Abstract  Blind image quality assessment (BIQA) methods can measure the quality of distorted images even without referencing the original images. This property is indispensable in the image processing field because reference images are normally not available in practice. Unlike the existing trained models, in our work, the training process is constructed as an end-to-end learning mechanism that minimizes the loss between the predicted score and the ground-truth score of the human vision system (HVS). Moreover, a convolutional neural network (CNN) takes distorted images as input and outputs the related score for each image. In this paper, we evaluate the proposed method on six publicly available benchmarks and the cross-database validation performance on the LIVE, CSIQ and TID2013 databases. The experimental results show that our proposed method outperforms other state-of-the-art methods.

Keywords: blind image quality assessment, convolutional neural network

1. Introduction

Nowadays, a huge number of images are produced daily for various purposes, for example, forecasting weather, detecting diseases and monitoring criminals. For these reasons, it is very important to keep the quality of such images at an acceptable visual level for end users after production and transmission. Furthermore, accurate measurement of the image quality is an important step in many image-based applications. To achieve this goal, effective image quality assessment (IQA) algorithms are necessary and have recently become a very hot research topic.

IQA methods can be categorized into subjective and objective methods. As human observers are the most important users in most multimedia applications, the most accurate and reliable way of assessing the quality of images is through subjective evaluation. Moreover, humans can efficiently assess the quality of images without using any reference images. However, subjective evaluations are expensive and time-consuming to apply in real-world applications. Therefore, the purpose of objective IQA is to design mathematical models that can predict the quality of images accurately and automatically similarly to human observers.

Generally, depending on the availability of a reference image, the IQA metrics can be divided into three groups. The first one is full-reference IQA (FR-IQA), where the information of the original image is fully available. The second one is reduced-reference IQA (RR-IQA), where the information of the original image is partly available. The third one is no-reference IQA (NR-IQA) or blind IQA (BIQA), where the information of the original image is unavailable. State-of-the-art FR-IQA measures, such as PSNR, SSIM [1], MS-SSIM [2], MAD [3], VIF [4] and FSIM [5], achieve a very high correlation with human perception. Nevertheless, the degree of improvement is limited by their insufficient consideration of the human visual system (HVS) properties. Moreover, since information of distorted images is normally not available in reality, BIQA methods are becoming more important than the FR-IQA and RR-IQA methods in the image processing field.

Most successful BIQA approaches use natural scene statistics (NSS)-based features. Mittal et al. proposed a blind or referenceless image spatial quality evaluator (BRISQUE) [6], which exploits an NSS model framework of locally normalized luminance co-
coefficients and measures the quantity of “naturalness” using the parameters of the model. Saad et al. proposed a blind image integrity notator using discrete cosine transform (DCT) statistics (BLIINDS-II) [7], which utilizes the DCT to predict the quality score of images. The Bayesian inference model extracts features using the DCT as the input to predict the quality score. Moorthy and Bovik developed distortion-identification-based image verity and integration evaluation (DIIVINE) [8], in which a wavelet transform using steerable pyramids is computed for two scales and along six orientations to compute NSS features. Mittal et al. [9] proposed a “completely blind” IQA model called the natural image quality valuator (NIQE), which adopts a set of natural-sense statistical features of an image’s space domain to fit a multivariate Gaussian (MVG) model without a human-rated score, taking the distance of the MVG model’s parameter as a measure of the image quality. Zhang et al. [10] extended the NIQE to integrated local NIQE (IL-NIQE) using a richer set of statistics features and a new pooling strategy. In most NSS-based BIQA methods, features are extracted by the wavelet transform or the DCT transform. These methods are usually very slow owing to the use of computationally expensive image transformations. In addition, the HVS has different perception for different locations in one image; thus, a good IQA model should consider the HVS’s perceptual characteristics.

As a recent development in BIQA methods, codebook representation for no-reference image quality assessment (CORNIA) [11], promotes extracting features from the spatial domain, which leads to a significant reduction in computation time. CORNIA demonstrates that it is possible to learn discriminant image features directly from the raw image pixels instead of using handcrafted features. A more recent technique, high order statistics aggregation (HOSA) [12], employs the k-means clustering of normalized image patches and describes them using low- and high-order statistics to obtain a small codebook. In HOSA, a soft assignment is used to build an image representation and support vector regression (SVR) is used for the mapping of features to subjective scores.

Currently, convolutional neural networks (CNNs) are attracting researchers’ attention and have achieved great success in various computer vision tasks, because this technique has shown superior performance on many standard object recognition benchmarks. However, all existing BIQA methods still have restrictions preventing them from obtaining their highest performance.

In this paper, therefore, we construct an end-to-end learning mechanism using a CNN to overcome such restrictions. One of the advantages of CNNs is that they can take raw images as the input and incorporate feature learning into the training process. Thus, in our work, we take distorted images labeled with the mean opinion score (MOS) as input and output the related score for each image. The experimental results show that our proposed method outperforms other state-of-the-art IQA methods.

The remainder of this paper is organized as follows. Section II presents the proposed method in detail. In Sect. III, we evaluate the performance of the proposed method by experiments. Finally, Sect. IV concludes
the paper.

2. Proposed Method

The architecture of our proposed method is shown in Fig. 1. It is an end-to-end learning mechanism that takes the whole image as an input and directly outputs a continuous score and then lets the network accept input images with an arbitrary size. In this paper, we utilize 64×64-size distorted images as inputs. Firstly, images are labeled with their related MOSs and then these labeled images are input to the CNN for supervised training. The usage of the CNN is motivated by the fact that it can learn relevant features from an image at different levels, similarly to a human brain. As the target scores of input images are assigned with MOS scores of humans, the trained CNN can realize reliable objective scores that correspond to the subjective scores obtained via the HVS.

In our CNN, two-dimensional convolution is utilized and the features of the input images are convolved with two kernels with a size of 2×8, where each kernel generates a feature map.

After the convolution operation, we use max pooling to cover the entire image as quickly as possible exponentially. The pooling size is 1×2 with a stride of 2 and it reduces the spatial dimensions of a CNN by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Therefore, the computational performance of the CNN becomes high and the overfitting problem is reduced because there is less spatial information. In our case, max pooling downsamples the features that have been extracted by the convolution layer.

Then, we apply a second convolution layer with 16 kernels of size 4×5 and then a batch normalization layer is inserted in our CNN. Batch normalization allows each layer of a network to learn by itself more independently of other layers. In this layer, batch normalization ensures that there is no activation that has gone excessively high or low. Furthermore, it reduces overfitting because it has a slight regularization effect. Similar to dropout, it adds some noise to the activations of each hidden layer. To increase the stability of a neural network, batch normalization normalizes the output of the previous activation layer by subtracting the batch mean and dividing by the batch standard deviation. Thus, inserting this layer after convolution improves the accuracy and accelerates the convergence of the network model.

The second pooling layer is 1×7, with a stride of 16 and then a second batch normalization layer is applied to improve the accuracy of the model.

In practice, images are naturally non-linear. When we look at any image, we find that it contains a large number of nonlinear features, for example, the transition between pixels, the borders, and the colors. The purpose of applying the rectifier function is to increase the nonlinearity in such images. In our network, therefore, we use a rectified linear unit (ReLU) in the fully connected layers instead of the traditional tanh or sigmoid function. The tanh function is mainly used for classification between two classes. Both the tanh and logistic sigmoid activation functions are used in feed-forward networks. The ReLU enables the network to be trained several times faster than by using traditional activation functions [13]. Note that the ReLU only allows nonnegative values to pass. Thus, we use parametric rectified linear unit (PReLU) in the pooling layer instead of the ReLU because there are many negative features after normalization. Unlike the ReLU, which always outputs 0 for input of less than 0, the PReLU multiplies input of less than 0 by a constant value to output results.

After the second pooling, the resulting feature maps are fully connected with one layer that consists of 142 feature nodes. All the nodes are fully connected, resulting in one node that produces the output.

Finally, the squared error loss function is utilized to output continuous values for each input image. The loss is minimized by using stochastic gradient descent with standard backpropagation [14].

Other CNNs may contain a larger or smaller numbers of feature nodes in the fully connected layer and greater or fewer fully connected layers. Engineers often experiment to figure out the configuration that produces the best results for their model. In our network, we set the layer structure to obtain high-accuracy performance.

In brief, the proposed system in Fig. 1 is a very simple idea but it saves computation time and cost while producing much higher performance for all image quality assessment databases than other previous state-of-the-art systems.

3. Experimental Results and Discussion

In this section, we compare the overall performance of our proposed method with that of other state-of-the-art FR-IQA and BIQA methods. As the FR-IQA methods, we utilize SSIM [1], MS-SSIM [2], PSNR, MAD [3], VIF [4] and FSIM [5]. As the BIQA methods, we utilize BLIINDS-II [7], BRISQUE [6], CORNIA [11], HOSA [12], RankIQa+FT [15], NIQE [9], IL-NIQE [10], DIQaM-NR [16], WaDIQaM-NR [16]
and DIIVINE [8]. The six standard benchmarks that we use are A57, TID2008, TID2013, CSIQ, LIVE and IVC. For detailed performance evaluation, we compare the performance of each type of distortion on the six databases. For cross-database evaluation, a model trained on the whole of the laboratory for image and video engineering (LIVE) database is evaluated on subsets of the Tampere image database (TID2013) and computational and subjective image quality (CSIQ) databases which contain only four types of distortion, white noise, Gaussian blur, JPEG and JPEG2000, shared between the three databases.

The A57 Database [17] has three reference images and 54 distorted images, including six types of distortion — flat allocation with equal distortion contrast at all scales (FA), JPEG compression (JPEG), JPEG2000 compression (JPEG2K), JPEG2000 compression with dynamic contrast-based quantization (JPEG2K+DCQ), Gaussian blurring (GB), and additive Gaussian white noise (WGN) — at five different levels. The subjective quality scores used for this database are DMOSs, ranging from 0 to 1.

The Tampere image database (TID2008) [18] includes 25 reference images and 1700 distorted images. Each reference image is distorted using 17 types of distortion, as shown in Table 1, with four different levels. The subjective quality scores provided for this database are MOSs, ranging from 0 to 9.

The TID2013 [19] includes 25 reference images, 24 types of distortion for each reference image, as shown in Table 2, and five different levels for each type of distortion. The whole database contains 3000 distorted images. The MOS is provided in this database and the scores range from 0 to 9.

The CSIQ database [20] contains 900 distorted images. Each image is distorted by six types of distortion — JPEG compression (JPEG), JPEG2000 compression (JPEG2K), global contrast decrements (CTDs), additive Gaussian white noise (WGN), additive pink Gaussian noise (PGN) and Gaussian blurring (GB) — from four to five different levels. Ratings with a 0-to-1 scale are given in the form of DMOSs.

The LIVE database [21] has 29 reference images and 779 distorted images, including five types of distortion — JPEG2K, JPEG, WGN, GB and transmission errors in the JPEG2000 bit stream using a fast-fading Rayleigh channel model (FF). The subjective quality scores provided in this database are DMOSs, ranging from 0 to 100.

Irvine Valley College (IVC) database [22] has 10 reference images and 185 distorted images, including four types of distortion: GB, JPEG2K, JPEG and locally adaptive resolution (LAR) coding. The subjective quality scores provided in this database are MOSs, ranging from 1 to 5.

To fairly compare the performance, we employ the widely used Pearson linear correlation coefficient

| No | Types                                      |
|----|-------------------------------------------|
| 1  | Additive Gaussian noise                    |
| 2  | Additive noise in color component          |
| 3  | Spatially correlated noise                 |
| 4  | Masked noise                               |
| 5  | High frequency noise                       |
| 6  | Impulse noise                              |
| 7  | Quantization noise                         |
| 8  | Gaussian blur                              |
| 9  | Image denoising                            |
| 10 | JPEG compression                           |
| 11 | JPEG2000 compression                       |
| 12 | JPEG transmission errors                   |
| 13 | JPEG2000 transmission errors               |
| 14 | Non eccentricity pattern noise             |
| 15 | Local block-wise distortions of different intensity |
| 16 | Mean shift (intensity shift)               |
| 17 | Contrast change                            |

Table 2 Types of distortions in TID2013

| No | Types                                      |
|----|-------------------------------------------|
| 1  | Additive Gaussian noise                    |
| 2  | Additive noise in color component          |
| 3  | Spatially correlated noise                 |
| 4  | Masked noise                               |
| 5  | High frequency noise                       |
| 6  | Impulse noise                              |
| 7  | Quantization noise                         |
| 8  | Gaussian blur                              |
| 9  | Image denoising                            |
| 10 | JPEG compression                           |
| 11 | JPEG2000 compression                       |
| 12 | JPEG transmission errors                   |
| 13 | JPEG2000 transmission errors               |
| 14 | Non eccentricity pattern noise             |
| 15 | Local block-wise distortions of different intensity |
| 16 | Mean shift (intensity shift)               |
| 17 | Contrast change                            |

| No | Types                                      |
|----|-------------------------------------------|
| 1  | Additive Gaussian noise                    |
| 2  | Additive noise in color component          |
| 3  | Spatially correlated noise                 |
| 4  | Masked noise                               |
| 5  | High frequency noise                       |
| 6  | Impulse noise                              |
| 7  | Quantization noise                         |
| 8  | Gaussian blur                              |
| 9  | Image denoising                            |
| 10 | JPEG compression                           |
| 11 | JPEG2000 compression                       |
| 12 | JPEG transmission errors                   |
| 13 | JPEG2000 transmission errors               |
| 14 | Non eccentricity pattern noise             |
| 15 | Local block-wise distortions of different intensity |
| 16 | Mean shift (intensity shift)               |
| 17 | Contrast change                            |
Table 3 PLCC and SROCC performance comparison on A57 database: FR-IQA methods are italicized.

| IQA Method | PLCC | SROCC |
|------------|------|-------|
| SSIM       | 0.802 | 0.807 |
| MS-SSIM    | 0.560 | 0.667 |
| PSNR       | 0.635 | 0.619 |
| MAD        | 0.501 | 0.850 |
| VIF        | 0.616 | 0.622 |
| FSIM       | 0.557 | 0.496 |
| BLIINDS-II | -     | -     |
| BRISQUE    | -     | -     |
| CORNIA     | -     | -     |
| HOSA       | -     | -     |
| RankIQA+FT | -     | -     |
| NIQE       | -     | -     |
| IL-NIQE    | -     | -     |
| DIQaM-NR   | -     | -     |
| WaDIQaM-NR | -     | -     |
| DIIVINE    | -     | -     |
| Proposed   | 0.947 | 0.923 |

The PLCC (PLCC) and Spearman’s rank ordered correlation coefficient (SROCC). They measure the correlation between the predicted scores and the ground-truth scores of images. PLCC is defined as

$$PLCC = \frac{1}{N-1} \sum_{n=1}^{N} \left( \frac{x_n - \bar{x}}{d_x} \right) \left( \frac{y_n - \bar{y}}{d_y} \right)$$

(1)

where $N$ is the data size, $x_n$ and $y_n$ are the single subjective and objective scores with index $n$, $d_x$ and $d_y$ denote the standard deviations of the subjective and objective scores, and $\bar{x}$ and $\bar{y}$ denote the mean values of the subjective and objective scores, respectively. SROCC is defined as

$$SROCC = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

(2)

where $D$ is the difference between the subjective and objective score ranks.

In Tables 3-8, we compare the PLCC and SROCC values of the proposed method with those of sixteen other state-of-the-art IQA methods based on six databases for the whole databases. The models, which are trained on 80% of each database, are tested on the other 20% of the database. Through the experimental results, we see that the proposed method significantly outperforms other state-of-the-art blind IQA methods and even full-reference IQA ones for all databases.

In Tables 9 and 10, we compare the PLCC and SROCC values, respectively, of our proposed method with those of other state-of-the-art FR-IQA methods for each type of distortion on the A57 database. For almost all types of distortion, the proposed method is more robust than even state-of-the-art FR-IQA methods.

Likewise, in Tables 11 and 12, we compare the PLCC and SROCC values, respectively, of our proposed method with those of other state-of-the-art FR-IQA methods for each type of distortion on the A57 database. For almost all types of distortion, the proposed method is more robust than even state-of-the-art FR-IQA methods.

Table 4 PLCC and SROCC performance comparison on TID2008 database: FR-IQA methods are italicized.

| IQA Method | PLCC | SROCC |
|------------|------|-------|
| SSIM       | 0.772 | -     |
| MS-SSIM    | 0.839 | -     |
| PSNR       | 0.536 | -     |
| MAD        | 0.831 | -     |
| VIF        | 0.806 | -     |
| FSIM       | 0.871 | -     |
| BLIINDS-II | -     | -     |
| BRISQUE    | -     | -     |
| CORNIA     | 0.837 | 0.813 |
| HOSA       | -     | -     |
| RankIQA+FT | -     | -     |
| NIQE       | -     | -     |
| IL-NIQE    | -     | -     |
| DIQaM-NR   | -     | -     |
| WaDIQaM-NR | -     | -     |
| DIIVINE    | -     | -     |
| Proposed   | 0.987 | 0.986 |

Table 5 PLCC and SROCC performance comparison on TID2013 database: FR-IQA methods are italicized.

| IQA Method | PLCC | SROCC |
|------------|------|-------|
| SSIM       | 0.650 | 0.627 |
| MS-SSIM    | 0.729 | 0.786 |
| PSNR       | 0.639 | 0.639 |
| MAD        | 0.807 | 0.838 |
| VIF        | 0.772 | 0.677 |
| FSIM       | 0.821 | 0.802 |
| BLIINDS-II | 0.628 | 0.536 |
| BRISQUE    | 0.651 | 0.562 |
| CORNIA     | 0.613 | 0.549 |
| HOSA       | -     | 0.728 |
| RankIQA+FT | -     | 0.780 |
| NIQE       | 0.426 | 0.317 |
| IL-NIQE    | -     | -     |
| DIQaM-NR   | 0.855 | 0.835 |
| WaDIQaM-NR | 0.787 | 0.761 |
| DIIVINE    | 0.654 | 0.549 |
| Proposed   | 0.981 | 0.977 |
IQA methods for each distortion type on the TID2008 database. Significantly, the proposed method outperforms even the state-of-the-art FR-IQA methods for all types of distortion.

In Table 13, we compare the SROCC values of our proposed method with those of other state-of-the-art BIQA methods for each distortion type on the TID2013 database to ensure a fair performance comparison. Unfortunately, we cannot show the PLCC performance comparison on TID2013 for each distortion type because there is still no description of such PLCC values in other papers. The experimental results demonstrated that for all types of distortion, the proposed method is more robust than other state-of-the-art BIQA methods.

In Table 14, we compare the SROCC values of our proposed method with those of three other BIQA methods for each distortion type on the CSIQ database because SROCC values have previously been given for the three BIQA methods for each distortion type on the CSIQ database. Through the experimental results, we can observe that our proposed method consistently outperforms other previous cutting-edge BIQA methods for almost all types of distortion on the CSIQ database.

Similarly, in Tables 15 and 16, we compare the PLCC and SROCC values, respectively, of our proposed method with those of other BIQA methods for each distortion type on the LIVE database. The proposed method is found to be considerably more robust than the other BIQA methods for some types of distortion.

In Tables 17 and 18, we compare the PLCC and SROCC values, respectively, of our proposed method with those of other FR-IQA methods for each distortion type on the IVC database. The proposed method clearly outperforms the FR-IQA methods for almost all types of distortion.

In Table 19, we extend the cross-database validation and compare the SROCC performance of the proposed method with other state-of-the-art BIQA methods. The model trained on the whole of the LIVE database is evaluated on subsets of the CSIQ and TID2013 databases for the same types of distortion (JPEG, JPEG2K, GB and WGN) for all three databases. In the experiments, the maximum number of epochs for the CNN is 950, the batch size is 16 and the learning rate is 0.1. Experimental results demonstrated that the proposed method is slightly weak for the CSIQ subset but it shows superior performance to all other methods on the TID2013 subset, thus providing independent cross-database validation.

### Table 6 PLCC and SROCC performance comparison on CSIQ database: FR-IQA methods are italicized.

| IQA Method | PLCC | SROCC |
|------------|------|-------|
| SSIM       | 0.859| 0.876 |
| MS-SSIM    | 0.784| 0.873 |
| PSNR       | 0.800| 0.806 |
| MAD        | 0.935| 0.928 |
| VIF        | 0.925| 0.919 |
| FSM        | 0.800| 0.919 |
| BLINDS-II  | 0.832| 0.780 |
| BRISQUE    | 0.817| 0.775 |
| CORNIA     | 0.781| 0.714 |
| HOSA       | -    | -     |
| RankIQA+FT | -    | -     |
| NIQE       | 0.725| 0.627 |
| IL-NIQE    | 0.865| 0.822 |
| DIQaM-NR   | -    | -     |
| WaDIQaM-NR | -    | -     |
| DIVINE     | -    | -     |
| Proposed   | 0.976| 0.964 |

### Table 7 PLCC and SROCC performance comparison on LIVE database: FR-IQA methods are italicized.

| IQA Method | PLCC | SROCC |
|------------|------|-------|
| SSIM       | 0.906| 0.913 |
| MS-SSIM    | 0.940| 0.952 |
| PSNR       | 0.856| 0.866 |
| MAD        | 0.967| 0.967 |
| VIF        | 0.960| 0.964 |
| FSM        | 0.954| 0.963 |
| BLINDS-II  | 0.927| 0.924 |
| BRISQUE    | 0.931| 0.933 |
| CORNIA     | 0.935| 0.942 |
| HOSA       | -    | -     |
| RankIQA+FT | 0.982| 0.981 |
| NIQE       | 0.908| 0.908 |
| IL-NIQE    | 0.906| 0.902 |
| DIQaM-NR   | 0.972| 0.960 |
| WaDIQaM-NR | 0.963| 0.954 |
| DIVINE     | 0.917| 0.916 |
| Proposed   | 0.982| 0.973 |

4. Conclusion

We developed an end-to-end CNN that can take input images of arbitrary size and directly output their predicted quality scores for BIQA. Experimental results verified that the proposed method achieves superior performance to other state-of-the-art BIQA methods and even full-reference IQA methods not only for the whole database but also for several types of distortion.
Table 8  PLCC and SROCC performance comparison on IVC database: FR-IQA methods are italicized.

| IQA Method | PLCC   | SROCC  |
|------------|--------|--------|
| SSIM       | 0.938  | 0.948  |
| MS-SSIM    | 0.940  | 0.952  |
| PSNR       | 0.870  | 0.876  |
| MAD        | 0.967  | 0.967  |
| VIF        | 0.960  | 0.964  |
| PSIM       | 0.954  | 0.963  |
| BLIINDS-II | -      | -      |
| BRISQUE    | -      | -      |
| CORNIA     | -      | -      |
| RankIQA+FT | -      | -      |
| NIQE       | -      | -      |
| IL-NIQE    | -      | -      |
| DIQaM-NR   | -      | -      |
| DiIIVINE   | -      | -      |
| Proposed   | 0.988  | 0.981  |

Table 9  PLCC performance evaluation of each distortion type on A57 database

| Distortion | MS-SSIM | MAD  | FSIM   | Proposed |
|------------|---------|------|--------|----------|
| FA         | 0.619   | 0.546| 0.623  | 0.960    |
| JPEG       | 0.520   | 0.473| 0.559  | 0.963    |
| JPEG2K     | 0.490   | 0.473| 0.484  | 0.935    |
| JPEG2K+DCQ | 0.456   | 0.524| 0.545  | 0.956    |
| GB         | 0.485   | 0.567| 0.576  | 0.861    |
| WGN        | 0.683   | 0.801| 0.843  | 0.794    |

Table 10  SROCC performance evaluation of each distortion type on A57 database

| Distortion | MS-SSIM | MAD  | FSIM   | Proposed |
|------------|---------|------|--------|----------|
| FA         | 0.672   | 0.656| 0.706  | 0.933    |
| JPEG       | 0.372   | 0.539| 0.672  | 0.967    |
| JPEG2K     | 0.422   | 0.556| 0.622  | 0.917    |
| JPEG2K+DCQ | 0.539   | 0.422| 0.372  | 0.833    |
| GB         | 0.489   | 0.422| 0.389  | 0.600    |
| WGN        | 0.589   | 0.389| 0.522  | 0.833    |

Table 11  PLCC performance evaluation of each distortion type on TID2008 database

| Distortion | MS-SSIM | MAD  | FSIM   | Proposed |
|------------|---------|------|--------|----------|
| 1          | 0.774   | 0.817| 0.783  | 0.952    |
| 2          | 0.798   | 0.827| 0.820  | 0.926    |
| 3          | 0.791   | 0.860| 0.794  | 0.971    |
| 4          | 0.809   | 0.757| 0.769  | 0.914    |
| 5          | 0.852   | 0.893| 0.838  | 0.968    |
| 6          | 0.723   | 0.042| 0.674  | 0.936    |
| 7          | 0.764   | 0.798| 0.783  | 0.974    |
| 8          | 0.908   | 0.923| 0.908  | 0.977    |
| 9          | 0.897   | 0.961| 0.932  | 0.992    |
| 10         | 0.900   | 0.949| 0.925  | 0.992    |
| 11         | 0.850   | 0.973| 0.955  | 0.996    |
| 12         | 0.814   | 0.856| 0.842  | 0.981    |
| 13         | 0.809   | 0.830| 0.788  | 0.979    |
| 14         | 0.666   | 0.824| 0.726  | 0.984    |
| 15         | 0.890   | 0.801| 0.841  | 0.980    |
| 16         | 0.683   | 0.571| 0.670  | 0.921    |
| 17         | 0.589   | 0.257| 0.729  | 0.983    |

Table 12  SROCC performance evaluation of each distortion type on TID2008 database

| Distortion | MS-SSIM | MAD  | FSIM   | Proposed |
|------------|---------|------|--------|----------|
| 1          | 0.803   | 0.839| 0.858  | 0.955    |
| 2          | 0.813   | 0.826| 0.852  | 0.922    |
| 3          | 0.828   | 0.867| 0.848  | 0.971    |
| 4          | 0.814   | 0.734| 0.802  | 0.898    |
| 5          | 0.845   | 0.887| 0.910  | 0.955    |
| 6          | 0.754   | 0.064| 0.746  | 0.937    |
| 7          | 0.796   | 0.816| 0.855  | 0.972    |
| 8          | 0.940   | 0.920| 0.947  | 0.981    |
| 9          | 0.923   | 0.943| 0.960  | 0.989    |
| 10         | 0.897   | 0.928| 0.928  | 0.976    |
| 11         | 0.883   | 0.971| 0.978  | 0.992    |
| 12         | 0.820   | 0.866| 0.871  | 0.982    |
| 13         | 0.839   | 0.840| 0.855  | 0.979    |
| 14         | 0.692   | 0.829| 0.750  | 0.974    |
| 15         | 0.896   | 0.796| 0.848  | 0.978    |
| 16         | 0.719   | 0.517| 0.670  | 0.869    |
| 17         | 0.554   | 0.272| 0.648  | 0.967    |

References

[1] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli: Image quality assessment: From error visibility to structural similarity, IEEE Trans. Image Process., Vol.13, No.4, pp.600-612, 2004.
[2] Z. Wang, E. P. Simoncelli and A. C. Bovik: Multiscale structural similarity for image quality assessment, Proc. 13th European Signal Processing Conference, 2005.
[3] E. C. Larson and D. M. Chandler: Most apparent distortion: Full-reference image quality assessment and the role of strategy, J. Electron. Imag., Vol.19, No.1, pp.011006:1-011006:21, 2010.
[4] H. R. Sheikh and A. C. Bovik: Image information and visual quality, IEEE Trans. Image Process., Vol.15, No.2, pp.430-444, 2006.
Table 13  SROCC performance evaluation of each distortion type on TID2013 database

| Distortion | BLII-NDS-II | BRISQUE | CORNIA | HO-SA | Rank-IQA+FT | Proposed |
|------------|-------------|---------|--------|-------|-------------|----------|
| JPEG2K     | 0.994       | 0.997   | 0.970  | 0.970 | 0.970       | 0.970    |
| JPEG       | 0.994       | 0.997   | 0.970  | 0.970 | 0.970       | 0.970    |
| WGN        | 0.994       | 0.997   | 0.970  | 0.970 | 0.970       | 0.970    |
| GB         | 0.994       | 0.997   | 0.970  | 0.970 | 0.970       | 0.970    |
| FF         | 0.994       | 0.997   | 0.970  | 0.970 | 0.970       | 0.970    |

Table 14  SROCC performance evaluation of each distortion type on CSIQ database

| Distortion | BLII-NDS-II | BRISQUE | CORNIA | Proposed |
|------------|-------------|---------|--------|----------|
| WGN        | 0.905       | 0.925   | 0.746  | 0.817    |
| GB         | 0.892       | 0.903   | 0.917  | 0.963    |
| CTD        | 0.012       | 0.024   | 0.302  | 0.972    |
| PGN        | 0.379       | 0.253   | 0.420  | 0.948    |
| JPEG       | 0.900       | 0.909   | 0.908  | 0.978    |
| JPEG2K     | 0.895       | 0.867   | 0.914  | 0.982    |

Table 15  PLCC performance evaluation of each distortion type on LIVE database

| Distortion | BLII-NDS-II | BRISQUE | CORNIA | Rank-IQA+FT | DII-VINE | Proposed |
|------------|-------------|---------|--------|-------------|----------|----------|
| JPEG2K     | 0.935       | 0.923   | 0.951  | 0.975       | 0.922    | 0.984    |
| JPEG       | 0.968       | 0.973   | 0.965  | 0.986       | 0.921    | 0.983    |
| WGN        | 0.980       | 0.985   | 0.987  | 0.994       | 0.988    | 0.986    |
| GB         | 0.938       | 0.951   | 0.968  | 0.988       | 0.923    | 0.989    |
| FF         | 0.896       | 0.903   | 0.917  | 0.960       | 0.888    | 0.994    |

Table 16  SROCC performance evaluation of each distortion type on LIVE database

| Distortion | BLII-NDS-II | BRISQUE | CORNIA | Rank-IQA+FT | DII-VINE | Proposed |
|------------|-------------|---------|--------|-------------|----------|----------|
| JPEG2K     | 0.929       | 0.914   | 0.943  | 0.970       | 0.913    | 0.966    |
| JPEG       | 0.942       | 0.965   | 0.955  | 0.978       | 0.910    | 0.981    |
| WGN        | 0.969       | 0.979   | 0.976  | 0.991       | 0.984    | 0.976    |
| GB         | 0.923       | 0.951   | 0.969  | 0.988       | 0.921    | 0.910    |
| FF         | 0.889       | 0.887   | 0.906  | 0.954       | 0.863    | 0.966    |

Table 17  PLCC performance evaluation of each distortion type on IVC database

| Distortion | MS-SSIM | MAD | FSIM | Proposed |
|------------|---------|-----|------|----------|
| GB         | 0.858   | 0.971 | 0.890 | 0.979    |
| JPEG2K     | 0.882   | 0.916 | 0.930 | 0.993    |
| JPEG       | 0.915   | 0.940 | 0.960 | 0.993    |
| LAR        | 0.927   | 0.936 | 0.935 | 0.980    |

[12] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu and D. Doermann: Blind image quality assessment based on high order statistics aggregation, IEEE Trans. on Image Process., Vol.25, No.9, pp.4444-4457, 2016.
[13] A. Krizhevsky, I. Sutskever and G. E. Hinton: Imagenet classification with deep convolutional neural networks, Neural Information Process. Systems, Vol.1, pp.1097-1105, 2012.
[14] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner: Gradient-
Table 18  SROCC performance evaluation of each distortion type on IVC database

| Distortion | MS-SSIM  | MAD  | FSIM  | Proposed |
|------------|---------|------|-------|---------|
| GB         | 0.853   | 0.963| 0.963 | 0.988   |
| JPEG2K     | 0.938   | 0.910| 0.962 | 0.973   |
| JPEG       | 0.914   | 0.896| 0.981 | 0.979   |
| LAR        | 0.947   | 0.949| 0.886 | 0.961   |

Table 19  SROCC performance in cross-database evaluation: All models are trained on the whole of the LIVE database and evaluated on subsets of the CSIQ and TID2013 databases for the four types of distortion shared with the LIVE database.

| Method      | CSIQ  | TID2013 |
|-------------|-------|---------|
| BRISQUE     | 0.899 | 0.882   |
| CORNIA      | 0.899 | 0.892   |
| DIQaM-NR    | 0.908 | 0.867   |
| WaDIQaM-FR  | 0.866 | 0.872   |
| Proposed    | 0.781 | 0.969   |

Yadanar Khing received her B.C.Sc. (Hons.) and M.C.Sc. degrees in computer science from University of Computer Studies, Yangon, Myanmar, in 2010 and 2012, respectively. In 2013, she joined Computer University, Pathein, Myanmar as a tutor. In 2014, she joined University of Computer Studies, Yangon (UCSY), Myanmar as a researcher. She is currently pursuing her Ph.D. degree in the field of digital image processing at Saitama University.

Yosuke Sugiuura received his B.E., M.E., Ph.D. degrees from Osaka University, Osaka, Japan in 2009, 2011, and 2013, respectively. In 2013, he joined Tokyo University of Science, Tokyo, Japan. In 2015, he joined Saitama University, Saitama Japan, where he is currently an Assistant Professor. His research interests include digital signal processing, adaptive filter theory, and speech information processing.

Tetsuya Shimamura received his B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1986, 1988, and 1991, respectively. In 1991, he joined Saitama University, Saitama, Japan, where he is currently a Professor. He was a visiting researcher at Loughborough University, U.K. in 1995 and at Queen’s University of Belfast, U.K. in 1996, respectively. Prof. Shimamura is an author and coauthor of six books. He serves as an editorial member of several international journals and is a member of the organizing and program committees of various international conferences. His research interests are in digital signal processing and its application to speech, image, and communication systems.

(Received February 25, 2019; revised April 1, 2019)