

## RESEARCH ARTICLE

# No Reference Image Quality Assessment Based on DCT and SOM Clustering

MOHAMMADREZA ZAMANI<sup>1</sup> AND FARAH TORKAMANI AZAR<sup>1</sup>

Faculty of Electrical Engineering, Shahid Beheshti University, Tehran 19839-69411, Iran

Corresponding author: Farah Torkamani Azar (f-torkamani@sbu.ac.ir)

**ABSTRACT** The quality of images and videos is an important research topic due to their wide applications. The research should match the human subjective evaluation of quality. There are three types of objective image quality assessment: full-reference, reduced-reference, and no-reference. No-reference image quality assessment is the most realistic because distorted images often have no reference. In this article, we propose an algorithm that uses the discrete cosine transform (DCT) of patches to extract a feature vector. We apply the Self-Organizing Map (SOM) at two levels. In the first level, we classify the patches into three types: good quality, noisy, and blurred. In the second level, we further classify the noisy and blurred patches into weak and severe distortions using two separate SOMs. Then, we used a straightforward neural network with supervised back-propagation to adjust the number of distorted patches of five classes in one image to assign a quality score. Our method of training Self-Organizing Maps (SOMs) with reference, noisy, and blurred image patches allowed us to estimate the quality scores of other types of distortions that matched subjective scores equally well. The Spearman Rank Order Correlation Coefficient (SROCC) performance of our method, when measured against the subjective scores of degraded images that were used in training, lies in the range of 0.88-0.92. Similarly, for other degradation types that were examined (10 in total), the SROCC performance is in the range of 0.76-0.89, which is higher than other methods. The experiments show that our scores are consistent with the subjective scores.

**INDEX TERMS** Feature vector, image quality assessment (IQA), no-reference image quality assessment.

## I. INTRODUCTION

The importance of images and their applications in our daily lives has grown significantly thanks to technological advancements. Yet, images may become distorted during compression, acquisition, enhancement, and transmission processes. To assess image quality, we use Image Quality Assessment (IQA) methods, which fall into two categories: subjective and objective. Subjective IQA methods, namely the mean opinion score (MOS) and difference mean opinion score (DMOS), rely on human opinions and integers to determine image quality, such as excellent, fair, or bad. Nonetheless, subjective methods are time-consuming and costly as they require human assessments. Objective metrics were developed as a replacement for subjective experiments.

The associate editor coordinating the review of this manuscript and approving it for publication was Jeon Gwanggil<sup>1</sup>.

Objective IQA algorithms predict image quality automatically and aim to minimize deviation from subjective image quality results. Based on the use of the reference image, objective IQA algorithms fall into three categories: Full Reference IQA (FR-IQA), Reduced Reference IQA (RR-IQA), and No Reference IQA (NR-IQA).

Objective measures such as mean square error (MSE) and peak signal-to-noise ratio (PSNR) have been found to have poor correlation with subjective quality measures [1], [2]. As a result, Full-Reference Image Quality Assessment (FR-IQA) methods are often used to evaluate image quality. These methods take into consideration factors such as the human visual system (HVS), image structure, and image statistics.

On the other hand, Reduced-Reference Image Quality Assessment (RR-IQA) methods do not require a full reference image to assess the quality of a distorted image.

Rather, they only need some information and features from the reference image [3], [4]. Assessing the quality of images in the real world can be challenging as the reference image is often not available. This is particularly true in cases of image restoration where distorted photos and videos are the only available records [5]. To overcome this challenge, No-Reference Image Quality Assessment (NR-IQA) methods have been developed. These methods predict the quality of the distorted image without any use or knowledge of its reference. NR-IQA methods can be classified into two types: distortion-specific and general-purpose. Distortion-specific methods evaluate the quality of images in a targeted way, based on the types of distortion present in the images, such as jpeg, blur, and contrast. Therefore, these methods require awareness of the specific types of distortion present in the images. For example, the width of edges and the edge sharpness criterion are used to express a Just-Noticeable Blur (JNB) model to evaluate the quality of distorted blur images [6], [7]. In [8], Zoran et al. analyzed the scale invariance property of images to estimate the quality of noisy images. Liu et al. presented a noise-level estimation model for images in [9]. This algorithm was used not only to estimate the quality of an image but also in the process of image denoising. Gu et al. presented a quality evaluation metric for contrast-distorted images in [10]. This metric is based on maximizing global and local information. It is common for images to have multiple distortions, and the type of these distortions is often unknown. Therefore, research is currently focused on developing general-purpose methods that can be used in real-world applications. These methods do not require any prior knowledge of the types of distortions in the images. Several studies have proposed different approaches to estimating the quality of images. Selecting features from networks is one of the important points for evaluating the quality of images. To remove harmful features using contrastive learning, the IQA metric (QFM-IQM) [11] was proposed. In [12], the authors have evaluated the quality of the distorted image by measuring three aspects of structure, naturalness, and perception. This paper focuses on capturing changes in the structure of distorted images as the first step in quality assessment. According to a study (LPSI), [13] statistical features are extracted from the binary patterns of local image structures to reduce the feature space dimensions and enable the method to generalize to different distortion types. In contrast, the HOSA study extracts local normalized image patches as local features through a network, constructs a codebook using the K-means clustering algorithm, and evaluates the image quality by calculating the mean, diagonal covariance, and skewness of the clusters [14]. In the other study, [15], a novel algorithm was proposed to assess image quality by generating a distortion map for natural scenes using the convolutional autoencoder (CAE) algorithm. The researchers in the study [16] utilized the technique of pixel-pair spatial correlation to select high-quality image patches from the LIVE II dataset [17]. Subsequently, they

applied the FastICA algorithm to learn the independent component quality features from these patches. Finally, the quality of the patches was obtained by determining the independent component coefficients from each image patch. In other research, Zhang and colleagues utilized the DBCNN to enhance image quality. They employed deep convolutional networks (CNN) to evaluate image quality for both synthetically and authentically distorted images [18]. In the OU BIQA study, [19], the authors for visual perception analyze natural scene statistics (NSS) and perceptual features of the human brain. For this purpose, a pristine multivariate Gaussian (MVG) model is used as reference information for quality assessment. Recently, the authors in [20] proposed another method called VNM, which involves designing a visual model that extracts features from test images that are similar to the human visual system and can be utilized for various applications. In the FQA-Net study, [21], an efficient neural network structure was used for blind image quality assessment. This neural network includes a convolution layer and a standard deviation measurement layer. The advantage of the FQA-Net model is that it reduces the number of parameters and output dimensions during the training process. In [22], Yang et al. did not use the mean opinion scores to train the model. This method uses some joint spatial and transform features as quality reduction criteria. After thorough analysis, both of these features are used to remove redundancy. In the study, [23], a model is used that integrates the three sections of feature extraction, feature selection, and regression using a support vector machine (SVM). The suggested model reduces the size of the feature space and enhances the performance of the regression model by utilizing the information gain attribute technique.

In this article, we present the proposed idea based on image patch DCT which is classified using SOM in Section II, followed by obtaining one quality score for the whole image. The experimental results are in Section III. Finally, we conclude in Section IV.

## II. THE PROPOSED IDEA

In this study, our first focus was on extracting suitable features that could distinguish different image types based on their content. We then presented a clustering-based method that can identify the type of image distortion and finally evaluated the image quality score to rank the images.

### A. PRE-PROCESSING AND FEATURE EXTRACTION

To accurately analyze the features of color images, which are in size  $M \times N \times 3$ , it is helpful to split them into smaller, non-overlapping local patches with dimensions of  $r \times r \times 3$ . We began this process by starting from the top left corner of the image. This approach is effective because it allows us to overcome the challenge of large image dimensions, which can make feature analysis more difficult. The value of  $r$  affects the computation complexity and the visibility of the patterns.

If  $r$  is too small, the complexity will be low, but the patterns will not be clear. If  $r$  is too large, the complexity will be high, but the patterns will be more visible. Hence, in this research, we chose  $r = 64$  to balance the value of  $r$ . All patches in this process have equal importance, regardless of their position. In [24], the authors proposed the use of 2-D Discrete Cosine Transform (DCT) to extract features from each color component of the image.

$$F(u, v) = \alpha(u, v) \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) \cdot \cos\left(\frac{\pi(2m+1)u}{2M}\right) \cos\left(\frac{\pi(2n+1)v}{2N}\right);$$

$$\alpha(u, v) = \begin{cases} \frac{1}{\sqrt{N}} & \text{if } u = 0, \quad 0 \leq v \leq N - 1 \\ \frac{1}{\sqrt{M}} & \text{if } v = 0, \quad 0 \leq u \leq M - 1 \\ \frac{1}{\sqrt{MN}} & \text{if } 0 < u \leq M - 1, \quad 0 < v \leq N - 1 \end{cases}$$

(1)

where  $f(m, n)$  is the image of size  $M \times N$  and  $F(u, v)$  is DCT of  $f(m, n)$ . Specifically, they computed three DCT matrices, denoted as  $P^j$  in size  $r \times r$ ,  $j = 1, 2, 3$ , of each color patch  $p$  in size of  $r \times r \times 3$ . In this article, we adopt a similar approach and focus on the significance of DCT coefficients in rectangular paths to reduce dimensions. As depicted in Figure 1, we select  $r$  paths in a  $r \times r$  matrix and extract a feature vector of size  $1 \times r$  for each local patch in a rectangular path. The feature vector can be obtained using eq. 2:

$$\mathbf{R}^j(k) = \sum_{(m,n) \in k^{th} path} |P^j(m, n)|;$$

for  $k = 1, \dots, r, \quad j = 1, 2, 3$  (2)

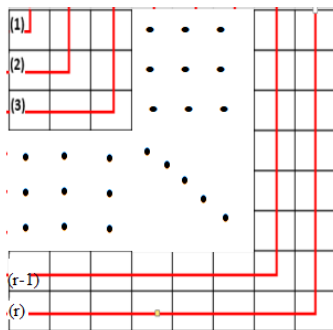


FIGURE 1. The rectangular paths in matrix  $R^j$ .

In our feature vector construction, we choose to utilize the absolute values of DCT coefficients. This is because we consider the weight magnitude to be especially significant. Within matrix  $P^j$ , the (1, 1) element corresponds to low frequency and possesses the greatest energy, resulting in a larger magnitude than the other elements. The other point we

would like to discuss is related to color spaces. The RGB color space is a color model that represents images using three components: red, green, and blue. These components are presented by matrices of equal importance, which require complex computations for analysis. To address this problem, we have utilized the LAB color space instead of RGB. In LAB color space, the lightness component is denoted by "L". L represents the grayscale version of the image and is more dominant than the other two color components. As a result, it requires less computational complexity than the RGB color space. To prove the effectiveness of these feature vectors in distinguishing different types of images, we show three examples of selected patches of reference, additive white gaussian noise (AWGN) distorted, and gaussian blur (GB) distorted images from image "I05.bmp" of the TID2013 database [25] (Further explanations about databases are in the appendix.), and their feature vectors are in Figure 2. Selected patches in the images are marked with a blue box.

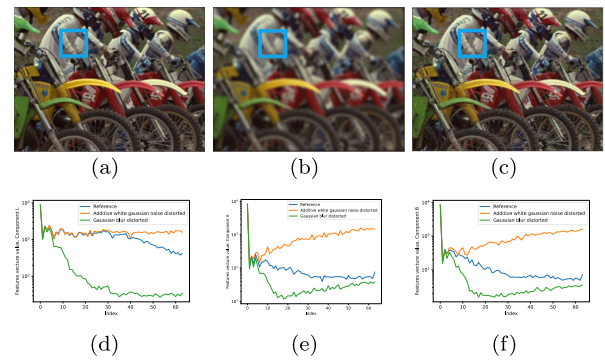
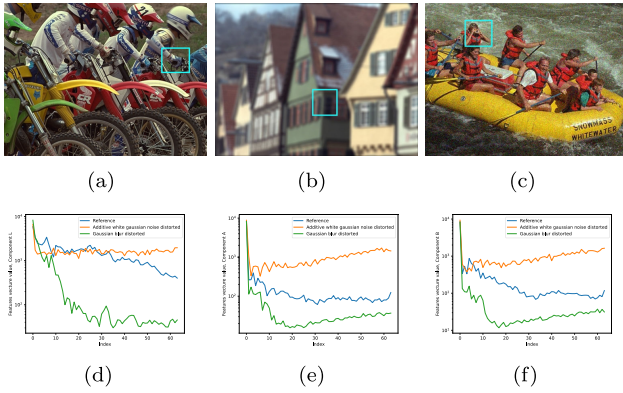


FIGURE 2. (a) reference, image "I05.bmp", (b) GB distorted image, "I05\_08\_05.bmp", (c) AWGN distorted image, "I05\_01\_05.bmp", (d) features vector value of "L" component, (e) features vector value of "A" component, (f) features vector value of "B" component.

The values of the feature vector for each patch depend on the patch's level of detail and frequency. As shown in Figure 2, the AWGN distorted patch has higher values than the reference patch in all three components, especially in the larger indices where the reference patch has very small coefficients. This is because the AWGN distorted patch has more details and frequencies than the reference patch. On the other hand, the GB distorted patch has lower values than the reference patch in all three components, because it lacks high frequencies. This feature extraction method can also distinguish between different types of image distortions, such as AWGN and GB. In Figure 2, three patches have been chosen from a single reference image and its degraded versions. As our paper focuses on no-reference image quality measurement, in Figure 3, we have used three different patches from images in the TID2013 dataset to demonstrate the same results as shown in Figure 2.

As mentioned earlier, we studied the results of Figures 2 and 3 and concluded that the light component in



**FIGURE 3.** (a) reference, image "105.bmp", (b) GB distorted image, "108\_08\_05.bmp", (c) AWGN distorted image, "114\_01\_05.bmp", (d) features vector value of "L" component, (e) features vector value of "A" component, (f) features vector value of "B" component.

the LAB color space plays a significant role. To ensure faster processing, we decided to use only the "L" component for the entire procedure and simulation. This led us to use only one feature vector of size  $1 \times r$  at first. Our next objective is to further reduce the number of feature values.

### B. PATCHES CLUSTERING USING SELF-ORGANIZING MAP

Our goal in this stage is to group the feature vectors obtained from the Discrete Cosine Transform (DCT) in a rectangular pattern for each local patch of the image.

The Self-Organizing Map (SOM), [26] is an effective competitive network for unsupervised learning in clustering algorithms. In Self-Organizing Maps (SOM), the number of output nodes is equivalent to the number of desired clusters. Each output node has a prototype vector of the same size as the inputs, which is initialized with random values close to zero as  $\mathbf{W}_i$ , where  $i$  ranges from 1 to  $L$  as the number of desired clusters. When one input vector is fed into the network, its similarity to each of the  $\mathbf{W}_i$  vectors is calculated. The winner cluster is determined as the cluster with the index  $i^*$  that has the highest similarity score. Now, it is time to adapt the prototype values. In Self-Organizing Maps (SOM), the prototype values of the winning neuron and its neighboring neurons are stimulated in the same direction as the input vector, while the other neurons are inhibited. After complete training, the nearby neurons become prototypes that are similar to each other, and the clusters are changed smoothly. In our method, we have only three clusters, and to have clusters with different prototypes that can distinguish various patch types, we defined training eqs. 3 accordingly.

$$\begin{aligned} \mathbf{W}_{i^*}(t) &= \mathbf{W}_{i^*}(t-1) + \eta(\mathbf{W}_{i^*}(t-1) - \text{input}(t)); \\ \mathbf{W}_j(t) &= \mathbf{W}_j(t-1) - h_{i^*,j}(\mathbf{W}_{i^*}(t-1) - \text{input}(t)); \\ \text{for } j &\neq i^* \\ h_{i^*,j} &= \varphi(\text{dist}(i^*, j)) \end{aligned} \quad (3)$$

where  $i^*$  is the inner index. The given equation represents an iterative learning process, where  $t$  denotes the iteration number.  $\eta$  is the learning rate which may change concerning the iteration number,  $\text{input}(t)$  is the input vector associated with iteration  $t$ , and  $h_{i^*,j}$  represents the learning rate associated with the neighbor distance that inhibits the losers and could be such as  $h_{i^*,j} = e^{-\text{dist}(i^*,j)/\sigma^2}$ .

To achieve this, the SOM is fed with the feature vectors from local patches of images in a random sequence. However, even though we have labels for the patches such as reference, noisy, and blurry, the training procedure does not utilize these labels for the feature vectors. The goal of clustering is to group feature vectors from reference, AWGN distorted, and GB distorted image patches into three separate clusters as effectively as possible. Ideally, clustering occurs when the distance between data points in each cluster is minimized and the distance between clusters is maximized. This process requires a large number of patches of AWGN distorted, GB distorted, and reference images. To train the SOM, we utilized  $64 \times 64$  patches from reference, GB distorted, and AWGN distorted images of the TID2013 [27] and CSIQ [28] datasets. The authors used a total of 605 images with 29040 patches for training. Of these patches, 27.28% were reference patches, 36.36% were GB distorted type, and 36.36% of patches were AWGN distorted type. For more information on SOM settings, please refer to Table 1.

**TABLE 1.** SOM parameters.

Learning rate ( $\eta$ )	0.01
variance of distance ( $\sigma^2$ )	0.5
Epochs	150
Distance metric (dist)	Euclidean

The predefined labels on patches help determine the cluster labels after the training procedure is done. When utilizing clustering algorithms, it is pivotal to identify the ideal number of clusters for a given input dataset to ensure accurate clustering and an optimal cluster number. This can be accomplished by utilizing the Within-Cluster Sum of Squares (WCSS) method. The sum of squares of the distances between each data point in every cluster and its corresponding centroids is represented by the WCSS value. The equation for calculating WCSS is depicted in eq. 4.

$$\text{WCSS} = \sum_{C_k} \left( \sum_{d_i \in C_k} \text{distance}(d_i, C_k)^2 \right) \quad (4)$$

where  $C$  is the cluster centroids and  $d$  is the data point in each cluster.  $k$  is the index of clusters, and  $i$  is the index of data points [29]. Eq. 4 demonstrates that the smaller the WCSS value, the higher the cluster number and compactness, resulting in a smaller distance of the data point from the center of the allocated cluster. The plot in Figure 4 shows the WCSS values for different numbers of clusters when using  $P^j$ ,  $j = 1$  of eq. 2 from patches of the TID2013 dataset.

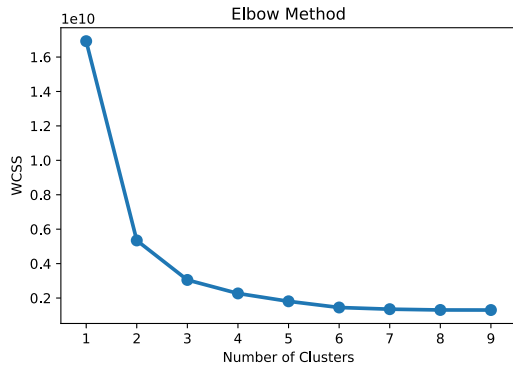


FIGURE 4. WCSS values in TID2013 dataset.

Based on Figure 4, the slope of the WCSS curve fails at the point of three clusters. Since increasing the number of clusters beyond 3 does not significantly decrease the value of WCSS, we chose three clusters for our process at first.

According to the results presented in Figures 2 and 3, the feature vectors exhibit variations in their end feature values. To streamline the method and save time, we have opted to use only the 12 end feature values of the 1 × 12 sized feature vector as input for the unsupervised neural network as a Self Organising Map (SOM). The resulting central vectors of clusters of SOM after training are shown in Figure 5. As can be seen, these central vectors are completely separate from each other and can easily distinguish three different categories.

C. CLASSIFICATION OF THE ENTIRE IMAGE

In the previous step, the feature vectors related to different patches of images were clustered into three clusters: reference, noisy, and blur. In this paper, the reference cluster means a very good quality patch. If all patches of an entire image are applied to a trained SOM, it means that an image will contain  $x_1$  percentage of all patches as reference patches,  $x_2$  percentage as noisy patches, and  $x_3$  percentage as blur patches that  $x_1 + x_2 + x_3 = 1$ . In fact, we will have a  $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$  vector for each image. In this step, the percentage of patches in each cluster was used to mitigate the impact of image size. When working with images, one common task in computer vision is image distortion identification. However, not all patches of an image are placed in a cluster, so a decision must be made regarding the type of distortion.

To classify the whole of one image more accurately, the decision tree classification method is used. The decision tree used the Gini criterion to partition data and measure the quality of the split in classification. In this process, the decision tree groups samples with similar labels or similar target values into right and left subgroups, like the leaves of a tree [30]. In fact, the deeper the leaves, the more appropriate model for data classification can be obtained. Details of the decision tree parameters are shown in Table 2.

This method divides images into three classes: reference, noisy, and blur. By making adjustments to parameters and

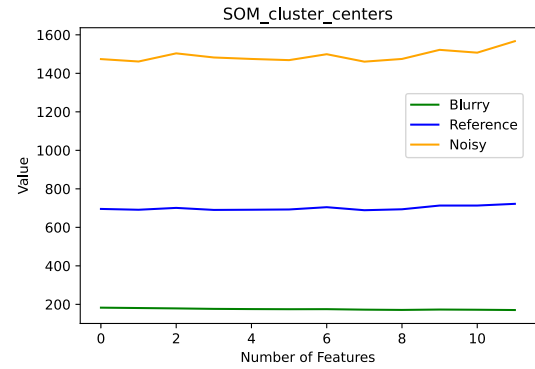


FIGURE 5. Center vectors of SOM clusters after training phase.

TABLE 2. Decision tree classifier parameters.

Criterion	Gini
Splitter	Best
Max-features	3
Max-depth	4

data distribution, the decision tree classification algorithm can improve the classification performance of unbalanced data. The datasets used in this method have five levels of distortion, and images with level 1 distortion (which have a low intensity of distortion) are always included in the collection of reference images.

D. IMPROVED PATCHES CLUSTERING

To further improve, we use another SOM level. The patches that are predicted as noisy or blurred in the first step are re-entered into other SOMs to classify them as having severe or weak destruction. Indeed, two other SOMs, one for noisy patches and another for blurry patches, are used. The expectation is that images with degrees of distortion 2 and 3 will be considered images with weak distortion, while images with degrees of distortion 4 and 5 will be considered images with severe distortion.

The central vectors of the reference patches from the first SOM and two degrees of noisy and blurred patches from the second SOMs are shown in Figure 6. According to Figure 6, the values of SOM cluster center vectors are completely different from each other, which improves the performance of data clustering. Figure 4 is also the confirmation of the continuation of the work, in which point 3 is the breaking point of the curve and point 5 is the beginning of the stabilization of the curve. Following Figure 6, we conducted a two-level classification process because the central vectors of the reference and weak distortions are more similar to each other than to the severe distortion types in clusters.

E. OBTAINING THE QUALITY SCORE FOR THE ENTIRE IMAGE

After the second SOMs, we know the percentage of patches in an image that belong to five clusters. Now, it's

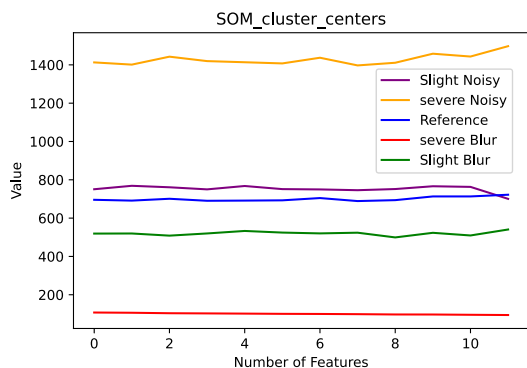


FIGURE 6. SOM five central vector of clusters.

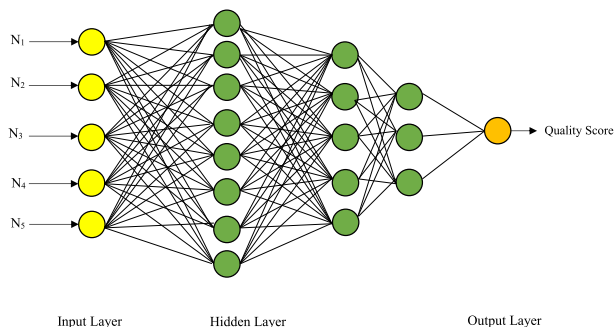


FIGURE 7. The designed neural network to obtain the quality score for each image.

time to obtain the quality score. A neural network for nonlinear regression is used to evaluate the image quality with one quantity score. And so we can calculate and evaluate the correlation between the values obtained from the algorithm and the MOS or DMOS values of datasets. The input of this network is the vectors obtained which specify the percentage of the number (to enhance the algorithm’s resilience for varying image dimensions) of severe noisy, weak noisy, reference, weak blurry, and severe blurry as  $[N_1, N_2, N_3, N_4, N_5]$ . It’s important to note that the percentage of good-quality patches, denoted as  $N_3$ , is obtained from the first SOM and the same  $x_2$ , while the percentage of other types of patches is obtained from two second-level SOMs (Indeed,  $N_1 + N_2 = x_1$  and  $N_4 + N_5 = x_3$  and  $\sum_{i=1}^5 N_i = 1$ ). The target values for MOS and DMOS were normalized between zero and one. In Figure 7, the neural network topology is shown. The hidden layers use the tansig activity function, while the output layer uses logsig. The number of hidden layers is 3. The first, second, and third hidden layers have 8, 5, and 3 nodes, respectively. To speed up convergence, we use the Adam algorithm as a neural network optimizer that changes the learning rates for every parameter according to the data encountered in training, [31]. More information is included in Table 3.

An overview of the overall proposed method is shown in Figure 8.

TABLE 3. Nonlinear regression network parameters.

Loss function	MSE
Optimizer	Adam
Epochs	500

### III. EXPERIMENTAL RESULTS

In this article, we implemented and performed the algorithm using Python 3.7.0. The implementation was done on a computer with an Intel Core i7 CPU having a clock speed of 2.6 GHz. The operating system installed on the computer was Windows 10 (64-bit) with 4 GB of RAM. First, the classification of the reference, AWGN distorted, and GB distorted images of the TID2013 and CSIQ datasets into three classes of reference, noisy, and blurred (after the first level SOM) was evaluated using confusion matrices. The results of the evaluation are presented in Tables 4 and 5 for the two datasets, respectively.

TABLE 4. TID2013 dataset confusion matrix.

Real \ Predicted	Reference	Noisy	Blur
Reference	53	14	8
Noisy	11	89	0
Blur	9	0	91

Based on the information provided in Tables 4 and 5, the highest value for the true-positive parameter in the confusion matrix is achieved. Also, the data related to noisy or blurred classes has the least number of misclassifications in each other’s classes. For example, out of 100 GB images in the TID2013 dataset, 91 images were classified correctly. The remaining images were of good quality but did not fall under the noisy class. Additionally, approximately 75% of the reference images from TID2013 were classified correctly. It is important to note that the incorrect classification of reference images may have been due to their specific contents. This is an interesting result that has also been observed in the CSIQ database, and it can be attributed to the use of SOM clustering with adaptation eqs. 3. Overall, this indicates that the classification process is performing exceptionally well. In further simulations, for the convenience of improving decision making, after classifying images into three classes: noisy, blurry, and reference, we then use specific simple formulas to determine if an image is weak or severely degraded as eqs. 5. The degraded image is:

$$\begin{cases}
 \text{severe noisy} & \text{if image is noisy} \ \& \ N_1 > N_2 \\
 \text{weak noisy} & \text{if image is noisy} \ \& \ N_1 < N_2 \\
 \text{weak blurry} & \text{if image is blurry} \ \& \ N_4 > N_5 \\
 \text{severe blurry} & \text{if image is blurry} \ \& \ N_4 < N_5
 \end{cases} \tag{5}$$

In Tables 6, 7, 8 and 9, the classification confusion matrices for the AWGN distorted and GB distorted images of both the

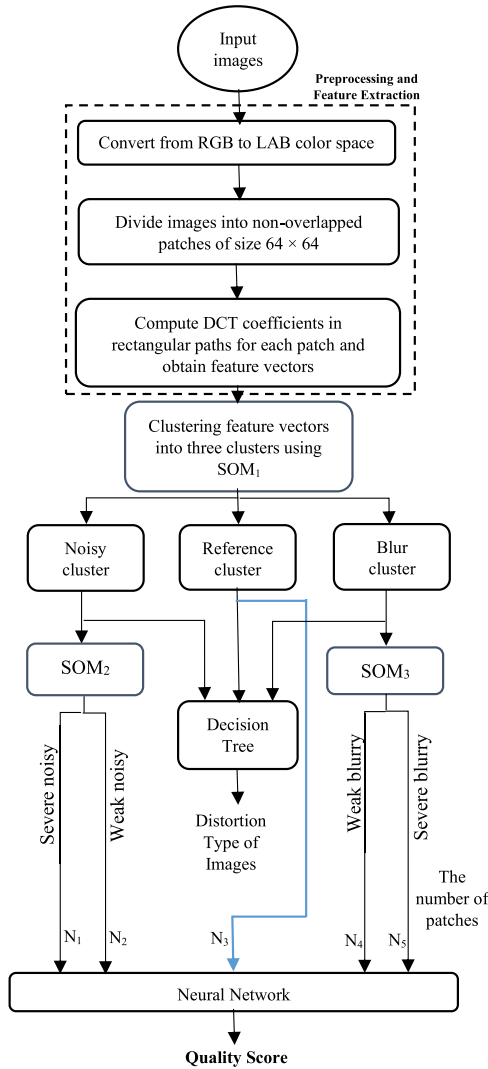


FIGURE 8. The overall flowchart of the proposed algorithm.

TABLE 5. CSIQ dataset confusion matrix.

Real \ Predicted	Reference	Noisy	Blur
Reference	59	10	21
Noisy	18	101	1
Blur	11	2	107

TID2013 and CSIQ datasets, are shown into two categories slight and severe. True-positive parameters on severe data have higher values than on slight data in Tables 6, 7, and 9. It could be because of that, the image quality score suffers due to severe noise or blurring, regardless of content. However, in cases where the degradation is mild, some areas of the image may appear to be severely noisy or blurry. For example, crowded regions with low amounts of noise may seem very noisy, while nearly uniform areas with slight blurring may appear to be severely blurred. The appropriate results obtained in the tables using simple eqs. 5 indicate the efficacy of the extracted feature vectors.

TABLE 6. Confusion matrix of AWGN distorted images of TID2013 dataset.

Real \ Predicted	Weak Noisy	Severe Noisy
Weak Noisy	32	9
Severe Noisy	7	41

TABLE 7. Confusion matrix of GB distorted images of TID2013 dataset.

Real \ Predicted	Weak Blur	Severe Blur
Weak Blur	36	5
Severe Blur	0	50

TABLE 8. Confusion matrix of AWGN distorted images of CSIQ dataset.

Real \ Predicted	Weak Noisy	Severe Noisy
Weak Noisy	35	7
Severe Noisy	12	47

TABLE 9. Confusion matrix of GB distorted images of CSIQ dataset.

Real \ Predicted	Weak Blur	Severe Blur
Weak Blur	37	10
Severe Blur	4	56

The Tables in 10, 11, and 12 report the classification assessment criteria, for each class and the overall classification. Accuracy is the percentage of correct predictions the model makes throughout the entire dataset, whereas recall indicates the percentage of true positive predictions among all the actual positive cases. Precision measures the percentage of true positive predictions among all the positive predictions made by the model. It's important to consider all three metrics in order to get a comprehensive evaluation of the model's performance [32].

It is observed that the overall classification criteria on the dataset named TID2013 are slightly higher than those on the CSIQ dataset. When comparing Tables 11, and 12, it can be noted that the classification of GB distorted images in both the TID2013 and CSIQ datasets met higher classification assessment criteria than that of AWGN distorted images.

To visualize the classification ability of the model for AWGN distorted and GB distorted data, some box plots are displayed in Figure 9. In fact, the range of MOS values corresponding to the images placed in each of the four classified classes at the end of the second level of SOMs is presented. These ranges do not overlap with each other, indicating the accurate classification of the proposed system. For instance, images that are in the weak noise of TID2013 have MOS values in the range of [4.2 – 5.8], while images with severe noise have MOS values in the range of [2.9 – 4.1].

**TABLE 10. Classification criteria on TID2013 and CSIQ datasets.**

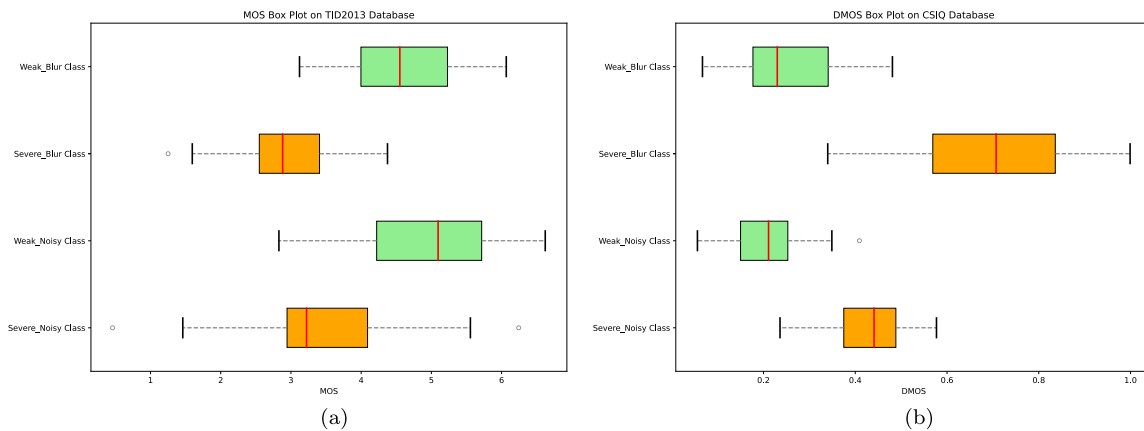
Classes	Recall		Accuracy		Precision	
	CSIQ	TID2013	CSIQ	TID2013	CSIQ	TID2013
Reference	0.67	0.73	0.82	0.85	0.66	0.70
Noisy	0.89	0.86	0.90	0.90	0.84	0.89
Blur	0.83	0.92	0.89	0.94	0.89	0.91
Overall	0.80	0.84	0.87	0.90	0.80	0.83

**TABLE 11. Classification criteria on AWGN distorted images of TID2013 and CSIQ datasets.**

Classes	Recall		Accuracy		Precision	
	CSIQ	TID2013	CSIQ	TID2013	CSIQ	TID2013
Weak Noisy	0.74	0.82	0.81	0.82	0.82	0.78
Severe Noisy	0.87	0.82	0.81	0.82	0.80	0.85
Overall	0.81	0.82	0.81	0.82	0.81	0.82

**TABLE 12. Classification criteria on GB distorted images of the TID2013 and CSIQ datasets.**

Classes	Recall		Accuracy		Precision	
	CSIQ	TID2013	CSIQ	TID2013	CSIQ	TID2013
Weak Blur	0.90	1	0.87	0.94	0.78	0.88
Severe Blur	0.85	0.91	0.87	0.94	0.93	1
Overall	0.88	0.96	0.87	0.94	0.86	0.94



**FIGURE 9. (a) box plot of MOS values of AWGN and GB images classified into weak and severe distortion classes in the TID2013 dataset (b) box plot of DMOS values of AWGN and GB images classified into weak and severe distortion classes in the CSIQ dataset.**

Finally, we calculated the quality scores for the entire images using the whole sections and neural network to compare with their MOS or DMOS available in datasets. Although our system has been trained using only reference, GB, and AWGN images, we analyzed the images of other degradation types in the TID2013 dataset with the proposed ideas. We selected nine other degradation types related to noise and blurring, as listed in Table 13. The authors presented the Spearman’s Rank Order Correlation Coefficient (SROCC) values between our calculated quality measures and their reported Mean Opinion Score (MOS) as eq. 6.

$$SROCC = 1 - \frac{6 \sum_{i=1}^L (d_i)^2}{L(L^2 - 1)} \quad (6)$$

where  $L$  is the number of distorted images,  $d_i$  is the difference between the predicted score and ground-truth score of the  $i^{th}$  image. In Table 13, we compared the SROCC’s of the achieved results of our method with those of different methods that some of them are state-of-the-art, [12], [16], [20], [22], and some of them are based on CNN, [18], [21]. As seen, our method demonstrates a stronger correlation with subjective scores when compared to other methods. As demonstrated, the proposed algorithm outperforms the mentioned methods with a simple and efficient approach. This work was repeated for other distortion types of the CSIQ dataset, and the results were reported in Table 14. Sample images along with their normalized MOS and the predicted quality scores in Figure 10 are displayed.



TABLE 13. SROCC performance on each individual distortion type of TID2013 dataset.

Type	[22]	FQA-Net [21]	VNM [20]	SNP-NIQE [12]	DBCNN [18]	PIH-IQA [16]	DistNet [15]	Proposed
Additive white gaussian noise	0.822	0.893	0.893	0.885	0.790	0.810	0.860	<b>0.909</b>
Additive noise in color components	0.697	0.740	0.644	0.733	0.700	0.756	0.780	<b>0.800</b>
Masked noise	0.694	0.817	0.816	0.740	0.646	0.613	0.560	<b>0.840</b>
Impulse noise	0.761	0.904	0.904	0.799	0.708	0.773	0.720	<b>0.913</b>
Gaussian blur	0.831	0.812	0.808	0.863	0.859	0.782	0.840	<b>0.890</b>
Image denoising	0.702	0.902	0.786	0.612	0.865	0.812	0.320	0.865
Non eccentricity pattern noise	0.017	0.840	0.829	0.014	0.270	0.339	0.330	<b>0.885</b>
Local block-wise	0.146	0.774	0.794	0.032	0.444	0.367	0.340	<b>0.798</b>
Mean shift	0.024	0.756	0.723	0.099	-0.009	0.412	0.410	<b>0.764</b>
Contrast change	0.185	0.861	0.859	0.156	0.548	0.536	0.320	<b>0.882</b>
Change of color saturation	0.104	0.780	0.845	0.106	0.631	0.569	0.520	<b>0.896</b>

TABLE 14. SROCC performance on each individual distortion type of CSIQ dataset.

Type	[19]	DistNet [15]	HOSA [14]	LPSI [13]	Proposed
Additive white gaussian noise	0.836	0.870	0.919	0.666	0.884
Gaussian blur	0.915	0.840	0.926	0.906	0.924
Additive pink gaussian noise	0.902	0.850	-	0.857	0.870



FIGURE 10. Sample images from TID2013 and CSIQ datasets with their normalized MOS or DMOS and predicted values (a) AWGN distorted image with distortion degree 3 from TID2013 dataset, normalized MOS = 0.558 and predicted MOS = 0.585 (b) Impulse noise distorted image with distortion degree 5 from TID2013 dataset, normalized MOS = 0.367 and predicted MOS = 0.353 (c) GB distorted image with distortion degree 2 from CSIQ dataset, normalized MOS = 0.103 and predicted MOS = 0.179.

TABLE 15. SROCC performance on each individual distortion type of LIVE II dataset.

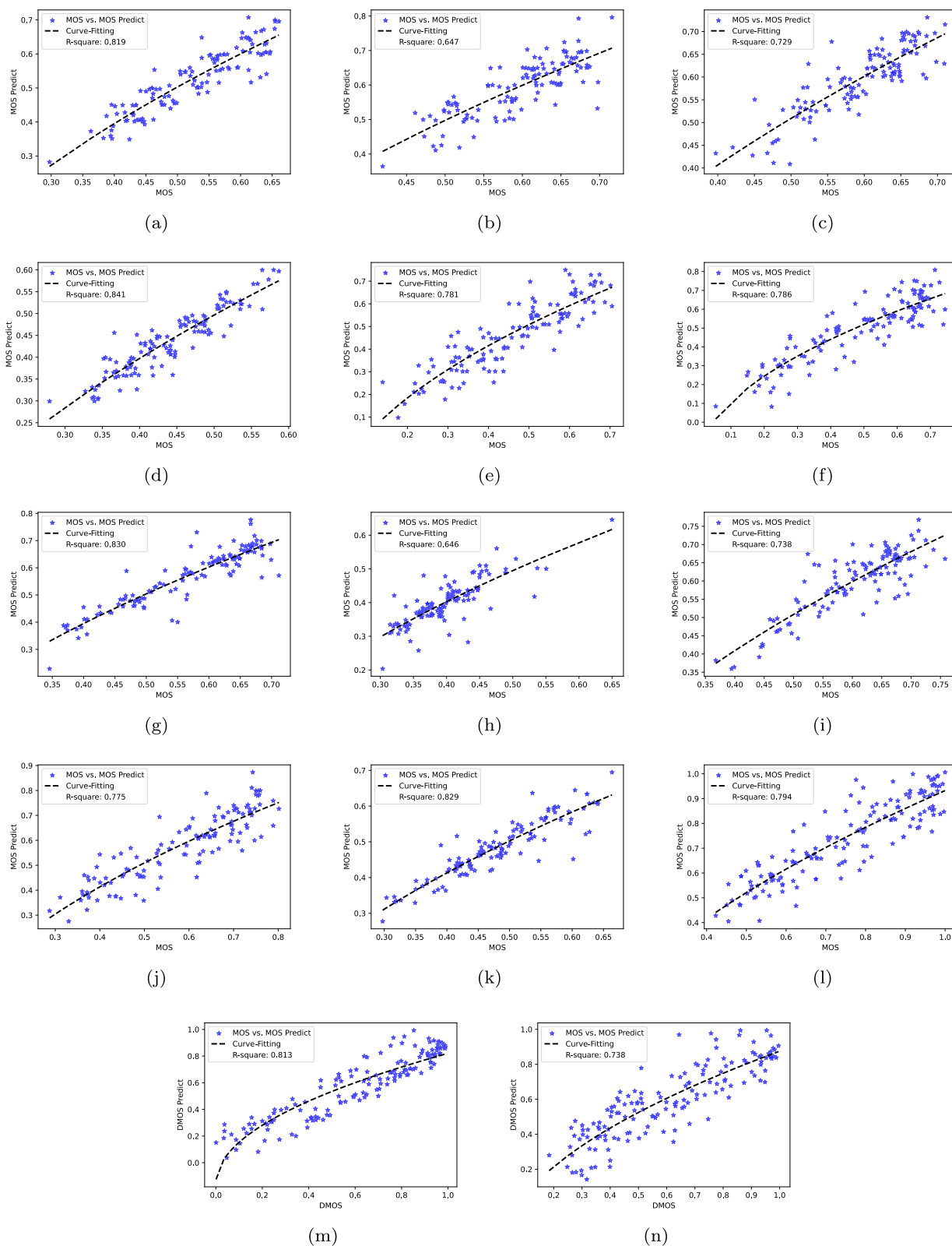
Type	[23]	[19]	DistNet [15]	HOSA [14]	LPSI [13]	Proposed
Additive white gaussian noise	0.972	0.977	0.970	0.973	0.955	<b>0.978</b>
Gaussian blur	0.959	0.939	0.900	0.952	0.915	<b>0.971</b>

Our proposed method accurately predicted the quality score of sample images with minimal error compared to their actual MOS, despite the different distortions they were subjected to. For example, the image 10(a) in Figure 10, that was selected from the TID2013 dataset, has a normalized subjective score of 0.558. By implementing the discrete cosine transform (DCT) and using the features extracted from the patches of this image and forming the feature vector for it, the regression network has predicted a quality score of 0.585 for this image. For another example, the 10(c) image, which has been destroyed by gaussian blur distortion, has a normalized subjective score of 0.103, while our predicted quality score for this image is 0.179, and this shows the effectiveness of our proposed algorithm in evaluating the quality of images and predicting their quality score. In fact,

TABLE 16. The running time of different NR-IQA methods.

Methods	Running Time (s)
[22]	1.73
FQA-Net	0.032
VNM	0.121
[19]	0.903
SNP-NIQE	3.685
PIH-IQA	0.7
DBCNN	0.256
HOSA	0.352
LPSI	0.081
<b>Proposed</b>	<b>0.232</b>

our algorithm obtains very effective features from different patches of the image in order to accurately predict the quality of the entire image. We applied our algorithm equally to all



**FIGURE 11.** Scatter plots of subjective scores versus objective scores delivered by our model on some types of distorted images (a) AWGN distorted images of TID2013 dataset (b) Additive noise in color components distorted images of TID2013 dataset (c) Masked noise distorted images of TID2013 dataset (d) Impulse noise distorted images of TID2013 dataset (e) GB distorted images of TID2013 dataset (f) Image denoising distorted images of TID2013 dataset (g) Non-eccentricity pattern noise distorted images of TID2013 dataset (h) Local block-wise distorted images of TID2013 dataset (i) Mean shift distorted images of TID2013 dataset (j) Contrast change distorted images of TID2013 dataset (k) Change of color saturation distorted images of TID2013 dataset (l) AWGN distorted images of CSIQ dataset (m) GB distorted images of CSIQ dataset (n) Additive pink Gaussian noise distorted images of CSIQ dataset.

**TABLE 17.** Types of distortion used in TID2013 dataset.

No.	Type of distortion
1	Additive Gaussian noise (AWGN)
2	Additive noise in color components is more intensive than additive noise in the luminance 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	JPEG 2000 compression
12	JPEG transmission errors
13	JPEG 2000 transmission errors
14	Non eccentricity pattern noise
15	Local block-wise distortions of different intensity
16	Mean shift (intensity shift)
17	Contrast change
18	Change of color saturation
19	Multiplicative Gaussian noise
20	Comfort noise
21	Lossy compression of noisy images
22	Image color quantization with dither
23	Chromatic aberrations
24	Sparse sampling and reconstruction

patches of an image to consider the effect of each patch on the quality of the image.

In pursuit of simulations, Figure 11 shows scatter plots of the predicted quality scores of images against the subjective quality ranking (MOS or DMOS) using nonlinear logistic monotonic mapping for curve fitting. Axis values in scatter plots are normalized values between zero and one. These plots are for 11 degraded types of the TID2013 dataset and three groups of the CSIQ dataset where each point represents a test image. The figures demonstrate the strong correlation between the subjective and objective scores generated by the proposed method, highlighting its effectiveness. Sometimes, in certain plots, the curve that is fitted appears almost linear, such as in the case of AWGN, additive noise in color components, masked noise, non-eccentricity pattern noise, local block-wise, and mean shift distorted images of the TID2013 dataset. In order to show the efficiency of our method, we also tested another dataset LIVE II [17], and the result of its correlation is still acceptable compared to other methods. The data results in Table 15 confirm this. Despite not using the LIVE II dataset in the training phase, our idea performed well for this new database. This is noteworthy.

To complete our simulations, Table 16 shows the running time of our method compared to other NR-IQA methods. Our method's running time is acceptable and faster than most models.

#### IV. CONCLUSION

An efficient method for measuring the quality of images without a reference is proposed in this paper. The method is based on DCT, SOM clustering, a decision tree classifier,

and a regression nonlinear neural network. The proposed algorithm clusters the feature vectors of image patches into one of three categories: reference, noisy, or blurred using an unsupervised SOM. A decision tree classifier is used to identify the type of distortion in an entire image based on the percentage of the number of patches in SOM clusters. To obtain the overall quality score for an entire image, two additional Self-Organizing Maps (SOMs) are created and trained using only the noisy or blurry patches identified in the first step. The SOMs in the second step are trained to cluster the patches into two categories: weak distortion and severe distortion. Finally, the percentage of patches in the entire image in each cluster (severe noisy, weak noisy, reference, weak blurry, and severe blurry) is used as input for a three hidden layer neural network regression, resulting in a quantity value for the quality score. Compared to other methods, our algorithm has a simple, convenient, and structured process. The simulation results indicate a reliable correlation between the predicted scores and the subjective quality scores for several types of distortions. They also demonstrate that our method is better at predicting scores than other NR-IQA methods. In this study, DCT basis was used to obtain features and comparison. In future work, we will try to use deep learning machines to improve our basis and obtain more effective features.

#### APPENDIX DEFINITION USING DATASETS

##### A. TID2013

The database named TID2013 consists of 25 reference images and 3000 distorted images. The reference images are obtained by cropping from the Kodak Lossless True Color

Image Suite (<https://r0k.us/graphics/kodak/>). The distorted images are created by applying 24 different types of distortions at five different levels to each reference image. Without any compression, the images are saved in bitmap format, and their file names indicate the reference image number, distortion type, and distortion level in the format of “iXX\_YY\_Z.bmp”. The mean opinion score for every distorted image may be found in the “mos.txt” file. The MOS was derived from 971 experiments that observers from Finland, France, Italy, Ukraine, and the United States of America conducted. Each observer performed either 524340 comparisons of the visual quality of distorted images or 1048680 evaluations of the relative visual quality of image pairs. The value ranges of the MOS from 0 (minimal quality) to 9 (maximal quality), with the MSE of each score being 0.018. Higher MOS values correspond to a higher visual quality of the image. Table 17 refers to the 24 degraded types of the TID2013 database.

### B. CSIQ

A Categorical Image Quality (CSIQ) Database is provided by Oklahoma State University’s Image Coding and Analysis Lab. This database consist of 30 original images that have been distorted at four to five different distortion levels using six different types of distortions. The distortions used in CSIQ include JPEG compression, JPEG-2000 compression, global contrast decrements, additive pink Gaussian noise, and Gaussian blurring. In total, there are 866 distorted versions of the original images. A linear displacement of the images across four calibrated LCD monitors arranged side by side with similar viewing distances from the observers served as the basis for the subjective ratings of the CSIQ images by 35 different observers. The database contains 5000 subjective ratings, which are reported as DMOS ranges from 0 (maximum quality) to 1 (minimum quality).

### C. LIVE II

An extensive experiment was conducted by LIVE Laboratory in collaboration with the Department of Psychology at the University of Texas at Austin to obtain scores from human subjects. They used 29 reference images, each with test versions that had five distortion types: white noise in the RGB components, Gaussian blur, JPEG, JPEG2000, and bit errors in JPEG2000 bitstream when transmitted over a simulated fast-fading Rayleigh channel with different levels. Each image was evaluated by about 20–29 human observers, and different subjects in different experiments evaluated each type of distortion using the same equipment. Seven experiments totaling 982 images, out of which 203 were reference images, were reviewed by human subjects. Finally, a Difference Mean Opinion Score (DMOS) value for each distorted image was computed to the full range from 0 (maximum quality) to 100 (minimum quality).

## REFERENCES

- [1] C. Shi and Y. Lin, “Full reference image quality assessment based on visual saliency with color appearance and gradient similarity,” *IEEE Access*, vol. 8, pp. 97310–97320, 2020.
- [2] D. Varga, “Saliency-guided local full-reference image quality assessment,” *Signals*, vol. 3, no. 3, pp. 483–496, Jul. 2022.
- [3] S. Dost, F. Saud, M. Shabbir, M. G. Khan, M. Shahid, and B. Lovstrom, “Reduced reference image and video quality assessments: Review of methods,” *EURASIP J. Image Video Process.*, vol. 2022, no. 1, pp. 1–31, Jan. 2022.
- [4] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, “Reduced-reference image quality assessment in free-energy principle and sparse representation,” *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 379–391, Feb. 2018.
- [5] R. A. Manap and L. Shao, “Non-distortion-specific no-reference image quality assessment: A survey,” *Inf. Sci.*, vol. 301, pp. 141–160, Apr. 2015.
- [6] R. Ferzli and L. J. Karam, “A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB),” *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.
- [7] W. Zhengzi, X. Zhihua, and H. Cuiqun, “A fast quality assessment of image blur based on sharpness,” in *Proc. 3rd Int. Congr. Image Signal Process.*, vol. 5, Oct. 2010, pp. 2302–2306.
- [8] D. Zoran and Y. Weiss, “Scale invariance and noise in natural images,” in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2209–2216.
- [9] X. Liu, M. Tanaka, and M. Okutomi, “Single-image noise level estimation for blind denoising,” *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5226–5237, Dec. 2013.
- [10] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, “No-reference quality metric of contrast-distorted images based on information maximization,” *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4559–4565, Dec. 2017.
- [11] X. Li, T. Gao, X. Zheng, R. Hu, J. Zheng, Y. Shen, K. Li, Y. Liu, P. Dai, Y. Zhang, and R. Ji, “Adaptive feature selection for no-reference image quality assessment using contrastive mitigating semantic noise sensitivity,” 2023, *arXiv:2312.06158*.
- [12] Y. Liu, K. Gu, Y. Zhang, X. Li, G. Zhai, D. Zhao, and W. Gao, “Unsupervised blind image quality evaluation via statistical measurements of structure, naturalness, and perception,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 4, pp. 929–943, Apr. 2020.
- [13] Q. Wu, Z. Wang, and H. Li, “A highly efficient method for blind image quality assessment,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 339–343.
- [14] 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. Image Process.*, vol. 25, no. 9, pp. 4444–4457, Sep. 2016.
- [15] S. V. R. Dendi, C. Dev, N. Kothari, and S. S. Channappayya, “Generating image distortion maps using convolutional autoencoders with application to no reference image quality assessment,” *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 89–93, Jan. 2019.
- [16] C. Zhang, X. Yang, X. Huang, G. Yu, and S. Chen, “No-reference image quality assessment based on quality patches in real time,” *EURASIP J. Image Video Process.*, vol. 2018, no. 1, pp. 1–10, Dec. 2018.
- [17] H. Sheikh. (2005). *Live Image Quality Assessment Database Release 2*. [Online]. Available: <http://live.ece.utexas.edu/research/quality>
- [18] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, “Blind image quality assessment using a deep bilinear convolutional neural network,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 36–47, Jan. 2020.
- [19] Y. Liu, K. Gu, X. Li, and Y. Zhang, “Blind image quality assessment by natural scene statistics and perceptual characteristics,” *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 16, no. 3, pp. 1–91, Aug. 2020.
- [20] H.-W. Chang, X.-D. Bi, and C. Kai, “Blind image quality assessment by visual neuron matrix,” *IEEE Signal Process. Lett.*, vol. 28, pp. 1803–1807, 2021.
- [21] H.-W. Chang, P.-J. Wang, C.-Y. Du, X.-D. Bi, and M.-H. Wang, “FQA-net: An efficient neural network for blind image quality assessment,” *J. Electron. Imag.*, vol. 31, no. 6, Nov. 2022, Art. no. 063027.
- [22] C. Yang, Q. He, and P. An, “Unsupervised blind image quality assessment via joint spatial and transform features,” *Sci. Rep.*, vol. 13, no. 1, p. 10865, Jul. 2023.
- [23] J. Ryu, “Adaptive feature fusion and kernel-based regression modeling to improve blind image quality assessment,” *Appl. Sci.*, vol. 13, no. 13, p. 7522, Jun. 2023.

- [24] F. Torkamani-Azar, "Reducing dimensions of the feature vector of an image based on blocking-DCT," Shahid Beheshti Univ., Iran, Tech. Rep., 2021, doi: [10.36227/techrxiv.16782523.v1](https://doi.org/10.36227/techrxiv.16782523.v1).
- [25] C. Yan, T. Teng, Y. Liu, Y. Zhang, H. Wang, and X. Ji, "Precise no-reference image quality evaluation based on distortion identification," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 17, no. 3, pp. 1–21, Oct. 2021.
- [26] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.
- [27] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, "Color image database TID2013: Peculiarities and preliminary results," in *Proc. Eur. Workshop Vis. Inf. Process. (EUVIP)*, 2013, pp. 106–111.
- [28] D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, Jan. 2010, Art. no. 011006.
- [29] M. Hamka and N. Ramdhoni, "K-means cluster optimization for potentiality student grouping using elbow method," in *Proc. AIP Conf.*, vol. 2578, 2022, Art. no. 060011.
- [30] J. Chen, R. Tan, and Y. Yang, "Research on an innovative feature importance recognition algorithm based on GINI-OOB index," in *Proc. IEEE Int. Conf. Image Process. Comput. Appl. (ICIPCA)*, Aug. 2023, pp. 862–866.
- [31] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," *Neural Comput. Appl.*, vol. 35, no. 23, pp. 17095–17112, Aug. 2023.
- [32] G. M. Foody, "Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient," *PLoS ONE*, vol. 18, no. 10, Oct. 2023, Art. no. e0291908.



**MOHAMMADREZA ZAMANI** received the B.Sc. degree in electrical engineering from Qazvin Islamic Azad University, Iran, in 2017, and the M.Sc. degree in telecommunication engineering from Shahid Beheshti University, Iran, in 2023. His research interests include objective and subjective image quality assessment and machine learning.



**FARAH TORKAMANI AZAR** received the B.S. degree in electrical engineering from the Amirkabir University of Technology, Tehran, Iran, in 1986, the M.S. degree in electrical engineering from Isfahan University of Technology, Isfahan, Iran, in 1991, and the Ph.D. degree from the University of New South Wales, Sydney, NSW, Australia, in 1995. She was an Academic Staff with Isfahan University of Technology, from 1995 to 2000. Since 2000, she has been a member of the Communication Department, Shahid Beheshti University, Tehran, where she is currently an Associate Professor. Her current research interests include the several aspects of signal processing, specially in image processing, compressive sensing, graph signal processing, and deep learning.

• • •