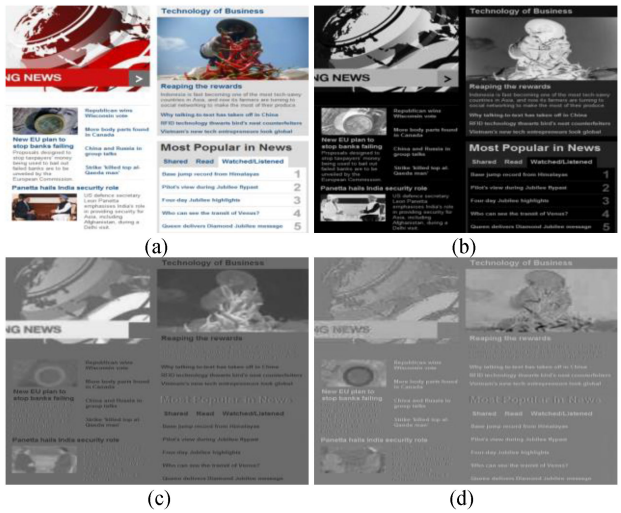



Fig. 3. Visualization of the original image and the components obtained with TD. (a) Original image. (b) Component $\mathbf{Y}^{(1)}$. (c) Component $\mathbf{Y}^{(2)}$. (d) Component $\mathbf{Y}^{(3)}$.

An RGB data cube can be represented by a three-order tensor$\chi \in R^{I_1 \times I_2 \times I_3}$, where $I_1$, $I_2$, and $I_3$ indicate the sizes of the red, green, and blue channels of the RGB image, respectively.

$$\chi \approx \varsigma \times_1 \mathbf{Y}^{(1)} \times_2 \mathbf{Y}^{(2)} \times_3 \mathbf{Y}^{(3)} \tag{2}$$

where $\mathbf{Y}^{(1)} \in \mathbb{R}^{I_1 \times I_1}$,$\mathbf{Y}^{(2)} \in \mathbb{R}^{I_2 \times I_2}$, and $\mathbf{Y}^{(3)} \in \mathbb{R}^{I_3 \times I_3}$ are the factor matrices(which are usually orthogonal) and can be thought of as the principal component in each mode. The elements of the core tensor $\varsigma \in \mathbb{R}^{I_1 \times I_2} \times I_3$ describe the level of interaction between the different components. Fig. 3 shows the visualization of the original image and the components obtained with TD. Clearly, the principal component $\mathbf{Y}^{(1)}$effectively fuses the brightness and color information while preserving the primary texture details. Hence, we employ the principal component, i.e., component $\mathbf{Y}^{(1)}$, as the carrier of dictionary learning in this paper according to the characteristics of HVS.

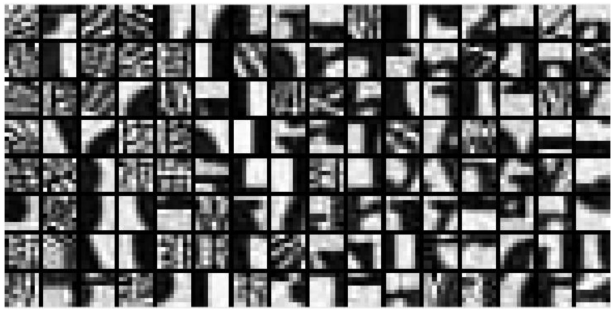


Fig. 4. Learned dictionary with component $\mathbf{Y}^{(1)}$ of TD by using the K-SVD algorithm.

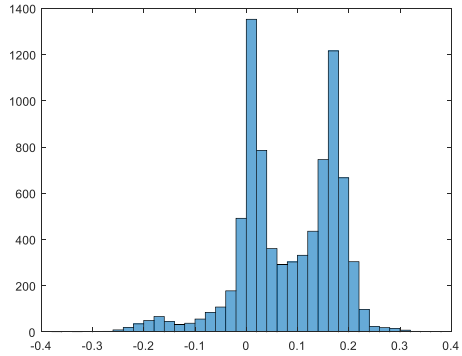



Fig. 5. Histogram of the numerical distribution for all atoms in the dictionary.

### B. Dictionary Learning

After obtaining the carrier data of SCIs, a critical question remains to be solved: how can a highly effective quality-aware feature vector or subspace be constructed while maintaining the internal structure of the carrier? Studies in neuroscience have shown that the receptive field properties of simple cells can be effectively represented via sparse representation [23]. As one model of sparse representation, dictionary learning always tries to capture the underlying and pristine properties of images by learning a set of basis vectors, which makes it easier to establish the intrinsic relationship between the feature and quality vectors. Aiming at this goal, in this paper, complete features are automatically generated without any prior knowledge about the functionalities of the HVS.

The construction of the dictionary plays an important role in the sparse representation of signals. To date, many dictionary learning algorithms have been proposed [24]–[26]. Among them, the K-SVD algorithm, as a generalization of the k-means clustering method, has been widely used due to its competitive performance in a variety of image processing applications [24]. Following our previous work [27], the dictionary is learned by using the K-SVD algorithm in this paper. First, the carrier data $\mathbf{Y}^{(1)}$, abbreviated as $\mathbf{Y}$ for simplicity, is used as input data. Then, with $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{N}$, where $\mathbf{y}_i \in \mathbb{R}^p$ is the $i$th $\sqrt{p} \times \sqrt{p}$ patch, an over complete dictionary matrix $\mathbf{D} \in \mathbb{R}^{p \times k}$ is learned with the K-SVD algorithm. Specifically, the target dictionary can be generated by solving the following optimization problem:

$$\min_{\mathbf{D},\mathbf{x}_i} \sum_{i=1}^{N} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 + \lambda \sum \|\mathbf{x}_i\|_1 \tag{3}$$

where $\mathbf{x}_i \in \mathbb{R}^k$ denotes the sparse code of $\mathbf{y}_i \in \mathbb{R}^p$ over $\mathbf{D} \in \mathbb{R}^{p \times k}$. Note that the overcomplete basis rather than the complete basis is used in the dictionary learning (that is, $k > p$). The influence of different values of $p$ and $k$ on the performance will be analyzed and elaborated in Section III.

Additionally, the learned dictionary is no longer required to be updated and can be directly used for feature encoding in the testing stage. In this paper, the learned dictionary is shown in Fig. 4, in which each atom represents different structural characteristics, and its corresponding reconstruction coefficient represents the strength of the structural characteristics for each patch.

Fig. 5 shows the histogram of the numerical distribution for all the atoms in the learned dictionary. Obviously, the learned dictionary contains positive and negative values. Hence, these values will directly affect the image reconstruction effect and put forward high requirements for subsequent feature coding and aggregation, which are illustrated in detail below.

### C. Sparse Representation

As an activation function, the purpose of feature coding is to obtain new feature representations with the transformation from the original feature space to the target dictionary space. In this paper, sparse coding is adopted to obtain the desirable feature representation, as it can represent an input signal as a sparse linear combination of the atoms in the dictionary, meaning that most coefficients are zeroes [28], [29].

Here, we use $\hat{\mathbf{Y}} = \{\hat{\mathbf{y}}_i\}_{i=1}^{N}$ as the symbol for the principal component of the tested SCI after preprocessing, and then the sparse code $\hat{\mathbf{x}}_i$ for an arbitrary patch can be calculated mathematically with the learned dictionary $\mathbf{D} \in \mathbb{R}^{p \times K}$ as follows.

$$\hat{\mathbf{X}} = \arg\min \|\hat{\mathbf{x}}_i\|_0, \quad \text{s.t.} \left\|\hat{\mathbf{Y}} - \mathbf{D}\hat{\mathbf{X}}\right\|_2^2 \leq T \tag{4}$$

where $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_N] \in \mathbb{R}^{p \times N}$. $T$ defines the predefined error threshold, which is empirically set to 5 in the implementation. Hereby, new feature representations can be obtained in the target dictionary space, and the activities associated with each atom are the resultant feature codes. The influence of different values of $T$ on the performance will be analyzed and elaborated in Section III.

Specifically, the batch-OMP algorithm [30] is used to obtain the optimal solution for Eq. (4) presented in this paper. Subsequently, the calculated sparse code $\hat{\mathbf{X}}$ directly serves as the patch-level feature vector for subsequent feature generation.

### D. Feature Generation

After sparse representation, the feature vectors for all the patches are calculated from a test image, which can completely capture the inherent essential characteristics of the carrier in a sparse representation. In this target dictionary space, these patch-level feature vectors should be aggregated to a final image-level

feature vector convenient for quality regression, as the goal of IQA is to estimate a single score for a whole image.

Considering the viewing behaviors of human eyes, both local and global properties of SCIs should be taken into account in the design of the IQA model. Following a similar research idea, we have conducted an intensive study of these patch-level feature vectors by considering both micro and macro aspects. Hereby, a macro-micro model is established in this paper to explicitly explain the causation of feature generation by combining the Bernoulli law of large numbers and sparse coding. The generation of micro and macro features will be described in detail below.

*1) Micro Feature Aggregation:* As mentioned above, all the atoms of the learned dictionary are directly used as the basic elements to characterize the images in the proposed method. Here, the properties of the dictionary and the associated sparse coefficients are analyzed. For the dictionary$\mathbf{D} \in \mathbb{R}^{p\times k}$, the test image can be represented with the sparse code as follows.

$$\hat{\mathbf{Y}} = \mathbf{D}\hat{\mathbf{X}} = \sum_{i=1}^{N} \mathbf{D}\hat{\mathbf{x}}_i \Rightarrow \sum_{i=1}^{N}\sum_{j=1}^{k} \alpha_{i,j} d_j \tag{5}$$

where $d_j \in \mathbb{R}^p$ denotes the *j*th atom of the learned dictionary, and each atom is a *p*-dimensional vector. $\alpha_{i,j}$ denotes the associated sparse coefficient of the *j*th atom for the *i*th patch. Note that the values of $\alpha_{i,j}$ directly reflect the strength of each atom for the image, meaning that the micro feature of the test image can be characterized with these coefficient vectors though feature aggregation.

Clearly, since the patch-level sparse coefficients cannot be directly used as the feature vectors because of the high dimensions, the pooling scheme should be effectively adapted to the properties of the dictionary used. As shown in Fig. 5, the learned dictionary contains positive and negative values for all the atoms, so the derived sparse codes also contain a large number of zero coefficients and few nonzero coefficients with positive and negative values. Fig. 6 shows the histogram of sparse codes when using the learned dictionary to reconstruct the image in Fig. 3(b) and illustrates the efficiency of image reconstruction with the small number of atoms used in sparse coding. Currently, the two most common local to global pooling operations are mean-pooling and max-pooling. However, considering the positive and negative values, as shown in Fig. 6, if these pooling operations are directly adopted here, it inevitably will make the aggregated features less distinguishable, consequently making the pooling meaningless and reducing the accuracy of the algorithm.

To solve this problem, a new pooling scheme is designed based on the log-normal distribution through the analysis of sparse coding in this paper. As shown in Fig. 6, there is no obvious regularity for the distribution of those sparse coefficients on the whole. As the atoms with zero coefficients are not concerned with the sparse reconstruction for each patch, and following our previous work [27], the main features for each image can still be characterized accurately while reconstructing the image with partial coefficients of nonzero value. Therefore, nonzero coefficients are selected as the target object of pooling, instead of all the sparse coefficients. Since each atom is directly used to

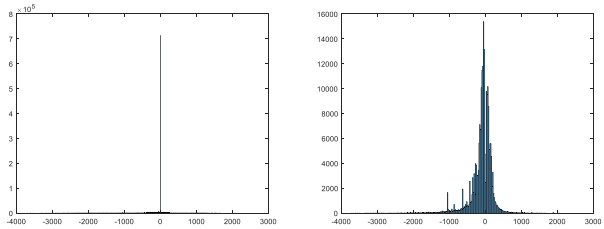



Fig. 6. Histogram of sparse code coefficients reconstructed by the learned dictionary for the image in Fig. 3(b). (a) With all values, (b) Without the '0' value.

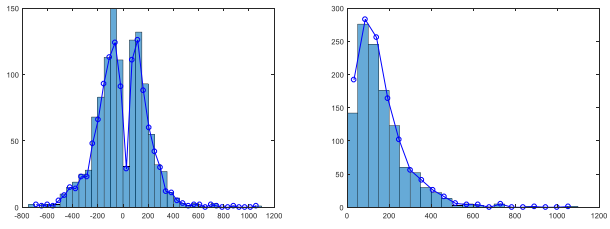



Fig. 7. Histogram of sparse code coefficients for atom 1 of the learned dictionary. (a) Original nonzero coefficients. (b) Absolute nonzero coefficients.

describe the basic micro features, all the pristine images are reconstructed with sparse coding, and their sparse coefficients are statistically analyzed according to each atom. Taking the image in Fig. 3(b) as an example, there are 1157 sparse coefficients with nonzero values (including 630 negative values and 527 positive values) and 6067 sparse coefficients of zero for atom 1 in the learned dictionary. Fig. 7 shows the histogram of the sparse code coefficients of the original nonzero values and that of the absolute nonzero values. Intuitively speaking, the value of atom 1 generally follows a log-normal distribution with the probability density function below:

$$f(x;\mu;\sigma) = \frac{1}{\sqrt{2\pi}x\sigma} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \tag{6}$$

$$E(x) = e^{\mu+\sigma^2/2} \tag{7}$$

where $\mu$ and $\sigma$ are the mean and standard deviation of the logarithm of the variables. $E(x)$ denotes the expectation of the log-normal distribution.

Additionally, similar test results are observed for the rest of the atoms. And the results of distorted images indicate that the distribution shifts to different degrees depending on the type and intensity of image distortion. Fig. 8 illustrates this phenomenon with some of the results, and it can be seen that the offset takes on an infinite variety of forms due to different distortions. Considering this fact, the micro features are aggregated with the log-normal distribution-based pooling scheme in this paper. With this scheme, each distortion can be distinguished much more equally and efficiently than with conventional schemes.

Specifically, with the patch-level feature codes $\hat{\mathbf{X}} \in \mathbb{R}^{p\times N}$ as input, the proposed pooling scheme works as follows. First, nonzero codes $\hat{\mathbf{x}}_i$ are extracted according to each atom. Then,

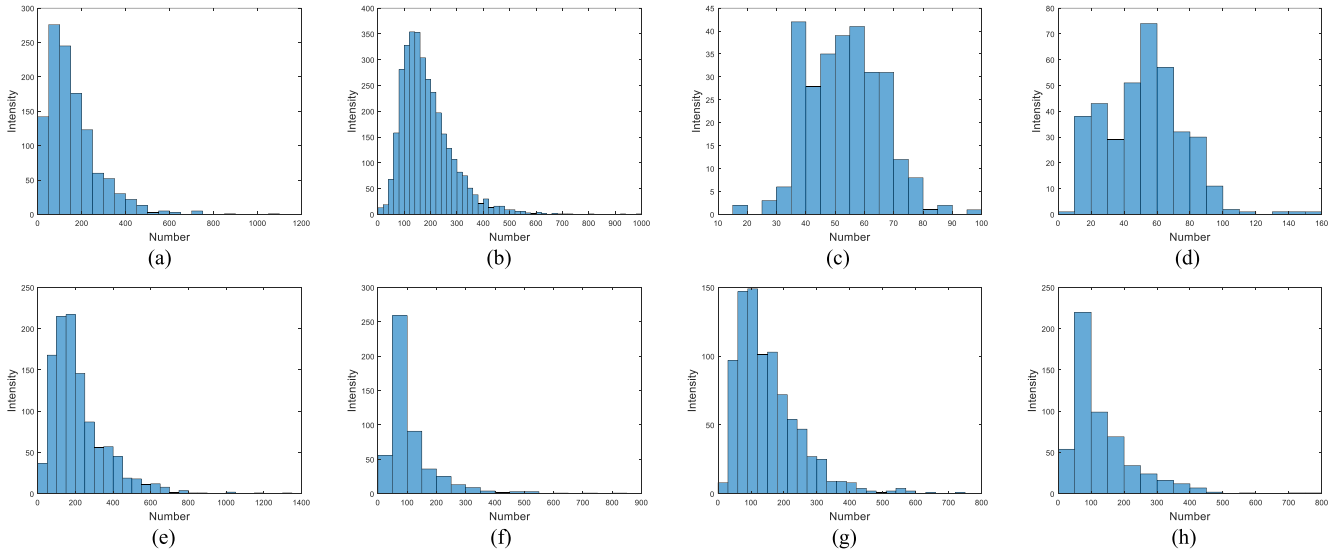



Fig. 8. Histogram of sparse coefficients for atom 1 with different distortion types. (a) Pristine image. (b) With Gaussian Noise. (c) With Gaussian Blur. (d) With Motion Blur. (e) With Contrast Change. (f) With JPEG. (g) With JPEG 2000. (h) With Layer Segmentation-based Coding.

their absolute values are used to calculate the mean and standard deviation. Finally, the proposed pooling scheme is expressed as

$$\mathrm{f}^{\mathrm{mic}} = e^{\mu+\sigma^2/2} \tag{8}$$

where $\mathrm{f}^{\mathrm{mic}} = [\mathrm{f}_1^{\mathrm{mic}}, \ldots, \mathrm{f}_k^{\mathrm{mic}}]$ represents the micro feature vectors for all the atoms. The performance of the proposed pooling scheme is compared with the mean-pooling and max-pooling schemes in Section III.

Besides, for the same pristine image, the numbers of the non-zero coefficients of each atom are changed for its different distorted images, which will affect the reliability of micro feature aggregation obtained above. In this regard, statistical information of each atom in sparse reconstruction is introduced logically as macro feature, which can not only make up for the deficiency of non-zero coefficient pooling in micro feature aggregation here, but also can measure the role of each atom in sparse reconstruction from a macro perspective. Specific analysis is shown in the next subsection.

*2) Macro Feature Generation:* In image processing, statistical information, which can characterize the properties of scenes from a macro perspective, has been widely investigated. For natural images, the NSS is widely used, but it does not apply to SCIs due to the artificial portions. This work tries to fill this gap with the use of sparse codes in the target dictionary space. As mentioned above, the test image can be reconstructed with almost lossless fidelity using partial atoms of different quantities and intensities when sparse codes with zero value are not considered. Hereby, the micro features are characterized by pooling the nonzero coefficient values of atoms (that is, the intensity for each atom). Intuitively, the composition of its quantity may reflect the macro characteristics. According to the Bernoulli law of large numbers, for a random event $Q$ in a random environment, the frequency of the event $Q$ will be stable near the probability of the event with the increase in the number of tests [31], [32]. That is,

$$\lim_{n\to\infty} P\left(\left|\frac{n_Q}{n} - s\right| < \varepsilon\right) = 1 \tag{9}$$

where $n_Q$ denotes the occurrence number of event $Q$ in the $n$-fold Bernoulli experiments. $P(Q)$ denotes the probability of event $Q$ in each test, and $P(Q) = s$. $\varepsilon$ is any positive number.

Based on the inference that probability converges to frequency, numerous experiments were conducted with the various distorted images to test the relationship between the frequency of each atom and quality degradation of the image in the target dictionary space, and the results indicate that the occurrence numbers of each atom changed with the type and intensity of the image distortion by using the probability statistics method. Fig. 9 shows the occurrence numbers for each atom with different types of distortion of the image in Fig. 3(b). Through comparison and analysis of each distortion type, it is shown that for identical images, the distributions of all the atoms have self-similar properties as a whole, while they exhibit subtle changes due to the different occurrence numbers for each distortion. Clearly, this phenomenon shows that there is a statistical law in the sparse space, whether in the artificial region or in the natural region. It does not depend on prior information about the functionalities of the HVS and changes with the degradation of image quality. With this observation, this statistical property of occurrence number can be used to characterize the intrinsic variations in the quality of SCIs from a macro perspective.

With the patch-level feature codes $\hat{\mathbf{X}} \in \mathbb{R}^{p\times N}$, macro features are generated as follows. First, the occurrence numbers of each atom are calculated according to feature codes $\hat{\mathbf{X}} \in \mathbb{R}^{p\times N}$. Then, the probability of each atom can be obtained through the normalization process.

$$f_i^{\mathrm{mac}} = n_i / \sum_{i=1}^{k} n_i \tag{10}$$

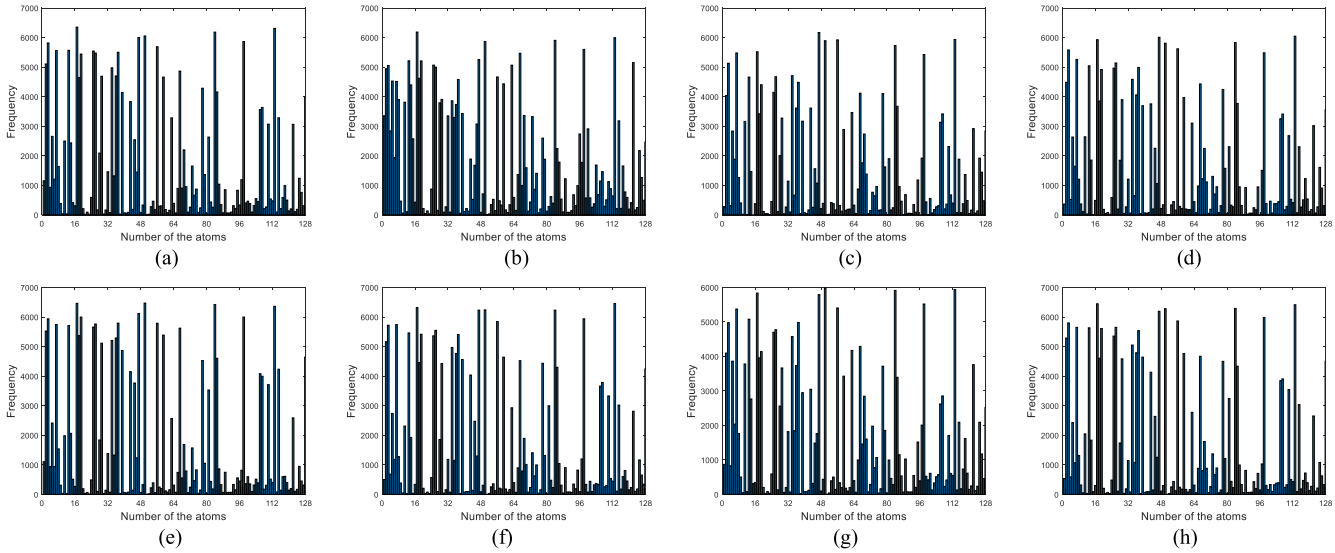



Fig. 9. Histogram of the occurrence number for each atom with different types of distortions. (a) Pristine image. (b) With Gaussian Noise. (c) With Gaussian Blur. (d) With Motion Blur. (e) With Contrast Change. (f) With JPEG. (g) With JPEG 2000. (h) With Layer Segmentation-based Coding.

TABLE I
BRIEF INTRODUCTION OF THREE PUBLIC SCI DATABASES

| | Image number | | Distortion | |
|---|---|---|---|---|
| | Reference | Distorted | Type | Level |
| SIQAD | 20 | 980 | 7 | 7 |
| SCD | 24 | 492 | 2 | / |
| SCID | 40 | 1800 | 9 | 5 |

where $\mathrm{f}^{\mathrm{mac}}=[\mathrm{f}_1^{\mathrm{mac}},\ldots,\mathrm{f}_k^{\mathrm{mac}}]$ represent the macro feature vectors for all the atoms. $n_i$ represents the occurrence numbers of atom *i* for the test image. Finally, the final image-level feature vector is constructed by directly combining the above two feature vectors, denoted as $\mathrm{f}=[\mathrm{f}^{\mathrm{mic}},\mathrm{f}^{\mathrm{mac}}]\in\mathbb{R}^{2k\times 1}$.

### *E. Quality Regression*

With the feature vectors extracted, quality regression is achieved using SVR in this paper to create a final comparison with the state-of-the-art BIQA methods. Specifically, an SVR model is first trained using a set of training SCIs. Then, the trained SVR model is used to evaluate the quality of testing SCIs. Here, SVR with a radial basis function kernel is adopted as the mapping function from normalized features to subjective quality scores by using the LIBSVM package [33]. In the implementation, the patch size *p* is set to 8×8, and the dictionary size *k* is set to 128, such that there are 256 features in total for each input SCI.

## III. EXPERIMENTAL RESULTS AND DISCUSSION

In this paper, three public SCI databases, the screen content image quality assessment database (SIQAD) [34], screen content database (SCD) [35] and screen content image database (SCID) [36], are selected as the test platforms to verify the effectiveness of the proposed method. Table I gives a brief introduction to these databases. For the distortion type, SIQAD includes Gaussian Noise (GN), Gaussian Blur (GB), Motion Blur (MB), Contrast Change(CC), JPEG, JPEG 2000 (J2K) and Layer Segmentation based Coding (LSC); SCD includes Screen Content Compression (SCC) and High Efficiency Video Coding (HEVC); SCID includes GN, GB, MB, CC, Color Quantization with Dithering (CQD), JPEG, J2K, HEVC and SCC.

Meanwhile, three commonly used criteria are adopted to evaluate the performance of the IQA method: Pearson's linear correlation coefficient (PLCC), Spearman's rank order correlation coefficient (SRCC) and root mean square error (RMSE). Generally speaking, PLCC and SRCC estimate a better prediction accuracy and monotonicity with higher values in the range of 0 and 1, and RMSE reflects a better prediction consistency with lower values, although there are some limitations for these three criteria [37], [38]. Additionally, a nonlinear logistic regression process with five parameters is applied to remove the nonlinearity of objective quality predictions as follows [39]:

$$f(x)=\beta_1\left(\frac{1}{2}-\frac{1}{1+e^{(\beta_2(x-\beta_3))}}\right)+\beta_4 x+\beta_5 \qquad (11)$$

where $(\beta_1,\ldots,\beta_5)$are the parameters to be fitted and x and $f(x)$ denote the original and the fitted quality scores, respectively.

For fairness, the database is randomly divided into training and testing subsets 1000 times, with 80% of the data used for training and the rest used for testing, and the median of the 1000 results is reported as the overall performance.

### *A. Overall Performance Comparison*

To investigate the effectiveness of the proposed method, we first compare the proposed method with the following full reference (FR) IQA methods, including five classic methods built for natural images (PSNR, SSIM [40], FSIM [41], VSI [42], and VIF [43]) and five top methods built for SCIs (SVQI [44], GFM [45], SQE [46], MDOGS [47], and EFGD [48]). All the results of the classic methods are obtained by running the codes released by

TABLE II
PERFORMANCE COMPARISONS WITH FR-IQA METRICS, USING THE SIQAD, SCD, AND SCID

| Method | SIQAD | | | SCD | | | SCID | | |
|---|---|---|---|---|---|---|---|---|---|
| | PLCC | SRCC | RMSE | PLCC | SRCC | RMSE | PLCC | SRCC | RMSE |
| PSNR | 0.5869 | 0.5608 | 11.5859 | 0.8610 | 0.8589 | 1.1273 | 0.7622 | 0.7512 | 9.1682 |
| SSIM | 0.5912 | 0.5836 | 11.5450 | 0.8696 | 0.8683 | 1.0953 | 0.7343 | 0.7146 | 9.6133 |
| FSIM | 0.5746 | 0.5652 | 11.6120 | 0.9019 | 0.9039 | 0.9585 | 0.7719 | 0.7550 | 9.0040 |
| VSI | 0.5403 | 0.5199 | 11.9380 | 0.8715 | 0.8719 | 1.0879 | 0.7550 | 0.7530 | 9.3470 |
| VIF | 0.8198 | 0.8065 | 8.1969 | 0.9028 | 0.9043 | 0.9542 | 0.8200 | 0.7969 | 8.1069 |
| SVQI | 0.8911 | 0.8836 | 6.4965 | **0.9158** | **0.9194** | **0.8909** | 0.8604 | 0.8386 | 7.2178 |
| GFM | 0.8828 | 0.8735 | 6.7234 | / | / | / | 0.8760 | **0.8759** | 6.8310 |
| SQE | **0.9040** | **0.8940** | **6.1150** | **0.9290** | **0.9310** | **0.8210** | **0.9150** | **0.9140** | **5.7610** |
| MDOGS | 0.8839 | 0.8822 | 6.6951 | / | / | / | / | / | / |
| EFGD | **0.8993** | **0.8901** | **6.2595** | / | / | / | **0.8846** | **0.8774** | **6.6044** |
| Proposed | **0.9162** | **0.9090** | **5.7111** | **0.9196** | **0.9123** | **0.8654** | **0.8811** | 0.8730 | **6.7031** |

TABLE III
PERFORMANCE COMPARISONS WITH THE STATE-OF-THE-ART NR IQA METHODS, USING THE SIQAD, SCD, AND SCID

| Method | SIQAD | | | SCD | | | SCID | | |
|---|---|---|---|---|---|---|---|---|---|
| | PLCC | SRCC | RMSE | PLCC | SRCC | RMSE | PLCC | SRCC | RMSE |
| SIQE | 0.7904 | 0.7593 | 8.7899 | 0.7168 | 0.7012 | 1.5470 | 0.6371 | 0.6034 | 10.9202 |
| OSM | 0.8306 | 0.8007 | 7.9331 | 0.7068 | 0.6804 | 1.5301 | / | / | / |
| NRLT | 0.8442 | 0.8202 | 7.5957 | / | / | / | 0.6625 | 0.6454 | 10.6452 |
| HRFF | 0.8520 | 0.8320 | 7.4150 | / | / | / | / | / | / |
| TFSR | 0.8618 | 0.8354 | 7.4910 | / | / | / | 0.8017 | 0.7840 | 8.8041 |
| LGFL | 0.8280 | 0.7880 | / | / | / | / | / | / | / |
| CLGF | 0.8331 | 0.8107 | 7.9172 | / | / | / | 0.6978 | 0.6870 | 10.1439 |
| PICNN | 0.8960 | **0.8970** | 6.7900 | / | / | / | **0.8270** | **0.8220** | **8.0130** |
| QODCNN | **0.9008** | 0.8888 | **6.2258** | / | / | / | / | / | / |
| IGMCNN | 0.8834 | 0.8634 | / | / | / | / | / | / | / |
| CSC | **0.9109** | **0.8976** | **5.8930** | **0.9182** | **0.9080** | **0.8721** | **0.8531** | **0.8377** | **7.3930** |
| Proposed | **0.9162** | **0.9090** | **5.7111** | **0.9196** | **0.9123** | **0.8654** | **0.8811** | **0.8730** | **6.7031** |

the authors to avoid unnecessary mistakes, and the other results are cited from the references. The performance comparisons are shown in Table II, and the top three performances of each category are highlighted in bold. From Table II, we can draw the following conclusions. First, the classic methods fail to estimate the quality of the SCIs since they do not consider the perceptual characteristics of SCIs. Thus, due to the peculiarity of SCIs, the existing methods for natural images cannot be directly used on SCIs. Second, for the rest of the FR-IQA methods built for SCIs, their performance is significantly improved, as the original reference images can provide more accurate and reliable feature measurements. However, the reference image is not available in most cases; thus, the application was restricted for these FR-IQA methods. Third, the proposed method results in a slightly better enhancement than these FR-IQA methods by adequately considering the properties of SCIs and dictionary learning.

Furthermore, the proposed method is compared with several state-of-the-art no reference (NR) IQA methods, including SIQE [12], OSM [13], NRLT [14], HRFF [15], TFSR [16], LGFL [17], CLGF [18], PICNN [19], QODCNN [20], IGMCNN [21] and CSC [27]. The performance comparisons are shown in Table III. Due to the limited test images and distortion types provided in the SCD, the experiments are mainly conducted with the SIQAD and SCID for these NR IQA methods. Among these methods, SIQE, OSM, NRLT, and HRFF are feature extraction-based methods, TFSR, LGFL, CLGF and CSC are sparse representation-based methods, and the others are neural network-based methods. As shown in Table III, the feature extraction-based methods and sparse representation-based methods clearly perform much worse than the FR IQA method above. While there may be several causes for this, the most important factor responsible is that these artificial features are subjective and independent, regardless of the reference image. Considering the randomness of composition of SCIs, the subjectivity of features leads to information loss and serious preference for partial distortion types, and the independence of features shows that the methods lack accurate and unified mathematical interpretation. Additionally, neural network-based methods, i.e., PICNN, QODCNN, and IGMCNN, avoid the shortcomings of artificial features, and their performance compares favorably with FR IQA methods. However, these methods still lack interpretability from an architectural point of view. In addition, the experimental results in Table III indicate that the proposed method, with the newly designed macro-micro model, can perform better, with an 8% improvement in PLCC compared to the feature extraction-based and sparse representation-based methods and provides a competitive performance compared with that of the neural network-based methods.

Besides, compared with our previous work CSC [27], the proposed method can only achieve slight improvement in the

TABLE IV
PLCC RESULTS OF DIFFERENT DISTORTION TYPES FOR THE PROPOSED AND COMPARED NR METHODS USING THE SIQAD

| Method | GN | GB | MB | CC | JPEG | J2K | LSC | Variance |
|---|---|---|---|---|---|---|---|---|
| SIQE | 0.8830 | 0.8033 | 0.7810 | 0.6030 | 0.8339 | 0.8535 | 0.8921 | 9.7016E-3 |
| OSM | / | / | / | / | / | / | / | / |
| NRLT | 0.9131 | 0.8949 | 0.8993 | 0.8131 | 0.7932 | 0.6848 | 0.7228 | 8.1681E-3 |
| HRFF | 0.9020 | 0.8900 | 0.8740 | 0.8260 | 0.7630 | 0.7540 | 0.7700 | 4.0773E-3 |
| TFSR | **0.9291** | **0.9367** | **0.9243** | 0.6563 | **0.8334** | **0.8347** | 0.8069 | 9.8440E-3 |
| LGFL | 0.9030 | 0.9110 | 0.8370 | 0.6600 | 0.7620 | 0.6680 | 0.6830 | 1.2005E-2 |
| CLGF | 0.8577 | 0.9082 | 0.8609 | 0.7440 | 0.6598 | 0.7463 | 0.5575 | 1.5525E-2 |
| PICNN | / | / | / | / | / | / | / | / |
| QODCNN | 0.9130 | **0.9250** | **0.8890** | **0.8370** | 0.8300 | 0.8180 | **0.8670** | **1.7700E-3** |
| IGMCNN | / | / | / | / | / | / | / | / |
| CSC | **0.9317** | 0.9148 | **0.8846** | **0.9229** | **0.9036** | **0.9143** | **0.9294** | **2.6689E-4** |
| Proposed | **0.9390** | **0.9156** | 0.8844 | **0.9231** | **0.9140** | **0.8949** | **0.9192** | **2.8280E-4** |

TABLE V
SRCC RESULTS OF DIFFERENT DISTORTION TYPES FOR THE PROPOSED AND COMPARED NR METHODSUSING THE SIQAD

| Method | GN | GB | MB | CC | JPEG | J2K | LSC | Variance |
|---|---|---|---|---|---|---|---|---|
| SIQE | 0.8280 | 0.7942 | 0.7748 | **0.8199** | **0.8388** | **0.8493** | **0.8843** | **1.2953E-3** |
| OSM | / | / | / | / | / | / | / | / |
| NRLT | 0.8966 | 0.8812 | **0.8919** | 0.7072 | 0.7698 | 0.6761 | 0.6978 | 9.7983E-3 |
| HRFF | 0.8720 | 0.8630 | 0.8500 | 0.6870 | 0.7180 | 0.7440 | 0.7400 | 5.9357E-3 |
| TFSR | **0.9144** | **0.9311** | **0.9148** | 0.6498 | 0.8377 | 0.8354 | 0.7948 | 9.6124E-3 |
| LGFL | 0.8790 | 0.8940 | 0.8320 | 0.4870 | 0.7440 | 0.6450 | 0.6660 | 2.1645E-3 |
| CLGF | 0.8478 | **0.9152** | 0.8694 | 0.5716 | 0.6778 | 0.7681 | 0.5842 | 1.9322E-2 |
| PICNN | / | / | / | / | / | / | / | / |
| QODCNN | 0.9050 | **0.9160** | **0.8710** | 0.7000 | 0.8150 | 0.7950 | 0.8820 | 5.8290E-3 |
| IGMCNN | / | / | / | / | / | / | / | / |
| CSC | **0.9143** | 0.8971 | 0.8708 | **0.9075** | **0.8848** | **0.8911** | **0.9064** | **2.2683E-4** |
| Proposed | **0.9201** | 0.8993 | 0.8703 | **0.9102** | **0.8966** | **0.8593** | **0.8867** | **4.3180E-4** |

SIQAD and SCD, but its experimental result is satisfactory in the SCID, as shown in Table III. The specific reasons are as follow. On the one hand, tensor decomposition can perfectly combine color and brightness information, thereby can limit information loss and optimize the structure of feature extraction. On the other hand, a macro-micro model is established to explicitly explain the causation of feature generation from unified mathematical principles. That is, the macro statistical feature ensures the integrity and interpretability of feature generation. And a log-normal distribution-based local pooling scheme is designed instead of that with empirical value that seems unreliable, which can further reduce the information loss in micro feature aggregation. Therefore, the performance of the proposed method has been improved to a certain extent in all the three databases, compared with the CSC [27]. But due to the limited samples, the performance improvement is directly proportional to the database size. In addition, the computational complexity is reduced effectively compared with the CSC, as image segmentation is not needed, sparse reconstruction is reduced to be conducted only once, and the feature vector still remains as 256. Similar conclusions have yielded for distortion types in the next subsection.

### *B. Performance Comparison of Distortion Types*

To verify the performance of the proposed method for each type of distortion, performance experiments are conducted on three SCI databases, as mentioned above. Due to space limitations, only the performance results of the state-of-the-art NR IQA methods with the SIQAD are illustrated in Tables IV –VI, and the top three metrics are highlighted in bold. In these tables, the values of variance reflect the fluctuation magnitude for all the distortion types, and lower variance values are desired because they correspond to better prediction consistency. From these experimental results, the conclusion can be obtained as follows. Most existing methods are sensitive to specific types of distortion. For example, TFSR showed the best performance for handling GB distortion, with a PLCC of 0.9367, but its PLCC is only 0.6563 for CC distortion. In contrast, our proposed method is able to efficiently handle GN, CC, JPEG, J2K, and LSC distortions.

For GB and MB distortions, the performance is slightly lower than that of the best NR methods, but still very competitive with them. The specific reasons are as follows. On the one hand, these blurring distortions are equivalent to smoothing the image especially for the texture part, which will make the artificial scenes of SCI fuzzy and reduce the subjective quality, when the intensity is larger than the just noticeable difference of human eyes. As people have different sensitivities to natural scenes and artificial scenes in SCIs due to individual differences, it is more difficult to have an accurate quality evaluation compared with other distortion types. On other hand, sparse coding technique is adopted to enhance the effect of feature extraction in this paper, thereby the characteristics of the learned dictionary play an important role in

TABLE VI
RMSE RESULTS OF DIFFERENT DISTORTION TYPES FOR THE PROPOSED AND COMPARED NR METHODS USING THE SIQAD

| Method | GN | GB | MB | CC | JPEG | J2K | LSC | Variance |
|---|---|---|---|---|---|---|---|---|
| SIQE | 19.0113 | 8.2689 | 8.6522 | 12.4155 | 7.3633 | 7.1150 | 6.5744 | 19.8017 |
| OSM | / | / | / | / | / | / | / | / |
| NRLT | / | / | / | / | / | / | / | / |
| HRFF | 6.2670 | 6.7380 | 6.4660 | **6.8740** | 5.8620 | 6.5010 | **5.4730** | **0.2442** |
| TFSR | **5.3105** | **5.2141** | **5.5266** | 10.5005 | **5.2541** | **6.6377** | 5.6217 | 3.7067 |
| LGFL | / | / | / | / | / | / | / | / |
| CLGF | / | / | / | / | / | / | / | / |
| PICNN | / | / | / | / | / | / | / | / |
| QODCNN | 6.1500 | 5.7720 | **5.7620** | 6.9390 | **5.4600** | 6.0000 | **4.3380** | 0.6184 |
| IGMCNN | / | / | / | / | / | / | / | / |
| CSC | **5.3292** | **5.3767** | 6.0794 | **5.0375** | 5.5912 | **5.4480** | **5.2539** | **0.1074** |
| Proposed | **5.0506** | **5.2992** | **6.1017** | **5.0238** | **5.3266** | **5.9826** | 5.5994 | **0.1639** |

TABLE VII
PERFORMANCE INDICES OF OUR PREVIOUS WORK CSC [27] AND THE PROPOSED METHODS WITH TWO TYPES OF DISTORTION TYPES USING THE SCD

| SCD | | HEVC | SCC | Variance |
|---|---|---|---|---|
| CSC | PLCC | 0.9004 | 0.8783 | 2.4400E-04 |
| | SRCC | 0.8825 | 0.8544 | 3.9500E-04 |
| | RMSE | 0.8964 | 1.0626 | 0.0138 |
| Proposed | PLCC | 0.9194 | 0.9024 | 1.4450E-4 |
| | SRCC | 0.9086 | 0.8839 | 3.0504 E-4 |
| | RMSE | 0.8101 | 0.9637 | 0.0118 |

image reconstruction, which is also the main reason for the performance degradation of blurring distortions. As artificial scenes contain more thin lines, sharp edges and little color variance, the atoms of the learned dictionary in this paper obviously contain more texture and edge information as shown in Fig. 4. From the perspective of reconstruction quality, it is obvious that texture, edge, and other information will be better represented in the image reconstruction. But compared with other distortion types, the blurring distortions will make the artificial scenes fuzzy and make the accuracy of reconstruction decrease. And then, for the robustness of distortions, the degradation of reconstruction accuracy further results in the performance of blur distortion being slightly lower than other distortions by about 3% as shown in Tables IV-V. Similar condition is also applies to the distortion of J2K with smoothing process. In summary, the variance values of the proposed method are dozens of times lower than the variance values of other methods in particular, which demonstrate a better generalization performance across different distortion types.

Furthermore, similar performances are obtained with the SCD and SCID, as shown in Tables VII and VIII. It can be found that, the proposed method achieves better performance than our previous work CSC [27] for each distortion type at various degrees. Thus, it is clear that the proposed method can more precisely and steadily evaluate and reflect various degenerations, which further verifies the effectiveness and robustness of the proposed method.

### C. Cross-Database Validation

In this subsection, the cross-database validation is conducted on two different databases to further verify the generalization ability of the proposed method. Consider the number of sample and distortion types in the three databases, the representative database (that is SIQAD) and the largest database (that is SCID) are selected as the training database and testing database, respectively. Following the common practice of Mittal et al [49] and Ye et al [50], the model is trained on one database with 6 types of noises, i.e., GN, GB, MB, CC, JPEG, and J2K, that are common in the two databases. And then, the other database is adopted to test the prediction ability of the trained model. Specifically, for each distortion type, all the data are utilized for model training and testing to reduce the impact of sample limitation. For the overall performance, 80% of the data for one database are randomly adopted for model training, and then the other database is utilized for model testing. This operation is repeated with 1000 times, and the median performance is reported in this paper, which can reduce dependence on the scale of database and further verify the generalization ability of the proposed method.

The experimental results of cross validation are presented in Table IX, including the overall performance and the performance for each type of distortion. And in this table, (a) means that the results are obtained by training with SIQAD and testing with SCID, and (b) means that the results are obtained by training with SCID and testing with SIQAD. From Table IX, we can draw the following conclusions. First, the performance of (b) is obviously better than that of (a) by about 10%, with the main reasons as follows. With the growth of database size and complexity, the data of image features are further supplemented, which is conducive to the subsequent relationship extraction and model generation. And then, the performance is enhanced effectively as shown in Table IX(b). Second, compared with NR methods listed in Table III, the proposed method obtains an outstanding performance for the SCID and achieves a competitive performance for the SIQAD, which effectively verifies the robustness of the proposed method. Meanwhile, the performance for each distortion type is relatively consistent, which also proves the stability of the proposed method. Note that, though the cross-validation performance of the proposed method is slight worse than the in-database performances of the neural network-based methods, such as PICNN, QODCNN, and IGMCNN, the performance of the proposed method is worthy of affirmation considering the interpretability. In summary, the cross validation and prediction results show that the proposed method is of powerful prediction ability and good stability.

TABLE VIII
PERFORMANCE INDICES OF OUR PREVIOUS WORK CSC [27] AND THE PROPOSED METHOD WITH NINE TYPES OF DISTORTION TYPES USING THE SCID

| SCID | | GN | GB | MB | CC | CQD | JPEG | J2K | HEVC | SCC | Variance |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CSC | PLCC | 0.9182 | 0.9296 | 0.9030 | 0.9067 | 0.9042 | 0.8992 | 0.9028 | 0.8363 | 0.8527 | 9.1700E-4 |
| | SRCC | 0.8962 | 0.9063 | 0.8725 | 0.8797 | 0.8955 | 0.8959 | 0.8923 | 0.8261 | 0.8396 | 7.7100E-4 |
| | RMSE | 4.9014 | 4.3107 | 5.1268 | 5.0742 | 4.9035 | 4.5375 | 4.2713 | 5.3346 | 5.1742 | 0.1495 |
| Proposed | PLCC | 0.9528 | 0.9434 | 0.9324 | 0.9394 | 0.9375 | 0.9075 | 0.8932 | 0.8400 | 0.8510 | 1.7167E-3 |
| | SRCC | 0.9359 | 0.9279 | 0.9171 | 0.9186 | 0.9308 | 0.9026 | 0.8924 | 0.8398 | 0.8406 | 1.3584E-3 |
| | RMSE | 3.7576 | 3.8754 | 4.3337 | 4.1297 | 3.9611 | 4.3948 | 4.4635 | 5.3053 | 5.1814 | 0.2985 |

TABLE IX
PERFORMANCE INDICES OF CROSS VALIDATION OF THE PROPOSED METHODS WITH SIX TYPES OF DISTORTION TYPES USING THE SIQAD AND SCID

| | (a) Training with SIQAD | | | (b) Training with SCID | | |
|---|---|---|---|---|---|---|
| | PLCC | SRCC | RMSE | PLCC | SRCC | RMSE |
| GN | 0.7768 | 0.6881 | 7.9149 | 0.8797 | 0.8606 | 7.0940 |
| GB | 0.7563 | 0.6989 | 7.7699 | 0.8357 | 0.8344 | 7.4125 |
| MB | 0.7351 | 0.7269 | 8.2052 | 0.8330 | 0.8229 | 7.4078 |
| CC | 0.7393 | 0.7605 | 8.2429 | 0.8426 | 0.8416 | 7.1574 |
| JPEG | 0.6928 | 0.6427 | 8.5272 | 0.8449 | 0.8451 | 7.6685 |
| J2K | 0.6434 | 0.6440 | 7.7403 | 0.8026 | 0.8223 | 7.7147 |
| Overall | 0.7349 | 0.7294 | 8.0391 | 0.8405 | 0.8321 | 7.4168 |

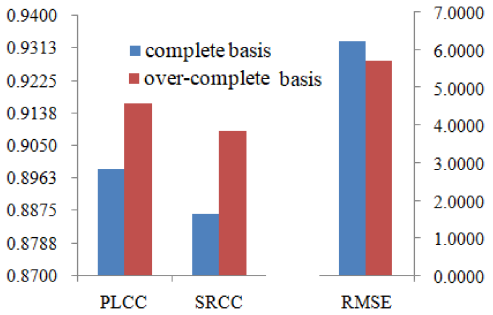



Fig. 10. Comparison of prediction performance with different bases for the SIQAD.

TABLE X
COMPARISON OF PREDICTION PERFORMANCE WITH DIFFERENT ATOM SIZES FOR THE SIQAD

| $p$ | PLCC | SRCC | RMSE |
|---|---|---|---|
| 4×4 | 0.8610 | 0.8438 | 7.2310 |
| 6×6 | 0.8895 | 0.8719 | 6.5195 |
| 8×8 | **0.9162** | **0.9090** | **5.7111** |
| 10×10 | 0.8957 | 0.8885 | 6.3617 |
| 12×12 | 0.8947 | 0.8789 | 6.3655 |
| 14×14 | 0.8727 | 0.8555 | 6.9649 |

TABLE XI
COMPARISON OF PREDICTION PERFORMANCE WITH DIFFERENT ERROR THRESHOLDS FOR THE SIQAD

| $T$ | PLCC | SRCC | RMSE |
|---|---|---|---|
| 1 | 0.9069 | 0.8930 | 6.0002 |
| 3 | 0.9037 | 0.8933 | 6.0974 |
| **5** | **0.9162** | **0.9090** | **5.7111** |
| 7 | 0.9059 | 0.8964 | 5.9469 |
| 9 | 0.8995 | 0.8896 | 6.1084 |
| 11 | 0.8938 | 0.8852 | 6.2064 |

### *D. Influencing Factors*

In this paper, there are two main factors that affect the performance of the algorithm: dictionary training and feature generation. For the former, the basis atom size $p$, and error threshold $T$ need to be optimized and selected. For feature generation, the impacts of the pooling scheme and feature selection should be considered. Here, the sensitivity of each factor is discussed by determining certain values in a proper interval around the determined value while settling the other values. Then, comparative experiments are conducted to validate the influence of the parameter setting.

*1) Dictionary Training:* Since the basis, atom size $p$, and error threshold $T$ directly affect the characteristics of the dictionary, we elaborate on these three factors as follows. First, the results with different bases for the dictionary for the SIQAD are shown in Fig. 10. It can be observed that the performance is improved with the overcomplete basis, as an overcomplete dictionary is more conducive to the expression of highly diversified data, thereby making it easier to capture the inherent essential characteristics of image [51], [52]. Therefore, this paper adopts the overcomplete basis (that is, $k > p$) for dictionary learning to capture more compact and effective features.

Second, Table X and Table XI show the corresponding results with different atom sizes and error thresholds, respectively, in which the values adopted in this paper are highlighted in bold. As mentioned in Section II.B, $p$ denotes the dimension of each atom in the dictionary. In sparse coding, the image patches are reconstructed by atoms, so the size of the atoms directly determines the size and number of patches for the test SCI. Clearly, a larger value of $p$ leads to fewer image patches but more complex sparse coding, which should allow trade-offs between them. As listed in Table X, the performance first increases and then decreases slightly with the increase of the value of $p$. Hence, $p$ is empirically set to 8×8 in the implementation as a default. Similarly, $T$ in Eq. (4) is the predefined error threshold, and a larger value means a lower reconstructed image quality and computational complexity. Considering the characteristics of HVS and sparse coding, if the value of $T$ becomes too large or too small, a loss of features may occur or the features may become similar, which would affect the performance. Here, we set $T$ to 5 as a default according to the actual results, as shown in Table XI.

*2) Feature Generation:* To evaluate the validity of the pooling scheme and contribution of the micro and macro features, we design four schemes by individually adopting max pooling, mean pooling, and log-normal pooling for micro features and

TABLE XII
COMPARISON OF PREDICTION PERFORMANCE WITH DIFFERENT POOLING SCHEMES AND FEATURE VECTORS

| Scheme | SIQAD | | | SCD | | | SCID | | |
|---|---|---|---|---|---|---|---|---|---|
| | PLCC | SRCC | RMSE | PLCC | SRCC | RMSE | PLCC | SRCC | RMSE |
| S_max | 0.6573 | 0.5767 | 10.7295 | 0.6563 | 0.6356 | 1.6644 | 0.6989 | 0.6973 | 10.1206 |
| S_mean | 0.8246 | 0.7878 | 8.0566 | 0.8536 | 0.8445 | 1.1428 | 0.6810 | 0.6628 | 10.3412 |
| S_micro | 0.9071 | 0.8983 | 6.0134 | 0.9140 | 0.9071 | 0.8948 | 0.8579 | 0.8465 | 7.2617 |
| S_macro | 0.8976 | 0.8859 | 6.2965 | 0.9195 | 0.9134 | 0.8683 | 0.8328 | 0.8211 | 7.8270 |
| Proposed | **0.9162** | **0.9090** | **5.7111** | **0.9196** | **0.9123** | **0.8654** | **0.8811** | **0.8730** | **6.7031** |

macro features separately, which are, respectively denoted as S_max, S_mean, S_micro and S_macro. The performances of these schemes on the three databases are reported in Table XII. From this table, we can observe that different pooling schemes have a great influence on performance due to the micro feature aggregation. Clearly, the commonly used max-pooling and mean-pooling will significantly degrade the performance due to the atoms with both positive and negative values. In contrast, the experimental results of S_micro are dramatically improved with the optimized log-normal pooling scheme because the feature codes follow the log-normal distribution law in the target dictionary space, and the distribution shifts to different degrees depending on the type and intensity of the image distortion. Furthermore, micro features can be efficiently aggregated with partial specific feature codes. In addition, the results of S_micro and S_macro in Table XII show that the macro feature has nearly the same effectiveness as the micro feature, which is consistent with the perception of the human eye. Hence, the best performance can be obtained by combining these two features and considering the complementary effect in the target dictionary space.

## IV. CONCLUSION

In this paper, a novel blind image quality assessment method based on a macro-micro model in a target dictionary space is proposed for screen content images (SCIs). Comparing with other existing quality assessment methods, the proposed method has several features. First, tensor decomposition is explored to avoid the loss of color information, and then a target dictionary is learned more effectively with the principal component. Second, a macro-micro model is established to characterize the micro and macro features in the target dictionary space, which can automatically generate complete features without depending on any prior knowledge about the functionalities of visual perception and the mechanism of SCI quality degradation. For the micro features, a log-normal pooling scheme is designed to enhance the effectiveness of feature aggregation by analyzing the distribution of sparse codes. And macro features can be extracted based on the Bernoulli law of large numbers to describe the statistical distribution and quality degradation of SCIs. Finally, the final quality-predictive features used for quality regression are the combination of the micro and macro features. Thorough experimental results on three public SCI databases demonstrated that the proposed method achieves better performance than the existing relevant full-reference and no-reference SCI quality assessment methods, especially in terms of generalization for distortion type and interpretability for feature generation.

## REFERENCES


[1] W. Kuang, Y. Chan, S. Tsang, and W. Siu, "Online-learning-based Bayesian decision rule for fast intra mode and CU partitioning algorithm in HEVC screen content coding," *IEEE Trans. Image Process.*, vol. 29, pp. 170–185, 2020.

[2] I. Kim and J. Jeong, "Hash rearrangement scheme for HEVC screen content coding," *IET Image Process.*, vol. 12, no. 4, pp. 479–484, 2018.

[3] T. Tang, L. Li, and J. Li, "Improved hierarchical quantisation parameter setting method for screen content coding in high efficiency video coding," *IET Image Process.*, vol. 13, no. 8, pp. 1382–1390, 2019.

[4] K. Chung, C. Huang, and T. Hsu, "Adaptive chroma subsampling-binding and luma-guided chroma reconstruction method for screen content images," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 6034–6045, Dec. 2017.

[5] K. Chung, Y. Liang, and C. Wang, "Effective content-aware chroma reconstruction method for screen content images," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1108–1117, Mar. 2019.

[6] L. Zhang, M. Li, and H. Zhang, "Fast intra bit rate transcoding for HEVC screen content coding," *IET Image Process.*, vol. 12, no. 5, pp. 738–744, 2018.

[7] A. Yang, H. Zeng, J. Chen, J. Zhu, and C. Cai, "Perceptual feature guided rate distortion optimization for high efficiency video coding," *Multidimensional Syst. Signal Process.*, vol. 28, no. 4, pp. 1249–1266, 2017.

[8] S. Ma *et al.*, "Nonlocal in-loop filter: The way toward next-generation video coding?," *IEEE Multimedia*, vol. 23, no. 2, pp. 16–26, Apr.–Jun. 2016.

[9] K. Gu *et al.*, "Saliency-guided quality assessment of screen content images," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1098–1110, Jun. 2016.

[10] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[11] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[12] K. Gu *et al.*, "No-reference quality assessment of screen content pictures," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4005–4018, Aug. 2017.

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

[14] Y. Fang, J. Yan, L. Li, J. Wu, and W. Lin, "No reference quality assessment for screen content images with both local and global feature representation," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1600–1610, Apr. 2018.

[15] 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 Trans. Multimedia*, vol. 21, no. 8, pp. 2057–2070, Aug. 2019.

[16] J. Yang, J. Liu, B. Jiang, and W. Lu, "No reference quality evaluation for screen content images considering texture feature based on sparse representation," *Signal Process.*, vol. 153, pp. 336–347, 2018.

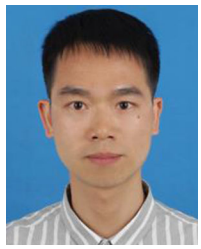
[17] W. Zhou *et al.*, "Local and global feature learning for blind quality evaluation of screen content and natural scene images," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2086–2095, May 2018.

[18] J. Wu, Z. Xia, H. Zhang, and H. Li, "Blind quality assessment for screen content images by combining local and global features," *Digital Signal Process.*, vol. 91, pp. 31–40, 2019.

[19] J. Chen, L. Shen, L. Zheng, and X. Jiang, "Naturalization module in neural networks for screen content image quality assessment," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1685–1689, Nov. 2018.

[20] X. Jiang, L. Shen, G. Feng, L. Yu, and P. An, "Deep optimization model for screen content image quality assessment using neural networks," CoRRabs/1903.00705, 2019. [Online]. Available: http://www.cs.technion.ac.il/users/wwwb/cgi-bin/tr-info.cgi/2008/CS/CS-2008-08

[21] G. Yue *et al.*, "Blind quality assessment for screen content images via convolutional neural network," *Digit. Signal Process.*, vol. 91, pp. 21–30, 2019.
[22] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
[23] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.
[24] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[25] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. Int. Conf. Comput. Vision*, 2011, pp. 543–550.
[26] D. Mandal and S. Biswas, "Generalized coupled dictionary learning approach with applications to cross-modal matching," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3826–3837, Aug. 2016.
[27] Y. Bai, M. Yu, Q. Jiang, G. Jiang, and Z. Zhu, "Learning content-specific codebooks for blind quality assessment of screen content images," *Signal Process*, vol. 161, pp. 248–258, 2019.
[28] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2009, pp. 1794–1801.
[29] Y. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2010, pp. 2559–2566.
[30] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit," 2008 Technical Report CS Technion. [Online]. Available: http://www.cs.technion.ac.il/users/wwwb/cgi-bin/tr-info.cgi/2008/CS/CS-2008-08
[31] Z. Chen and P. Wu, "Strong laws of large numbers for Bernoulli experiments under ambiguity," in *Proc. NL-MUA*, 2011, pp. 19–30.
[32] P. Nowak and O. Hryniewicz, "Strong laws of large numbers for IVM-events," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 12, pp. 2293–2301, Dec. 2019.
[33] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
[34] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4408–4421, Nov. 2015.
[35] S. Wang *et al.*, "Subjective and objective quality assessment of compressed screen content images," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 6, no. 4, pp. 532–543, Dec. 2016.
[36] Z. Ni *et al.*, "SCID: A database for screen content images quality assessment," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, 2017, pp. 774–779.
[37] L. Krasula, K. Fliegel, P. L. Callet, and M. Klima, "On the accuracy of objective image and video quality models: New methodology for performance evaluation," in *Proc. 8th Int. Conf. Quality Multimedia Exp. (QoMEX)*, 2016, pp. 1–6.
[38] A. Aldahdooh *et al.*, "Improved performance measures for video quality assessment algorithms using training and validation sets," *IEEE Trans. Multimedia*, vol. 21, no. 8, pp. 2026–2041, 2019.
[39] P. G. Gottschalk and J. R. Dunn, "The five-parameter logistic: A characterization and comparison with the four-parameter logistic," *Analytical Biochemistry*, vol. 343, no. 1, pp. 54–65, 2005.
[40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[41] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[42] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, Oct. 2014.
[43] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[44] K. Gu *et al.*, "Evaluating quality of screen content images via structural variation analysis," *IEEE Trans. Visualization Comput. Graph.*, vol. 24, no. 10, pp. 2689–2701, Oct. 2018.
[45] Z. Ni *et al.*, "A gabor feature-based quality assessment model for the screen content images," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4516–4528, Sep. 2018.
[46] Y. Zhang, D. M. Chandler, and X. Mou, "Quality assessment of screen content images via convolutional-neural-network-based synthetic/natural segmentation," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5113–5128, Oct. 2018.
[47] Y. Fu *et al.*, "Screen content image quality assessment using multi-scale difference of gaussian," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2428–2432, Sep. 2018.
[48] 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. 5285–5295, 2019.
[49] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.
[50] P. Ye, J. Kumar, and D. Doermann, "Beyond human opinion scores: Blind image quality assessment based on synthetic scores," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2014, pp. 4241–4248.
[51] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Comput.*, vol. 12, no. 2, pp. 337–365, 2000.
[52] A. Agarwal, A. Anandkumar, P. Jain, P. Netrapalli, and R. Tandon, "Learning sparsely used overcomplete dictionaries," in *Proc. Conf. Learning Theory, COLT*, 2014, pp. 123–137.



**Yongqiang Bai** received the B.S. and the M.S. degrees from Zhengzhou University, China, in 2006 and 2009 respectively, and the Ph.D. degree in signal and information processing from Ningbo University, China, in 2019. He is now a researcher with the College of Information and Intelligence Engineering, Zhejiang Wanli University, China. His research interests mainly include data hiding, multimedia communication and image processing.

**Zhongjie Zhu** received the Ph.D. degree in electronics science and technology from Zhejiang University, China, in 2004. He is currently a Professor with the College of Information and Intelligence Engineering, Zhejiang Wanli University, China. His research interests mainly include video compression and communication, image analysis and understanding, watermarking and information hiding, and 3D image signal processing.

**Gangyi Jiang** (Senior Member, IEEE) received the M.S. degree from Hangzhou University, China, in 1992, and the Ph.D. degree from Ajou University, Korea, in 2000. He is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, China. His research interests mainly include digital video compression and communication, multi-view video coding, image-based rendering, and image processing.

**Huifang Sun** (Fellow, IEEE) Graduated from Harbin Military Engineering Institute, Harbin, China, and received the Ph.D. degree from the University of Ottawa, Ottawa, ON, Canada. He was an Associate Professor with Fairleigh Dickinson University, in 1990. He joined Sarnoff Corporation, in 1990 as a Member of technical staff and was promoted to a Technology Leader of Digital Video Communication. In 1995, he joined Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA, and was promoted as the Vice President and Deputy Director, in 2003 and currently is a Fellow of MERL. He has coauthored two books and published more than 150 journal and conference papers. He holds more than 61 U.S. patents. He was an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and was the Chair of Visual Processing Technical Committee of the IEEE Circuits and System Society. He received the Technical Achievement Award for optimization and specification of the Grand Alliance HDTV Video Compression Algorithm, in 1994 at Sarnoff Lab.