Signal Processing: Image Communication 111 (2023) 116899

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

A CNN-based no reference image quality metric exploiting content saliency

Kamal Lamichhane [a,*], Marco Carli [a], Federica Battisti [b]

[a] Roma Tre University, Department of Industrial, Electronic and Mechanical Engineering, Via Vito Volterra 62, Roma, Italy
[b] University of Padova, Department of Information Engineering, Via Gradenigo 6/B, Padova, Italy

ARTICLE INFO

Keywords:
Artificial intelligence
Convolutional neural network
Deep learning
Image quality assessment
Natural scene statistics
Saliency

ABSTRACT

Assessing the quality of images is a challenging task. To achieve this goal, images must be evaluated by a pool of subjects following a well-defined protocol or an objective quality metric must be defined. In this work, an objective quality metric based on deep neural network is proposed. The metric takes into account the human vision system by computing the saliency map and natural scene statistics features of the image under test. The neural network is composed by two modules: the convolutional layers and the regression units. The first one is trained by using preprocessed distorted images. The feature weights of the first module are smoothed by exploiting the estimated saliency map. The latter module is fit with the ground truth quality scores of the input image and the scaled feature weights obtained from first module by using visual sensitivity factor of image obtained using natural scene statistics features. The performances of the proposed metric have been evaluated by using four datasets: LIVEIQA, TID2013, CSIQ, and KADID10K. The achieved results show the effectiveness of the proposed system in closely matching the predicted quality scores with the ground truth ones.

1. Introduction

In recent years we have seen the proliferation of multimedia services characterized by the fact that the information is shared through videos, images, and audio. Despite advances in both hardware and software, it is worth remembering that imaging devices, storage systems, coding algorithms, and transmission can cause artefacts that affect image quality [1]. The ability to reliably measure the delivered quality is a key element in the development of such systems. In the literature, many authors have dealt with the quality assessment problem, trying to define an all-purpose quality measure matching, as close as possible, the human judgement [2–9].

These methods can be classified into three categories according to the availability of the original content during the evaluation step: Full Reference (FR), Reduced Reference (RR), and No Reference (NR) Image Quality Assessment (IQA). That is, those methods exploit the availability of the original image, of some of its characteristics, or of no information respectively. In FR IQA a two steps procedure is generally adopted: a local quality measurement is followed by a pooling phase [10]. The local quality measurement is performed by evaluating spatial (e.g., Mean Square Error (MSE) or Structural Similarity Index Measurement (SSIM)), scale (pixel or patch based), and orientation features. Both spatial or transform domains have been exploited. FR IQA approaches offer the most competitive results compared to the other approaches [5,6]. Recently, innovative frameworks based on Deep Neural Network (DNN) have been proposed. A DNN is an artificial neural network inspired by human biology trying to mimic the way neurons of the human brain processes and understand inputs from human senses. Convolutional Neural Network (CNN) is a class of DNN that has become dominant in various computer vision tasks. It was originally designed for image analysis. The basic operations in a CNN are convolution and pooling. The first one is responsible for extracting the feature maps from the dataset while the latter is used for reducing the dimensionality of the feature maps [11]. CNN automatically and adaptively learns spatial hierarchies of features through back-propagation and it is designed using multiple building blocks, such as convolutional, pooling, and fully-connected layers. Machine Learning (ML) [12] approaches, either conventional ML or DNN [13] have been proposed leading to improved results. Conventional ML based NR IQA approaches extract low-level features [3,14–16] from distorted images. These features are then used in a ML-based regression step to predict the overall quality score. The NR IQA techniques are generally less accurate in predicting the human judgement when compared to FR IQA. Though, DNN based NR IQA techniques offer very competitive results close to the ones obtained by using FR IQA. In case of DNN-based no-reference Fully Deep Image Quality Assessment (FDIQA) technique, the low and high level features present in the image are automatically extracted during the training of the DNN. The weight of these features is optimized to improve the performance of the quality metric using a backward propagation method. In this work, we propose a DIQASN. It is based on the exploitation of information on the content saliency and Natural Scene Statistics (NSS) estimated from the distorted images.

* Corresponding author.
E-mail addresses: Kamal.Lamichhane@uniroma3.it (K. Lamichhane), marco.carli@uniroma3.it (M. Carli), federica.battisti@unipd.it (F. Battisti).

https://doi.org/10.1016/j.image.2022.116899
Received 8 November 2021; Received in revised form 18 October 2022; Accepted 15 November 2022
Available online 21 November 2022

The saliency map is estimated using the foreground extraction method by selecting the significantly bright pixels of the image. NSS feature reflects the variation of visual perception from the mean value of the visual intensity of the image. In the proposed metric, the CNN is split into two modules: (1) convolutional layers and (2) regression unit. This approach helps to minimize the overfitting problem of the CNN during training, even with small size datasets [6]. In the convolutional layers unit, the features are estimated by the CNN and used in the regression unit for predicting the quality score. To increase the learning capability of the convolutional layers unit, feature weights of convolutional layers are optimized by a loss function. For this purpose, the saliency map of the corresponding input image is used. In the next step, the visual sensitivity factor estimated on the NSS features of distorted images is used as input to the regression unit for improving the DIQASN model performance. To measure the similarities between the predicted subjective scores and ground truth quality scores, two standard statistical measures SRCC, and PLCC are used. LIVE_IQA [17], CSIQ [18], and TID2013 [19] are the three datasets used for training and testing our model. The model's performance have been evaluated with and without considering the saliency map and NSS features for understanding their contribution on the overall performance of the model. Furthermore, the proposed model performance has been evaluated on the large-size KADID10K [20] dataset and compared with state-of-the-art models. The main features of the proposed quality metric are:

1. It is a no-reference approach based on a fully deep neural-network. A pre-processed version of the image and the estimated saliency map are used as input to the learning model.
2. The adoption of the saliency estimation model extracts the visually most relevant area of the image by exploiting the histogram back-projection method [21], the multiscale pyramidal mean shift filter [22], and the *grabcut* [23] algorithm.
3. The obtained results show that the use of visual sensitivity for scaling the feature weights of convolutional layers, obtained using the global average pooling function helps to improve the overall performances.

The rest of the paper is organized as follows: in Section 2 state-of-the-art quality assessment models and the main limitation of CNN are described. In Section 3 the proposed model is introduced. In Section 4 the analysis of the performed experiment for assessing the metric performances is reported. Finally, in Section 5 the concluding remarks are drawn.

2. State of the art

2.1. Full reference IQM

With the advent of ML approaches, new Image Quality Metrics (IQM) based on ML have been developed with comparable or better performance than the well-known Peak Signal to Noise Ratio (PSNR) and SSIM. A FR IQM based on image features extracted from a pre-trained CNN model, AlexNet, is proposed in [24]. The feature maps of the distorted and reference images are extracted at multiple layers, and feature similarity is computed at each layer. A pooling function is then applied to obtain the overall quality score. The benefits of using saliency maps into FR IQMs are investigated in [25]. The comparison is performed among the pixel-based quality metrics (PSNR, MSE, and SSIM) and IQMs exploiting saliency maps obtained using Itti [26], Achanta [27], and GAFFE [28] computational models. The results show that the performance of quality metrics can be improved by adding information on content saliency. The precision of the saliency map, the characteristics of quality metric, and the type of distortions determine the performance gain of the IQM. A deep CNN-based FR model, DCNN, designed using eleven CNN layers having skip connections and two fully connected layers, is used to predict perceptual quality scores in [29]. To train, validate, and test the CNN model, a large-scale dataset is prepared by labeling the images with the human preference probability obtained from distorted and reference images, based on the proposed, pairwise learning framework. Kim et al. [5] propose a CNN-based FR IQM, where the behavior of the HVS is learned from the underlying data distribution of the input database. This model seeks the optimal visual weight based on understanding the database information from the error map between distorted and reference images. A dual-path based CNN is proposed in [30]. The two paths are used to analyze the distorted and the reference image respectively. At first, the same set of characteristics is extracted, then, by sharing the weight between the paths, the characteristics of both paths are concatenated. Finally, a regression unit is used to predict the image quality score.

2.2. Reduced reference IQM

A RR metric, FSI (Free-energy principle and Sparse representation-based Index for image quality assessment), is proposed in [31]. The internal generative model mimicking the perception of the scene by the brain is built on the sparse representation of reference and distorted images. The difference between the entropy of the prediction discrepancies for reference and distorted images is defined as the quality index. In [32], a RR-Entropies Differencing (RRED) approach is proposed by measuring the differences between entropies of wavelet coefficients of reference and distorted images. Wang et al. proposed an IQM based on contourlet transform of images in [33]. Different scales and direction sub-bands of distorted and reference images are considered.

A structural degradation model is proposed in [34] by considering the SSIM. In more details, the quality score of an image is obtained as a Support Vector Machine (SVM) based integration of distance between the structural degradation information of the reference and distorted images. In [35], a stochastic RR-IQM is proposed to assess the quality of an image based on deep learning, namely Restricted Boltzmann Machine Similarity Measure (RBMSim).

2.3. No reference IQM

A NSS based blind image quality evaluator (BRISQUE) model is proposed in [36]. The authors describe the natural scene based statistical model of locally normalized luminance coefficients in the spatial domain and the model for pairwise products of these coefficients. Statistical features are used to demonstrate the relation with the human quality judgement. A mapping feature-to-quality function is defined.

To design a No Reference Image Quality Assessment (NRIQA) model using DNN, the feature extraction methods adopted in conventional NRIQA systems, are generally replaced with deep learning models for automatically collecting low level and high level features. In this direction, in [4] the blind evaluator based on convolution network (BIECON) model is proposed. To cope with the lack of training data and ground truth images, reliable Full Reference Image Quality Assessment (FRIQA) metrics are used as intermediate local targets for each image patch.

A Multiple Instance Regression (MIR) based deep blind IQM is proposed in [9] with the assumption that each instance (patch) has a probability to be the prime instance of the image, which is responsible for the image label. Then, the global quality score is computed by the weighted summation of the local quality scores of the instances, where the weights are the probabilities of the instances to be the prime one. DeepIQA [37] model is designed using unpreprocessed image patches as an input and estimates the quality without employing any domain knowledge, inspired by the data driven concept proposed in [38]. SESANIA [39] is a NRIQA model based on deep learning which uses sum of subband coefficient amplitudes (SSCA) as primary features to differentiate natural and distorted image. The statistical property of most natural image in Shearlet domain is relatively constant. However in distorted image, it becomes discontinuous.

Neural networks designed with convolutional and Fully Connected Neural Network (FCNN) layers are incorporated in the Deep Image Quality Assessor (DIQA) [6] for learning objective error map and the corresponding prediction of the subjective scores. To further enhance the accuracy of objective error map, the reliability map of distorted image, generated by measuring the textural strength, is considered. Similarly, handcrafted features were employed to define the loss function required during the training of the module designed using fully connected network layers. This process helps in enhancing the accuracy of the subjective score. Extraction of human perceptual error map based on the visual sensitivity map [4] is also incorporated to evaluate the performance of the DIQA model.

The statistical evaluation of an IQA model is performed in [40] to verify the added value of computational saliency in objective IQM. For this purpose, different state-of-the-art saliency models and the best-known IQMs are considered. From the quantitative results, it is concluded that the difference in predicting human fixations between saliency models is sufficient to yield a noticeable difference in performance gain when integrating these saliency models to IQMs. From the performed analysis it can be noticed that in recent years the number of NR approaches has increased compared to FR and RR. This is explained with the advancements in ML and DNN techniques and with the fact that NR metrics do not require reference information at the receiver side thus resulting more suitable in real applications. However, the performance of most of the proposed models has been analyzed only on small datasets, and cross analysis is performed on a limited scale. In this context, a new DNN-based NR approach, DIQASN, is proposed. The model performance is evaluated with both small and large-size datasets and it is compared with state-of-the-art models. In addition, the performance of the model on cross-dataset tests are reported.

2.4. CNN and its limitations

CNN is a class of deep neural networks, used in the analysis of visual imagery [41,42] (e.g., image classification, medical image analysis [43], image and video quality assessment).

The performance of a neural network system depends on the task to be performed, learning complexity [44], and the size of the dataset used for training the model. A commonly used dataset in computer vision based application is ImageNet [45]. It consists of more than 1.2 million labeled data.

The common dataset used in IQA is LIVE_IQA [17] dataset, which consists of 982 distorted images and 29 reference images. The main drawback of the small size training dataset is that it may cause overfitting problem. To overcome this issue, in [46] the authors suggest to perform early stopping, network reduction or pruning, expansion of training data or data augmentation, and regularization. A different approach is presented in [6] where a two-stage training is proposed.

3. Proposed model

The proposed DIQASN model is shown in Fig. 1. Inspired by [6,47], it is composed by three main blocks. First, a pre-processing step is applied to the input image. Secondly, a CNN is employed for extracting spatial features. Finally, a regression block is used for predicting the quality scores of the image used under test. The following subsections illustrate each section of the DIQASN model in detail.

3.1. Pre-processing

In this stage, the luminance component of the input image is considered. A low pass filtered version is obtained through the three-stage function defined in [6]: gaussian filtering, sub-sampling of a factor 4, and interpolation to the original size. Let I_d be the input image, I_{dg} its luminance component, and I_{dl} the corresponding low pass filtered image. The pre-processed image I_{dp} [4] is obtained as:

$$I_{dp} = |I_{dg} - I_{dl}|. \tag{1}$$

This pre-processing is necessary for two reasons. First, the distortions in images usually only partially affect low frequencies. Hence, pre-processing helps removing unnecessary information that may hinder the CNN's learning process during the training phase. Likewise, it helps improving the model efficiency by reducing the overall computational cost and increasing the prediction accuracy. However, there is the drawback of losing information contained in original data. To cope with this issue, the saliency map and natural scene statistics are used, as detailed in the following.

3.2. Texture map

Textures are relevant elements in the exploration of a scene and they capture human attention. For this reason the presence of textures is considered. The texture map T is defined by:

$$T = \frac{R}{\frac{1}{HW}\sum_{i,j} R_{i,j}}, \tag{2}$$

where, H and W are height and width of I_{dp}, $R_{i,j}$ represents the value of R at location (i, j), and R is defined as [6]:

$$R = \frac{2}{1 + exp(-\beta(I_{dp}))} - 1, \tag{3}$$

where, β is a constant used to control the saturation of the texture map. A higher level of β correspond to larger homogeneous regions in R. The recommended value of β in [6] is between 0.6 and 1.4.

Zero values in the texture map T correspond to pixels in I_{dp} having no spatial information.

3.3. Saliency map

In order to improve the information conveyed by the texture map, we process the image to give more relevance to pixels belonging to edges in salient areas of the image. To estimate the image saliency, we rely on the fact that, for a human subject, objects in the foreground are generally more relevant than those in the background or than the background itself. Therefore, as a first step, we calculate a foreground background separation. First, the image is filtered by using multi-scale pyramidal mean shift filter [22] to generate a smooth image by grouping close pixels together. Then, we measure each pixel's similarity to the background through histogram back-projection as introduced in [21]. The back-projected image is smoothed by using the mean shift operation and its contrast is enhanced using histogram equalization. The negative of the resulting image is computed to obtain a coarse saliency map S_c. Then, bounding boxes are defined around the edges of S_c by exploiting the GrabCut algorithm in [23]. The obtained output is the saliency map S.

3.4. Convolutional layers

DIQASN is composed of 8 convolutional layers which are responsible for extracting spatial features from the filtered image. More specifically, given the pre-processed image I_{dp}, the convolutional layers return a feature map of size $\frac{H}{4} \times \frac{W}{4} \times 128$.

Then, the predicted map I_{pred} is evaluated by applying a 2D convolutional layer to the feature map. More precisely, a 1×1 convolutional layer is employed in order to convert the high dimensional feature map into a grayscale image of dimensions $\frac{H}{4} \times \frac{W}{4}$.

The first training stage of DIQASN focuses on the filters of the convolutional layers. We evaluate the ground truth as

$$I_{gth} = |I_{rp} - I_{dp}|^P, \tag{4}$$

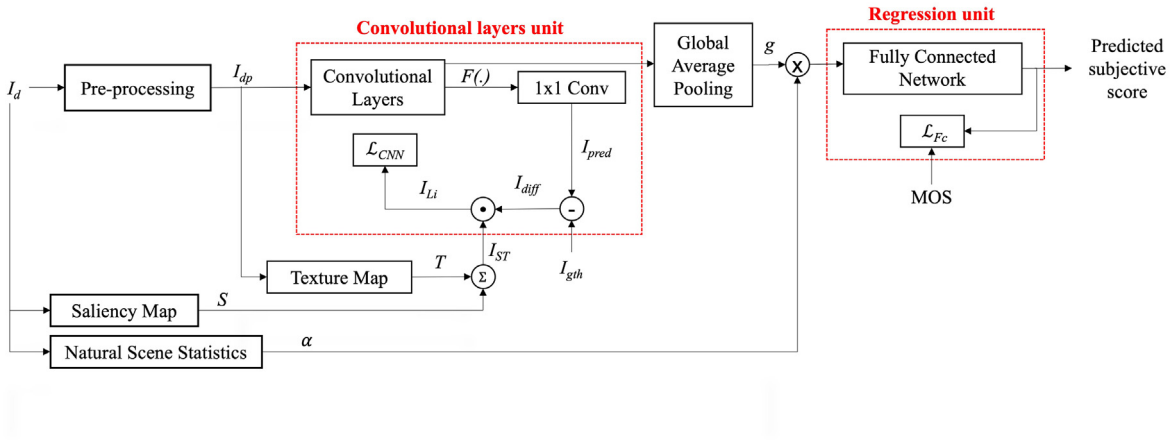


Fig. 1. Block diagram of proposed DIQASN model.

Table 1
Description of DIQASN's convolutional layers. The @ symbol stands for the number of the output channels. Last convolutional layer is used in order to obtain I_{pred}.

Conv2D 3×3 @ 48 channels
ReLU
Conv2D 3×3 @ 48 channels, stride $= (2, 2)$
ReLU
Conv2D 3×3 @ 64 channels
ReLU
Conv2D 3×3 @ 64 channels, stride $= (2, 2)$
ReLU
Conv2D 3×3 @ 64 channels
ReLU
Conv2D 3×3 @ 128 channels
ReLU
Conv2D 3×3 @ 128 channels
ReLU
Conv2D 3×3 @ 128 channels
ReLU
Conv2D 1×1 @ 1 channel

where I_{rp} is the pre-processed pristine image and I_{dp} is the pre-processed distorted image. It is important to underline that the reference image is only used to compute the ground truth. Then, the hyper parameter P, that is in the range $(0, 1)$, is introduced in order to generate a more strict I_{gth}. In this work, the adopted value of P is 0.2.

By means of the weighted texture map I_{ST}, we evaluate the difference between the ground truth and I_{pred}:

$$I_{Li} = (I_{pred} - I_{gth}) \bigodot I_{ST}, \tag{5}$$

where $\bigodot$ stands for the point-wise matrix multiplication. To compute the difference between the two images, the ground truth image I_{gth} is downsampled by a factor of 4. Finally, the weights of the convolutional stage are learnt by minimizing the L_2 norm of I_{Li}. More specifically, our loss function $\mathcal{L}_{CNN}$ is defined as:

$$\mathcal{L}_{CNN}(I_{Li}) = \|I_{Li}\|_2^2. \tag{6}$$

The minimization process is performed by means of the gradient descent algorithm, exploiting the derivative of the loss function with respect to the weights of the convolutional layers. Thus, the purpose of I_{diff} is to determine the closeness between I_{pred} and I_{gth} during the training process of the convolutional layer unit.

We have to remark that each non-strided convolutional layer of the model performs a zero-padding operation, preserving the size of the image. More details about the parameters of the CNN are available in Table 1. Moreover, all the images involved in the convolutional layers unit are depicted in Fig. 2.

3.5. Natural scene statistics

Based on [36], we exploit the evidence that there is a difference in the distribution of pixel intensities between pristine and distorted images. This difference is more evident when normalized pixel intensities are considered. In fact, the intensity distribution of the distorted normalized image significantly deviates from the Gaussian distribution that is typical of pristine images. For this reason, we compute the normalized image by using the Mean Subtracted Contrast Normalization (MSCN) approach:

$$\hat{I} = \frac{(I_d - \mu_l)}{(\sigma_l + 1)}, \tag{7}$$

where, μ_l and σ_l are the local mean and standard deviation that are computed as:

$$\mu_l(i, j) = \sum_{x=-X}^{X} \sum_{y=-Y}^{Y} w(x, y) I_d(i + x, j + y), \tag{8}$$

$$\sigma_l(i, j) = \sqrt{\sum_{x=-X}^{X} \sum_{y=-Y}^{Y} w(x, y).[I_d(i + x, j + y) - \mu(i, j)]^2}, \tag{9}$$

where, $w = \{w(x, y) \mid x = -X, \ldots, X, y = -Y, \ldots, Y\}$ indicates a $2D$ circularly symmetric Gaussian weighting function and (i, j) are pixel coordinates.

In our approach we used two NSS feature descriptors of $\hat{I}$: the mean, μ_{nss}, and the standard deviation, σ_{nss}. These feature descriptors are used in the proposed method to compute the visual sensitivity factor α[3], that is defined as the ratio $\frac{\sigma_{nss}}{\sigma_{nss}}$.

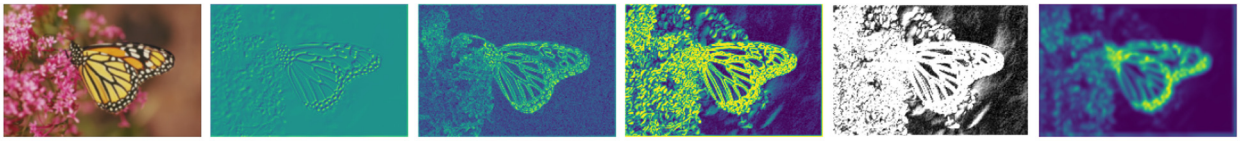
Let $GAP(.)$ be the Global Average Pooling operator which computes the average of each channel of the feature map. Given the multi-channel feature map of size $\frac{H}{4} \times \frac{W}{4} \times 128$, the resulting vector has 128 features. Then, a scale factor α is applied to each element of the feature vector. Let us denote with $F(.)$ the convolutional layers, the feature vector $\mathbf{g}$ is obtained as:

$$\mathbf{g} = \alpha GAP(F(I_{dp})). \tag{10}$$

The feature vector $\mathbf{g}$ is then fed to three fully connected layers (with 128, 64, and 1 neurons, respectively) in order to regress the MOS score.

3.6. Regression unit

The regression unit of the proposed model aims at predicting the perceived quality of the input image obtained as output of the convolutional layers. The second training stage of DIQASN focuses on learning the weights of the fully connected layers by computing the MSE between the predicted MOS and the ground truth. More precisely,



(a) (b) (c) (d) (e) (f)

Fig. 2. All the images involved in the convolutional layers unit: (a) a distorted image I_d; (b) the corresponding pre-processed image I_{dp}; (c) the ground truth image I_{gth}; (d) the texture map T; (e) the weighted texture map I_{ST}; (f) the predicted map I_{pred}.

Table 2
List of distortions present in LIVE_IQA, TID2013 and CSIQ datasets.

S. No.	Dataset	Distortions
1	LIVE_IQA	JPEG, JPEG2K, Gaussian blur, Additive white gaussian noise and Fast fading Rayleigh.
2	TID2013	Additive white gaussian noise, Additive noise in color component, Spatially correlated noise, Masked noise, High-frequency noise, Impulse Noise, Quantization noise, Gaussian blur, Image denoising, JPEG and JPEG2K & their transmission errors, Non-eccentricity pattern noise, Local block-wise distortions of different intensity, Mean shift, Contrast change, Change of color saturation, multiplicative Gaussian noise, Comfort noise, Lossy compression of noisy images, Image color quantization with dither, Chromatic aberration, Sparse sampling, and reconstruction.
3	CSIQ	JPEG and JPEG2K, Gaussian blur, Global contrast decrements, Additive white gaussian noise(AWGN) & Additive pink gaussian noise.

let $\mathcal{L}_{FC}$ be the loss function that trains the regression unit, $\hat{y}$ be the predicted MOS and y be the ground truth:

$$\mathcal{L}_{FC} = |\hat{y} - y|^2. \tag{11}$$

The weights are updated with the same algorithm used for training the convolutional layers.

4. Experiments and analysis

4.1. Datasets

To evaluate the performance of the proposed DIQASN model LIVE_IQA [17], TID2013 [18], and CSIQ [19] datasets are used. LIVE_IQA dataset is based on 29 high resolution RGB color images. 982 distorted images are obtained by applying five distortions with varying level of distortion. The subjective quality of the distorted images is measured in terms of the Differential Mean Opinion Score (DMOS). Lower values of the subjective score indicate higher perceptual quality, whereas higher values correspond to a lower visual quality. TID2013 dataset includes 3000 distorted images and 25 reference images. Distorted images are computed by applying 24 types of distortions and 5 levels of distortions on each reference image. MOS has been collected. CSIQ dataset is based on 30 original images, 6 types of distortions with 4 to 5 distortion levels thus resulting in 866 distorted images. The subjective ratings are given in terms of DMOS. The distortions present in the selected datasets are reported in Table 2.

4.2. Implementation details

In order to update the weights of each block, a NADAM [48] optimizer with learning rate $\lambda = 2e^{-4}$ is employed. The batch size is set empirically to 1. The implementation is based on Tensorflow-Keras with 4 NVIDIA GTX 2080Ti GPU. Additional parameters used during model implementation are reported in Table 4. For the training and testing of the proposed model, we randomly split the dataset in 80% images for the training and 20% for the test set. To avoid overfitting several counter measures have been adopted: (i) we used data augmentation in the convolutional layers unit based on the horizontal flipping of the pre-processed image, I_{dp}, which increases the number of training images by twofold, (ii) we implemented a small-size CNN to minimize the complexity of the convolutional layers unit and, consequently, require a smaller number of training data, (iii) we trained the model with pre-processed image in their original size to be able to exploit a larger number of features with respect to its subsampled version, and (iv) we exploited early stopping [49].

4.3. Performance evaluation

To evaluate the performance of the algorithm, SRCC, PLCC, Kendall Rank Order Correlation Coefficient (KRCC), and MSE are used. Let $\mathbf{s_g}$ and $\mathbf{s_p}$ be the vectors of ground truth quality scores and predicted subjective scores of the test set for n images, respectively. Let $\mu_{\mathbf{s_g}}$ and $\mu_{\mathbf{s_p}}$ be the mean of $\mathbf{s_g}$ and $\mathbf{s_p}$. Then, the general expression for SRCC, PLCC [50], and MSE are as follows:

$$SRCC(\mathbf{s_g}, \mathbf{s_p}) = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)}, \tag{12}$$

$$PLCC(\mathbf{s_g}, \mathbf{s_p}) = \frac{\sum_i (s_{p_i} - \mu_{\mathbf{s_p}})(s_{g_i} - \mu_{\mathbf{s_g}})}{\sqrt{\sum_i (s_{p_i} - \mu_{\mathbf{s_p}})^2}\sqrt{\sum_i (s_{g_i} - \mu_{\mathbf{s_g}})^2}}, \tag{13}$$

$$MSE(\mathbf{s_g}, \mathbf{s_p}) = \frac{1}{n}\sum_{i=1}^{n} (s_{g_i} - s_{p_i})^2, \tag{14}$$

where d_i is the ith element of the difference vector between the MOS and the predicted vector scores. In addition, the KRCC metric is computed as:

$$KRCC(\mathbf{s_g}, \mathbf{s_p}) = \frac{(n_c - n_d)}{\frac{n(n-1)}{2}}, \tag{15}$$

where n_c and n_d represent the number of concordant and discordant pairs between $\mathbf{s_g}$ and $\mathbf{s_p}$.

The scatter plots of the predicted subjective scores versus MOS of the test images for the three datasets are reported in Fig. 3. It is possible to appreciate the good ability of DIQASN to predict the quality scores on considered datasets. In fact the values are all close to the diagonal (representing the perfect match between MOS and predicted values) and are all close to each other.

We evaluated the impact of adopting the information extracted from the saliency map and the NSS features through the computation of SRCC and PLCC between the ground truth and the predicted MOSs for the three datasets, LIVE_IQA, TID2013, and CSIQ. Table 3 reports the obtained results. It can be noticed that, for all considered datasets, the use of saliency and NSS features slightly improves the SRCC and PLCC scores with respect to the case in which no features are exploited.

An additional test was performed to analyze the variability of the correlation coefficients, SRCC and PLCC, with respect to the images used as training and testing. As already mentioned, we randomly split the dataset in 80% images for the training and 20% for the test set. We repeated this operation for 10 runs and we evaluated the SRCC and PLCC with their confidence intervals. The box plot in Fig. 4 shows that the confidence intervals of the SRCC and PLCC coefficients are very small, indicating that DIQASN has a strong stability.

To evaluate the proposed model performances with respect to the state-of-the-art, we compared them with:

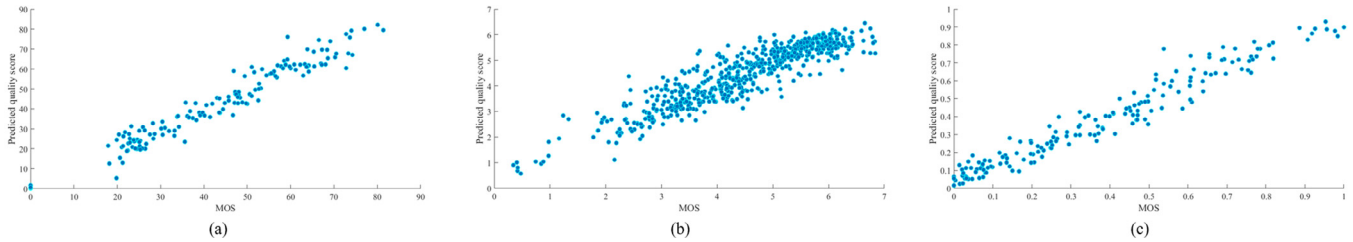


Fig. 3. Scatter plot between predicted quality scores, and ground truth quality scores of test images (20% of total images) for different datasets. (a) LIVE_IQA dataset, (b) TID2013 dataset, and (c) CSIQ dataset.

Table 3
SRCC and PLCC between the ground truth and predicted MOSs when different features are used.

Features	LIVE_IQA		TID2013		CSIQ	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
Saliency and NSS	**0.979**	**0.990**	**0.912**	**0.920**	**0.973**	**0.975**
Saliency only	0.979	0.989	0.906	0.918	0.973	0.974
NSS only	0.979	0.990	0.911	0.913	0.969	0.973
No Saliency and no NSS	0.978	0.987	0.904	0.910	0.968	0.970

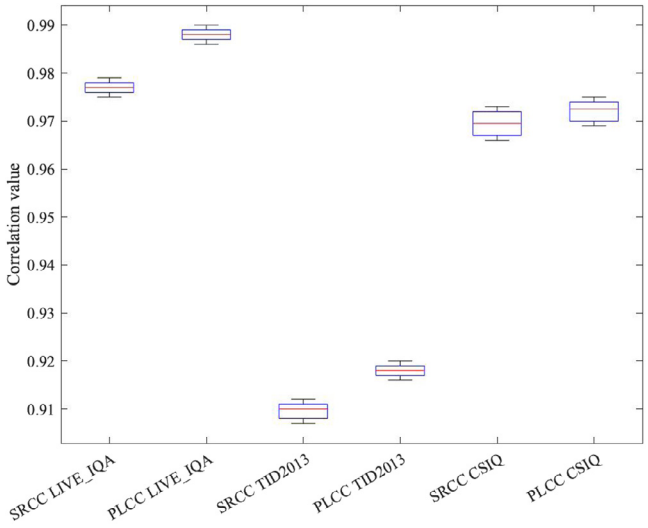


Fig. 4. Variability analysis of the SRCC and PLCC scores.

Table 4
Model parameters.

S. No.	Module	Parameters/variables	Value
1	Image normalization	Gaussian kernel σ	16×16 7/6
2	Convolutional layers unit	Convolution kernel Input image size Batch size Max pooling stride Trainable parameters	3×3 Original size 1 2×2 381,972
3	Texture map	P β	0.2 0.6
4	Saliency map	Pixel weights for segmentation	[200 255]
5	Regression unit	Trainable parameters	24,061

- 8 FR IQA metrics: PSNR [51], SSIM [7], GAFFE-SSIM [25], PieAPP [29], FR-DCNN [30], DeepQA-s, DeepQA [5], an AlexNet based FR IQM [24];
- 4 RR IQA metrics: FSI [31], RRED [32], SDM [34], RBMSim [35];
- 9 NR IQA metrics: BRISQUE [36], BIECON [52], CNN [38], DMIR [9], DeepIQA [37], DIQA [6], DIQA_BASE [6] MGDNN [53], SESANIA [39].

To perform this evaluation, we trained and tested the DIQASN model by randomly dividing the three datasets into two subsets with 80% for training and 20% for testing without overlapping.

The performance comparison is shown in Table 5. From the collected results, it is possible to notice that DIQASN outperforms the considered NR and RR metrics, while it shows comparable SRCC and PLCC values with two FR metrics. As previously mentioned, the DIQASN model is an extended version of DIQA leading to improved performance. One of the main differences between DIQA and DIQASN is in the training step. DIQASN presents a simplified approach with respect to DIQA thus leading to a decrease in the computational complexity. The first training stage of DIQA is performed by using image patches extracted from the normalized form of the original input image. DIQASN instead uses the entire input image and the weighted texture map as inputs to the learning model. Furthermore, while in DIQA the regression unit is designed using two fully connected layers having feature channels 130:128, and 128:1, in DIQASN the regression unit is designed using three fully connected layers having feature channels: 128:128, 128:64 and 64:1. Furthermore, we use a different approach for coping with the information loss caused by the normalization process. In DIQA, the mean and the standard deviation extracted from the texture map and the low-pass frequency component of the input image are used as additional feature sources. Those are concatenated with the output of the global average pooling layer before moving to the second training stage. In DIQASN we simplify this procedure by adopting the visual sensitivity factor of the input image to scale the output of the global average pooling layer before forwarding it to the second training stage.

We also investigated the performances of the model with respect to different distortions. The results are reported in Table 6 in terms of the best scores of SRCC, PLCC, and KRCC obtained by measuring the similarity between the predicted and ground truth quality scores for the different classes of distortions available in the adopted datasets. The distortions for which the metric provides the best and worst results are reported in bold.

In DIQASN, a gaussian filtering is adopted for pre-processing the input image and to obtain I_{dp}, which is then used to train the model. In this way, DIQASN is trained to recognize the presence of relevant edges and textures. This has a positive impact on three types of distortions: Gaussian Blur (GB), Additive White Gaussian Noise (AWGN) and High Frequency Noise (HFN). For this reason, the model is able to accurately estimate their presence since that largely affect the edges and textures in the images.

The Quantization Noise (QN) is usually caused by the operations of image registration and γ-correction and it affects the local contrast and the spatial frequency [18]. The Mean Shift (MS) distortion produces an intensity shift and it is common in the image acquisition phase [54]. Both QN and MS affect the luminance component of the image, thus leading to a degradation in the system performances. In fact, all the steps in the DIQASN model rely on the features extracted from the luminance component.

Table 5
Performance comparison of DIQASN in terms of SRCC and PLCC with state-of-the-art quality metrics.

Model	Quality metric	Dataset type					
		LIVE_IQA		TID2013		CSIQ	
		SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
FR	PSNR	0.904	0.868	0.672	0.496	0.719	0.670
	SSIM	0.938	0.855	0.650	0.680	0.821	0.747
	GAFFE-SSIM	0.933	–	–	–	–	–
	PieAPP	–	–	**0.945**	**0.946**	**0.973**	**0.975**
	FR-DCNN	0.975	0.977	–	–	–	–
	AlexNet based FR IQM	–	0.91	–	0.84	–	0.92
	DeepQA	**0.981**	**0.982**	**0.939**	**0.947**	**0.961**	**0.965**
	DeepQA-s	0.977	0.975	0.766	0.818	0.957	0.956
RR	FSI	0.883	0.882	0.58	0.611	0.918	0.927
	RRED	0.943	0.939	–	–	–	–
	SDM	0.936	0.933	–	–	–	–
	RBMSIM	0.95	0.97	–	–	0.95	0.95
NR	BRISQUE	0.690	0.648	0.260	0.233	0.438	0.507
	BIECON	0.958	0.962	0.721	0.765	0.825	0.838
	CNN	0.956	0.963	–	–	–	–
	DMIR	0.967	0.971	0.796	0.821	0.823	0.881
	DeepIQA	0.960	0.972	–	–	–	–
	DIQA_BASE	0.963	0.964	0.800	0.803	0.812	0.791
	DIQA	0.975	0.977	0.825	0.850	0.884	0.915
	MGDNN	0.951	0.949	–	–	–	–
	SESANIA	0.934	0.948	–	–	–	–
	DIQASN	**0.977**	**0.988**	**0.910**	**0.918**	**0.969**	**0.972**

Table 6
Analysis of the DIQASN performance with respect to the specific distortion.

Dataset	Distortion types	SRCC	PLCC	KRCC
LIVE IQA	**Additive White Gaussian noise (AWGN)**	**0.990**	**0.992**	**0.949**
	Fast fading Rayleigh (FFR)	0.985	0.990	0.923
	Gaussian blur (GB)	0.989	0.991	0.940
	JPEG compression (JPEG)	**0.944**	**0.959**	**0.875**
	JPEG 2000 compression (JPEG2K)	0.979	0.987	0.901
CSIQ	Additive white gaussian noise (AWGN)	0.988	0.991	0.935
	Gaussian Blur (GB)	**0.993**	**0.994**	**0.950**
	Global contrast decrements (GCD)	0.986	0.989	0.925
	Additive pink gaussian noise (FNOISE)	0.977	0.980	0.897
	JPEG compression (JPEG)	0.925	0.928	0.831
	JPEG 2000 compression (JPEG2K)	**0.919**	**0.923**	**0.819**
TID2013	Additive White Gaussian noise (AWGN)	0.942	0.944	0.866
	Additive noise in color components (ANC)	0.940	0.942	0.835
	Chromatic aberrations (CA)	0.914	0.916	0.817
	Contrast change (CC)	0.841	0.842	0.705
	Comfort noise (CN)	0.907	0.910	0.790
	Change of color saturation (CCS)	0.832	0.844	0.692
	Gaussian Blur (GB)	0.946	0.949	0.878
	High frequency noise (HFN)	**0.948**	**0.951**	**0.903**
	Image color quantization with dither (ICQD)	0.933	0.935	0.848
	Image denoising (IDN)	0.937	0.939	0.850
	Impulse noise (IN)	0.900	0.932	0.827
	JPEG compression (JPEG)	0.940	0.943	0.856
	JPEG 2000 compression (JPEG2K)	0.942	0.944	0.870
	JPEG transmission errors (JPEG-TE)	0.912	0.915	0.828
	JPEG2K transmission errors (JPEG2K-TE)	0.864	0.866	0.740
	Local block-wise distortions (LBW)	0.885	0.887	0.752
	Lossy compression of noisy images (LCN)	0.843	0.845	0.869
	Multiplicative Gaussian noise (MGN)	0.940	0.942	0.853
	Masked noise (MN)	0.927	0.929	0.813
	Mean Shift (MS)	0.790	0.794	0.630
	Non eccentricity pattern noise (NEPN)	0.902	0.910	0.778
	Quantization noise (QN)	**0.741**	**0.744**	**0.563**
	Spatially correlated noise (SCN)	0.946	0.949	0.882
	Sparse sampling and reconstruction (SSR)	0.945	0.947	0.872

To further evaluate the performance of the proposed DIQASN model, it has been trained using one dataset and tested on the other two.

Table 7
Performance of the DIQASN model in cross-dataset tests.

Train	Test	SRCC	PLCC
LIVE_IQA	TID2013	0.50	0.53
	CSIQ	0.70	0.71
TID2013	LIVE_IQA	0.85	0.82
	CSIQ	0.68	0.74
CSIQ	LIVE_IQA	0.65	0.83
	TID2013	0.45	0.53

Table 8
Performance of BRISQUE, BIECON, and DIQASN models in terms of PLCC in cross-dataset tests.

Train	Test	BRISQUE	BIECON	DIQA	DIQASN
LIVE_IQA	TID2013	0.494	0.506	**0.524**	**0.529**
	CSIQ	0.689	**0.744**	0.706	**0.711**
TID2013	LIVE_IQA	0.798	**0.859**	0.817	**0.823**
	CSIQ	0.692	0.660	**0.737**	**0.744**
CSIQ	LIVE_IQA	**0.889**	0.761	0.825	**0.833**
	TID2013	0.528	0.482	**0.529**	**0.531**

The result of the cross-datasets evaluated by measuring the similarities between the predicted and ground truth quality scores in terms of SRCC and PLCC is reported in Table 7.

In [55] the cross-dataset performance of two NR IQA methods, BRISQUE and BIECON, is presented in terms of PLCC. The DIQA model performance is also evaluated based on [56]. A comparative study including the DIQASN model is shown in Table 8. It can be noticed that also in this case, the DIQASN model is one of the best performing models. Though, the aforementioned datasets have different content, i.e. other reference image types, adopted distortions, and number of distorted images. Hence, it results in a relatively lower value of correlation coefficients during the cross dataset test of the proposed model.

Finally, we evaluated the DIQASN performances on a real distortion database, KADID10K [20]. It is composed by 81 pristine images, each degraded by 25 distortions with 5 severity levels, all provided with the DMOS. The test has been performed by randomly splitting the images in 60% for training, 20% for validation, and 20% for test. The considered FR IQA metrics for the comparative study are SSIM, FSIM, and VSI [57].

Table 9
DIQASN performances on the KADID10K dataset.

Model	Quality metric	SRCC	PLCC	KRCC
FR IQA	SSIM	0.723	0.724	0.527
	FSIM	0.851	0.854	0.865
	VSI	**0.878**	**0.879**	**0.691**
NR IQA	BRISQUE	0.554	0.519	0.368
	CORNIA	0.580	0.541	0.384
	InceptionResNetV2	0.734	0.731	0.546
	DIQASN	**0.881**	**0.88**	**0.699**

Similarly, BRISQUE, CORNIA [58], and InceptionResNetV2 [59] are the NR IQA metrics considered during comparative study with our proposed model's performance. The obtained results in terms of SRCC, PLCC, and KRCC are reported in Table 9. From the obtained results, we conclude that, also for the real distortion database, the proposed model outperforms the FR IQA and NR IQA state-of-the-art metrics.

5. Conclusions

In this contribution a new quality metric for 2D images is presented, DIQASN. It exploits the characteristics of the human visual system, through saliency information and NSS features. The use of saliency information accounts for perceptually relevant areas in the image whereas, the NSS features allow to measure the changes in naturalness of the image due to the presence of artifacts. DIQASN shows competitive results with respect to FR, NR, and RR IQA metrics on four datasets. In particular, DIQASN outperforms NR, and RR IQA approaches for which no or limited information about the original content is available. This is an encouraging result since our NR approach can reach the same result as FR ones.

CRediT authorship contribution statement

Kamal Lamichhane: Conception and design of study, Writing – original draft. **Marco Carli:** Analysis and/or interpretation of data, Writing – review & editing. **Federica Battisti:** Analysis and/or interpretation of data, Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgment

All authors approved version of the manuscript to be published.

References

[1] A. Punchihewa, D.G. Bailey, Artefacts in image and video systems; classification and mitigation, in: Proceedings of Image and Vision Computing New Zealand, 2002.
[2] P.R. Rajarapollu, V.R. Mankar, Bicubic interpolation algorithm implementation for image appearance enhancement, Int. J. Comput. Sci. Technol. 8 (2) (2017) 23–26.
[3] M.A. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: A natural scene statistics approach in the DCT domain, IEEE Trans. Image Process. 21 (8) (2012) 3339–3352.
[4] J. Kim, S. Lee, Fully deep blind image quality predictor, IEEE J. Sel. Top. Sign. Proces. 11 (1) (2017) 206–220.
[5] J. Kim, S. Lee, Deep learning of human visual sensitivity in image quality assessment framework, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1676–1684.
[6] J. Kim, A. Nguyen, S. Lee, Deep cnn-based blind image quality predictor, IEEE Trans. Neural Netw. Learn. Syst. 30 (1) (2019) 11–24.
[7] W. Zhou, A.C. Bovik, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[8] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: A feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.
[9] D. Liang, X. Gao, W. Lu, J. Li, Deep blind image quality assessment based on multiple instance regression, Neurocomputing 431 (2021) 78–89.
[10] Z. Wang, Q. Li, Information content weighting for perceptual image quality assessment, IEEE Trans. Image Process. 20 (5) (2010) 1185–1198.
[11] W. Zhu, Y. Ma, Y. Zhou, M. Benton, J. Romagnoli, Deep learning based soft sensor and its application on a pyrolysis reactor for compositions predictions of gas phase components, in: Computer Aided Chemical Engineering, Vol. 44, Elsevier, 2018, pp. 2245–2250.
[12] T.O. Ayodele, Types of machine learning algorithms, New Adv. Mach. Learn. 3 (2010) 19–48.
[13] A. Canziani, A. Paszke, E. Culurciello, An analysis of deep neural network models for practical applications, 2016, arXiv Preprint arXiv:1605.07678. arXiv–1605.
[14] L. Nanni, S. Ghidoni, S. Brahnam, Handcrafted, vs. non-handcrafted features for computer vision classification, Pattern Recognit. 71 (2017) 158–172.
[15] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: From natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20 (12) (2011) 3350–3364.
[16] W. Xue, X. Mou, L. Zhang, A.C. Bovik, X. Feng, Blind image quality assessment using joint statistics of gradient magnitude and laplacian features, IEEE Trans. Image Process. 23 (11) (2014) 4850–4862.
[17] H.R. Sheikh, Live image quality assessment database release 2, 2005, http://live.ece.utexas.edu/research/quality.
[18] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, et al., Image database TID2013: Peculiarities, results and perspectives, Signal Process., Image Commun. 30 (2015) 57–77.
[19] E.C. Larson, D.M. Chandler, Most apparent distortion: Full-reference image quality assessment and the role of strategy, J. Electron. Imaging 19 (1) (2010) 011006.
[20] H. Lin, V. Hosu, D. Saupe, Kadid-10k: A large-scale artificially distorted iqa database, in: 2019 Eleventh International Conference on Quality of Multimedia Experience, QoMEX, 2019, pp. 1–3.
[21] M.J. Swain, D.H. Ballard, Color indexing, Int. J. Comput. Vis. 7 (1) (1991) 11–32.
[22] P.J. Burt, Fast filter transforms for image processing, Comput. Graph. Image Process. 16 (1) (1981) 20–51.
[23] C. Rother, V. Kolmogorov, A. Blake, GrabCut interactive foreground extraction using iterated graph cuts, ACM Trans. Graph. 23 (3) (2004) 309–314.
[24] S.A. Amirshahi, M. Pedersen, S.X. Yu, Image quality assessment by comparing cnn features between images, J. Imag. Sci. Technol. 60 (6) (2016) 60410–1 – 60410–10.
[25] M.C.Q. Farias, W.Y.L. Akamine, On performance of image quality metrics enhanced with visual attention computational models, Electron. Lett. 48 (11) (2012) 631–633.
[26] L. Itti, C. Koch, Computational modelling of visual attention, Nat. Rev. Neurosci. 2 (3) (2001) 194–203.
[27] R. Achanta, F. Estrada, P. Wils, S. Süsstrunk, Salient region detection and segmentation, in: International Conference on Computer Vision Systems, Springer, 2008, pp. 66–75.
[28] U. Rajashekar, I. van der Linde, A.C. Bovik, L.K. Cormack, GAFFE: A gaze-attentive fixation finding engine, IEEE Trans. Image Process. 17 (4) (2008) 564–573.
[29] E. Prashnani, H. Cai, Y. Mostofi, P. Sen, Pieapp: Perceptual image-error assessment through pairwise preference, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1808–1817.
[30] Y. Liang, J. Wang, X. Wan, Y. Gong, N. Zheng, Image quality assessment using similar scene as reference, in: European Conference on Computer Vision, Springer, 2016, pp. 3–18.
[31] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, W. Gao, Reduced-reference image quality assessment in free-energy principle and sparse representation, IEEE Trans. Multimed. 20 (2) (2018) 379–391.
[32] R. Soundararajan, A.C. Bovik, RRED indices: Reduced reference entropic differencing for image quality assessment, IEEE Trans. Image Process. 21 (2) (2011) 517–526.
[33] X. Wang, G. Jiang, M. Yu, Reduced reference image quality assessment based on contourlet domain and natural image statistics, in: 2009 Fifth International Conference on Image and Graphics, 2009, pp. 45–50.
[34] K. Gu, G. Zhai, X. Yang, W. Zhang, A new reduced-reference image quality assessment using structural degradation model, in: 2013 IEEE International Symposium on Circuits and Systems, ISCAS, 2013, pp. 1095–1098.
[35] D.C. Mocanu, G. Exarchakos, H.B. Ammar, A. Liotta, Reduced reference image quality assessment via boltzmann machines, in: 2015 IFIP/IEEE International Symposium on Integrated Network Management, IM, 2015, pp. 1278–1281.

[36] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[37] S. Bosse, D. Maniry, T. Wiegand, W. Samek, A deep neural network for image quality assessment, in: IEEE International Conference on Image Processing, ICIP, 2016, pp. 3773–3777.
[38] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1733–1740.
[39] Y. Li, L. Po, X. Xu, L. Feng, F. Yuan, C. Cheung, K. Cheung, No-reference image quality assessment with shearlet transform and deep neural networks, Neurocomputing 154 (2015) 94–109.
[40] W. Zhang, A. Borji, Z. Wang, P. Le Callet, H. Liu, The application of visual saliency models in objective image quality assessment: A statistical evaluation, IEEE Trans. Neural Netw. Learn. Syst. 27 (6) (2016) 1266–1278.
[41] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[42] M.V. Valueva, N.N. Nagornov, P.A. Lyakhov, G.V. Valuev, N.I. Chervyakov, Application of the residue number system to reduce hardware costs of the convolutional neural network implementation, Math. Comput. Simulation 177 (2020) 232–243.
[43] A. Qayyum, S.M. Anwar, M. Awais, M. Majid, Medical image retrieval using deep convolutional neural network, Neurocomputing 266 (2017) 8–20.
[44] D.K. Chaturvedi, Factors affecting the performance of artificial neural network models, in: Soft Computing: Techniques and its Applications in Electrical Engineering, 2008, pp. 51–85.
[45] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, Li Fei-Fei, ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.
[46] X. Ying, An overview of overfitting and its solutions, J. Phys. Conf. Ser. (2019) 1168-2.
[47] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
[48] T. Dozat, Incorporating nesterov momentum into adam, in: Proceedings Workshop Track, ICLR, 2016, pp. 1–4.
[49] J. Brownlee, Better deep learning: train faster, reduce overfitting, and make better predictions, in: Machine Learning Mastery, 2018.
[50] ITU, Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II (fr-tv2), 2020, https://www.itu.int/md/T01-SG09-C-0060. (Accessed 29 May 2020).
[51] J. Korhonen, J. You, Peak signal-to-noise ratio revisited: Is simple beautiful? in: 2012 Fourth International Workshop on Quality of Multimedia Experience, 2012, pp. 37–38.
[52] H. Oh, S. Ahn, J. Kim, S. Lee, Blind deep S3D image quality evaluation via local to global feature aggregation, IEEE Trans. Image Process. 26 (10) (2017) 4923–4936.
[53] Y. Lv, G. Jiang, M. Yu, H. Xu, F. Shao, S. Liu, Difference of gaussian statistical features based blind image quality assessment: A deep learning approach, in: IEEE International Conference on Image Processing, ICIP, 2015, pp. 2344–2348.
[54] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, C.-C.J. Kuo, Color image database TID2013: Peculiarities and preliminary results, in: European Workshop on Visual Information Processing, EUVIP, 2013, pp. 106–111.
[55] D. Yang, V. Peltoketo, J. Kamarainen, CNN-based cross-dataset no-reference image quality assessment, in: 2019 IEEE/CVF International Conference on Computer Vision Workshop, ICCVW, 2019, pp. 3913–3921.
[56] O. Ricardo, Deep cnn-based blind image quality predictor in python, 2020, https://towardsdatascience.com/deep-image-quality-assessment-with-tensorflow-2-0-69ed8c32f195. (Accessed 29 May 2020).
[57] L. Zhang, Y. Shen, H. Li, Vsi: A visual saliency-induced index for perceptual image quality assessment, IEEE Trans. Image Process. 23 (10) (2014) 4270–4281.
[58] P. Ye, J. Kumar, L. Kang, D. Doermann, Unsupervised feature learning framework for no-reference image quality assessment, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012.
[59] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-first AAAI Conference on Artificial Intelligence, 2017.