Article

Full-Reference Image Quality Assessment Based on an Optimal Linear Combination of Quality Measures Selected by Simulated Annealing

Domonkos Varga

Ronin Institute, Montclair, NJ 07043, USA; domonkos.varga@roninstitute.org

Abstract: Digital images can be distorted or contaminated by noise in various steps of image acquisition, transmission, and storage. Thus, the research of such algorithms, which can evaluate the perceptual quality of digital images consistent with human quality judgement, is a hot topic in the literature. In this study, an image quality assessment (IQA) method is introduced that predicts the perceptual quality of a digital image by optimally combining several IQA metrics. To be more specific, an optimization problem is defined first using the weighted sum of a few IQA metrics. Subsequently, the optimal values of the weights are determined by minimizing the root mean square error between the predicted and ground-truth scores using the simulated annealing algorithm. The resulted optimization-based IQA metrics were assessed and compared to other state-of-the-art methods on four large, widely applied benchmark IQA databases. The numerical results empirically corroborate that the proposed approach is able to surpass other competing IQA methods.

1. Introduction

Nowadays, people increasingly communicate through media in form of audio, video, and digital images. Therefore, image quality assessment (IQA) has found many applications and become a hot research topic in the research community [1]. Namely, IQA methods evaluate the perceptual quality of digital images and support, among others, image enhancement [2], restoration [3], steganography [4], or denoising algorithms [5]. Further, IQA is also necessary in the benchmarking of many image processing or computer-vision algorithms [6–8]. In the literature, IQA is classified into two groups, i.e., subjective and objective IQA. Specifically, subjective IQA deals with the collection of users' quality ratings for a set of digital images either in a laboratory [1] or in an online crowd-sourcing experiment [9]. Moreover, images' perceptual quality is expressed as a mean opinion score (MOS), which is the arithmetic mean of individual quality scores. As a result, subjective IQA provides quality labelled images with objective IQA as training or test data [10]. Namely, objective IQA deals with algorithms and mathematical models that are able to predict the quality of a given image. Conventionally, objective IQA is divided into three classes [11]—full-reference (FR) [12], reduced-reference (RR) [13], and no-reference (NR) [14]—with respect to the availability of the reference (distortion-free) images. As the names indicate, FR-IQA methods have full access to the reference images. In contrast, NR-IQA algorithms evaluate image quality without any information about the reference images [15], and RR-IQA algorithms have partial information about them.

1.1. Contribution

The development of objective FR-IQA algorithms can also involve fusion-based strategies that already take existing FR-IQA metrics and try to create a "super evaluator". Recently, many complex fusion-based approaches have been published in the literature [16–19]. The main contribution to this paper is also a fusion-based approach. Namely,

we demonstrate a solution based on a linear combination of several already existing FR-IQA metrics optimized with a simulated annealing (SA) algorithm using a root mean square error (RMSE) objective, which is able to produce well-performing fusion-based FR-IQA metrics. To be more specific, a linear combination of 16 FR-IQA metrics is used in an optimization problem to select FR-IQA metrics and find their weights via an SA algorithm that minimizes the RMSE of the prediction. Unlike the approach of Oszust [20], we apply simulated annealing instead of a genetic algorithm for performing the fusion of FR-IQA metrics. Namely, simulated annealing usually achieves better results in the case of continuous function approximation than basic genetic algorithms because they choose one or two genes at a given location [21]. The proposed fusion-based metrics was evaluated on large, popular, and widely accepted IQA benchmark databases, such as LIVE [22], TID2013 [23], TID2008 [24], and CSIQ [25].

1.2. Organization

The rest of this paper is organized as follows. In Section 2, an overview about the current state of FR-IQA is given. Next, the proposed fusion-based metric is introduced in Section 3. Our experimental results, together with the description of the applied benchmark IQA databases, evaluation environment, and performance indices, are given in Section 4. Finally, a conclusion is drawn in Section 5.

2. Literature Review

In this paper, we follow the classification of FR-IQA algorithms presented in [26]. To be specific, Ding et al. [26] categorized existing FR-IQA algorithms into five distinct classes, i.e., error visibility, structural similarity, information theoretic, learning-based, and fusion-based methods.

Error visibility methods measure a distance between the pixels of the distorted and the reference images to quantify perceptual quality degradation. The representative method of this class of FR-IQA is the mean squared error (MSE) method, which measures the average of the squares of the errors. In other words, it is the average squared difference between the reference and the distorted images in the context of FR-IQA [27]. Another well-known example is the peak signal-to-noise ratio (PSNR), which is commonly applied to assess the quality of the reconstruction of lossy compression codecs [28]. Although both MSE and PSNR have low computational costs and their physical meaning is clear and well understood, they often mismatch with subjective perceptions of visual quality.

Structural similarity methods measure the similarity between the corresponding regions of the distorted and reference images using sliding-windows in the images and correlation measures. The representative and first published method of this class is the structural similarity index (SSIM) [29], which has become extremely popular in the field with many extensions and applications [30]. The theorem of SSIM has become extremely popular in the research community and inspired many variants. For example, the wavelet domain structural similarity [31] carries out SSIM in the wavelet domain to quantify perceptual quality. This work was extended by Sampat et al. [32] into the complex wavelet domain. In [33], information content was utilized as weights in the pooling process of local image quality scores. In contrast, Wang et al. [34] extended SSIM to multi-scale processing to improve perceptual quality estimation. Li and Bovik [35] elaborated an FR-IQA metric by taking the average of SSIMs computed over three different regions of an image, such as edges, textures, and smooth regions. Kolaman and Yadid-Pecht [36] found an extension of SSIM to color images by representing red, green, and blue color channels with quaternions. Later, SSIM was also extended to hyperspectral images [37].

Information theoretic methods approach the FR-IQA task from the point of view of information communication. For example, Sheikh et al. [38,39] compared the information content of the reference and distorted images. Namely, perceptual quality was quantified by how much information is similar between the reference and distorted images. In contrast, Larson and Chandler [25] classified image distortions as near-threshold and

supra-threshold. The authors elaborated two quality indexes for both distortion types. Finally, the overall perceptual quality was determined based on the quality scores of near-threshold and supra-threshold distortions.

As the terminology suggests, learning-based methods rely on a specific machine learning algorithm to create a quality model from training images. Next, the obtained quality model is tested on previously unseen images. For instance, Liang et al. [40] implemented a special convolutional neural network containing two paths, one for the reference image and the other for the distorted image. Further, this network was trained on 224×224-sized image patches sampled simultaneously from the reference and distorted images. As a consequence, the perceptual quality of a distorted image was estimated by the average score of the considered patches. Kim and Lee [41] devised a similar network, but it predicts a visual sensitivity map that is multiplied by an error map calculated directly from the reference and the distorted images to estimate perceptual image quality. Ahn et al. [42] further improved the idea of Kim and Lee [41] by implementing an end-to-end trained convolutional neural network with three inputs, i.e., reference image, distorted image, and spatial error map. Similar to [41], a distortion-sensitivity map was predicted from the inputs and was later multiplied by the spatial error map to give an estimation for the perceptual image quality. In contrast to the previously mentioned methods, Ding et al. [43] extracted a set of feature maps from the reference and the distorted images using the Sobel operator, log Gabor filter, and local pattern analysis. Subsequently, the extracted feature maps were compared, and from the resulting similarity scores a feature vector was compiled that was mapped onto perceptual quality scores with a trained support vector regressor. Tang et al. [44] took a similar approach, but the authors employed a different set of features (phase congruency maps [45], gradient magnitude maps, and log Gabor maps). Further, the similarity scores of the feature maps were mapped onto perceptual quality with a trained random forest regressor.

Fusion-based FR-IQA methods utilize existing FR-IQA metrics to create a new FR-IQA algorithm. First, Okarma [46] suggested the idea of combined methods. Namely, the author proposed a combined metric using the product and power of MS-SSIM [34], VIF [38], and R-SVD [47]. This approach was developed further in [19], where the optimal exponents in the product were determined by using MATLAB's fminsearch command. In [48], Oszust took a similar approach, but the author applied the scores of traditional FR-IQA metrics as predictor variables in a lasso regression. Instead of lasso regression, Yuan et al. [49] used kernel ridge regression in a similar layout. The work of Lukin et al. [50] exhibits the properties of both learning-based and fusion-based methods. Specifically, the authors created a training and a test set from the images of an IQA benchmark database. Next, the scores of several traditional FR-IQA metrics were used as image features, and a neural network was trained to estimate perceptual image quality. Amirshahi et al. [51] elaborated a special fusion-based FR-IQA metric relying on a pretrained convolutional neural network. Namely, the authors ran a reference-distorted image pair through an AlexNet [52] network and compared the activation maps with the help of a traditional FR-IQA metric. Next, the resulted scores were aggregated to obtain a single score for the perceptual image quality. Bakurov et al. [53] revisited the classical SSIM [29] and MS-SSIM [34] metrics by applying evolutionary and swarm intelligence optimization methods to find optimal hyperparameters for SSIM and MS-SSIM instead of the original settings. Fusion-based metrics were also proposed for remote sensing images [54], stitched panoramic images [55], and 3D image quality assessment [18].

For more detailed studies about FR-IQA, we refer readers to the book of Xu et al.'s [56] and to the study of Pedersen and Hardeberg [57]. Further, Zhang et al. [58] provide an evaluation of several state-of-the-art FR-IQA algorithms on various IQA benchmark databases. Zhai and Min provided an comprehensive overview of classical algorithms in [59]. For the quality assessment of screen content images [60], Min et al. gave an overview in [61].
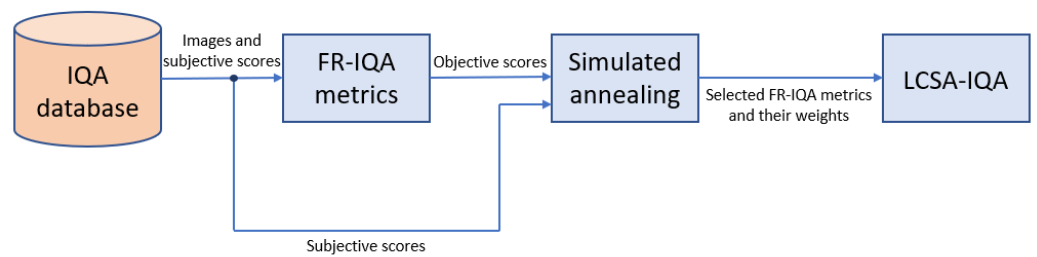
3. Proposed Method

As already mentioned, an FR-IQA metric should deliver perceptual quality scores consistent with the human judgement using both the distorted and reference images. Let us express the aggregated decision of n different FR-IQA metrics by a weighted sum as:

$$Q = \sum_{i=1}^{n} \alpha_i q_i, \tag{1}$$

where $q_i (i = 1, 2, \ldots, n)$ stands for the quality scores provided by the FR-IQA metrics. Further, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ is a real vector of weights whose values are found via an optimization procedure to ensure an effective fusion of FR-IQA metrics. Namely, an optimization fusion was carried out in our study using $n = 16$ open-source FR-IQA metrics, such as FSIM [62], FSIMc [62], GSM [63], IFC [38], IFS [64], IW-SSIM [33], MAD [25], MS-SSIM [34], NQM [65], PSNR, RFSIM [66], SFF [67], SR-SIM [12], SSIM [29], VIF [39], and VSI [68].

In the literature, Pearson's linear correlation coefficient (PLCC), Spearman's rank-order correlation coefficient (SROCC), Kendall's rank order correlation coefficient (KROCC), and root mean square error (RMSE) are often considered to characterize the consistency between the ground-truth quality scores of an IQA benchmark database and the quality scores predicted by an FR-IQA metric [22]. From these performance indices, RMSE was applied as an objective function in the proposed optimization based metric. Figures 1 and 2 depict flowcharts where the compilation of the proposed fusion-based metrics and its application for FR-IQA are demonstrated.

Figure 1. In the offline optimization stage, the proposed fusion-based metric is obtained by using 20% of the reference with its corresponding distorted counterparts. Next, a simulated annealing (SA) optimization process selects FR-IQA metrics and provides them with weights. The resulting metric is codenamed as LCSA-IQA to refer to the fact that is the linear combination of selected FR-IQA metrics where the weights were assigned using simulated annealing.
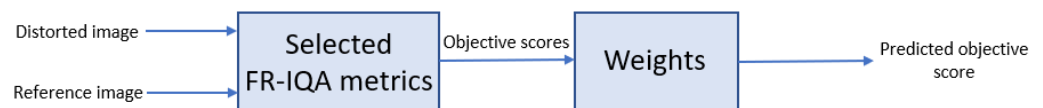
Figure 2. The optimal linear combination of the selected FR-IQA metrics is applied to estimate perceptual image quality.

Formally, the optimization problem can be written as

$$\min_{\alpha} RMSE(F(\mathbf{Q}_p, \boldsymbol{\beta}), \mathbf{S}),$$
$$\text{subject to } \alpha_i \in \mathbb{R}, n \in \mathbb{N}, \beta \geq 0, \tag{2}$$

where $\mathbf{Q}_p$ is vector containing the quality scores of a set of images obtained by Equation (1) and $\mathbf{S}$ contains the corresponding ground-truth scores. Further, prior to the calculation of

RMSE, a non-linear regression is also applied [22] since a non-linear relationship exists between the ground-truth and predicted scores. Formally, it can be written

$$Q = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(Q_p - \beta_3)}} \right) + \beta_4 Q_p + \beta_5, \tag{3}$$

where $\beta_1, ..., \beta_5$ stand for the parameters of the regression model. In addition, Q and Q_p are the fitted and predicted scores, respectively. Since we use four large, widely accepted IQA benchmark databases, i.e., LIVE [22], TID2013 [23], TID2008 [24], and CSIQ [25], in this paper, four optimization-based fusion FR-IQA metrics are proposed, respectively. To this end, approximately 20% of the reference images were randomly selected from a given benchmark IQA database. More precisely, **Q** and **S** were compiled based on those distorted images whose reference counterparts were randomly selected. Although 20% is a common choice for parameter setting in the literature [69,70], there are also researchers who applied 30% [62] or 80% [71] for parameter tuning. However, we evaluate all the fusion based metrics on all the databases to demonstrate results independent from the database.

Next, the optimization problem was solved described by Equation (2) to determine the α_i weights for Equation (1). Since the number of possible solutions increases exponentially with number of the considered FR-IQA metrics, simulated annealing (SA) [72,73] was used to solve the above-described optimization task. Namely, SA is a probabilistic optimization technique for estimating the global optimum of a given function. The stochastic nature of this algorithm enables the usage of nonlinear objective functions where many other methods do not operate well. SA was inspired by the physical model of heating a material and then slowly decreasing the temperature to eliminate imperfections from the material. Hence, minimizing the system's energy is the main goal. More precisely, the SA randomly generates a new point at each iteration. Based on a probability distribution with a scale proportional to the temperature, the new point's distance from the present point or the size of the search is determined. All new points that reduce the objective are accepted by the algorithm, but points that increase the objective can also be accepted with a pre-defined probability. Due to this property of the method, SA is prevented from being stuck in local minima in early iterations. In our implementation, the SA was performed using MATLAB R2020a with a Global Optimization Toolbox using $\alpha_i = 0$ for $i = 1, 2, ..., n$ as initial point and defining no lower or upper bounds for the method. After 100 runs of SA, the best solution—$\boldsymbol{\alpha}_d^{best}$—was selected, where d denotes the database from which 20% of the reference images was chosen randomly.

In the end of the SA optimization processes using LIVE [22], TID2013 [23], TID2008 [24], and CSIQ [25] databases, the following FR-IQA metrics can be obtained, which are code-named LCSA, referring to the fact that they are linear combinations of FR-IQA measures selected by simulated annealing:

$$\begin{aligned} LCSA1 = \boldsymbol{\alpha}_{LIVE}^{best} = {}&-561.0123 \cdot VSI + 281.826 \cdot FSIMc - 116.1501 \cdot IFC - 846.6376 \cdot MAD \\ &+349.6191 \cdot MSSSIM - 262.6766 \cdot NQM + 41.6348 \cdot PSNR - 308.9426 \cdot SSIM \\ &+722.4479 \cdot VIF, \end{aligned} \tag{4}$$

$$\begin{aligned} LCSA2 = \boldsymbol{\alpha}_{TID2013}^{best} = {}&1774.8368 \cdot VSI + 467.5433 \cdot FSIMc - 332.1863 \cdot GSM - 63.4379 \cdot IFC \\ &+84.7954 \cdot IWSSIM - 346.5585 \cdot MAD - 126.5188 \cdot NQM + 381.0923 \cdot PSNR \\ &-626.9841 \cdot SSIM + 380.3341 \cdot VIF + 524.6484 \cdot IFS + 342.7968 \cdot SFF, \end{aligned} \tag{5}$$

$$\begin{aligned} LCSA3 = \boldsymbol{\alpha}_{TID2008}^{best} = {}&1253.2402 \cdot VSI + 217.0877 \cdot IWSSIM - 168.1779 \cdot MAD \\ -75.6832 \cdot NQM + 276.9035 \cdot PSNR - 28.5915 \cdot RFSIM - {}&454.7619 \cdot SSIM + 203.0893 \cdot VIF \\ &+500.4323 \cdot IFS - 153.3686 \cdot SFF, \end{aligned} \tag{6}$$

$$LCSA4 = \alpha_{CSIQ}^{best} = 266.3256 \cdot FSIM - 119.8937 \cdot FSIMc - 15.6937 \cdot IWSSIM$$
$$-529.1806 \cdot MAD - 656.4991 \cdot MSSSIM - 73.009 \cdot NQM + 381.0923 \cdot PSNR \qquad (7)$$
$$-626.9841 \cdot SSIM + 380.3341 \cdot VIF + 524.6484 \cdot IFS + 342.7968 \cdot SFF.$$

The corresponding β vectors are as follows:

$$\beta_{LIVE} = (106.1735, 36.8421, 30.0447, 15.7705, 139.3613), \qquad (8)$$

$$\beta_{TID2013} = (56.413, 193.7249, 14.9834, 147.7736, 89.8778), \qquad (9)$$

$$\beta_{TID2008} = (13.4153, 115.9834, 45.4464, 22.0253, 269.7624), \qquad (10)$$

$$\beta_{CSIQ} = (13.5361, 105.4132, 70.1095, 150.7645, 11.5291). \qquad (11)$$

4. Results

In this section, our experimental results are presented. First, the applied IQA benchmark databases and evaluation protocol are described in Section 4.1. Next, Section 4.2 presents a comparison to other competing state-of-the-art methods on four large IQA benchmark databases, i.e., LIVE [22], TID2013 [23], TID2008 [24], and CSIQ [25].

4.1. Applied IQA Benchmark Databases and Evaluation Protocol

The main properties of the applied IQA benchmark databases are outlined in Table 1. These databases consist of a set of reference images, whose visual quality are considered perfect and flawless. Further, distorted images are generated artificially from the reference images using different distortion types (i.e., JPEG compression noise, JPEG2000 compression noise, salt and pepper, motion blur, Gaussian, Poisson, etc.) at different distortion levels. Figure 3 depicts the empirical MOS distributions of the applied benchmark databases.
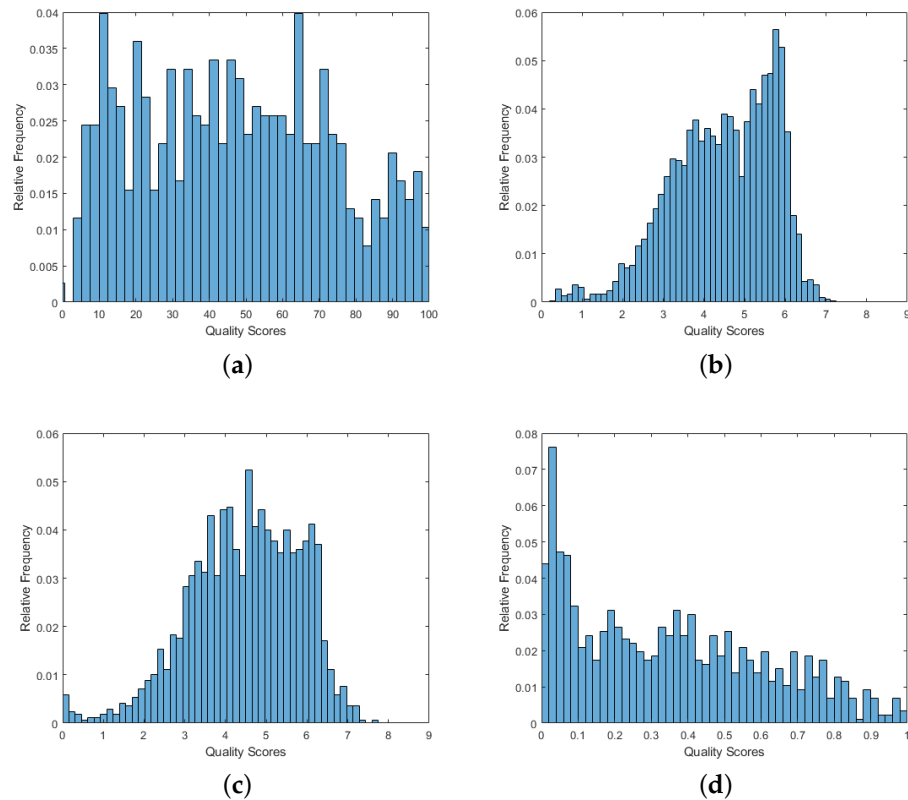
Figure 3. Empirical MOS distributions in the used benchmark IQA databases: (**a**) LIVE, (**b**) TID2013, (**c**) TID2008, and (**d**) CSIQ.

Table 1. Summary of benchmark databases used in this study.

	LIVE [22]	TID2013 [23]	TID2008 [24]	CSIQ [25]
No. of reference images	29	25	25	30
No. of distorted images	779	3000	1700	866
No. of distortions	5	24	17	6
No. of levels	5	5	4	4-5
No. of observers	161	917	838	35
Resolution	$\sim 768 \times 512$	512×384	512×384	500×500

In the literature, PLCC, SROCC, and KROCC is widely used and accepted to characterize the performance of FR-IQA methods. They are measured between the ground-truth scores of an IQA benchmark database and the predicted scores. Moreover, prior to the calculation of PLCC a non-linear regression is also applied [22] since a non-linear relationship exists between the ground-truth and predicted scores. This non-linear relationship was also defined by Equation (3). Further, Q and Q_p are the fitted and predicted scores, respectively. PLCC between vectors $\mathbf{x}$ and $\mathbf{y}$ with length m is defined as

$$PLCC(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}^T \mathbf{y}}{\sqrt{\bar{\mathbf{x}}^T \bar{\mathbf{y}} \bar{\mathbf{y}}^T \bar{\mathbf{x}}}}, \tag{12}$$

where $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ are the mean subtracted version of vectors $\mathbf{x}$ and $\mathbf{y}$, respectively. On the other hand, SROCC can be defined as

$$SROCC(\mathbf{x}, \mathbf{y}) = 1 - \frac{6 \sum_{i=1}^{m} (x_i - y_i)^2}{m(m^2 - 1)}, \tag{13}$$

where x_i and y_i are the ith entries of vectors $\mathbf{x}$ and $\mathbf{y}$, respectively. In contrast, KROCC uses the number of concordant pairs (m_c) and the number of discordant pairs (m_d) between vectors $\mathbf{x}$ and $\mathbf{y}$ and is defined as

$$KROCC(\mathbf{x}, \mathbf{y}) = \frac{m_c - m_d}{\frac{1}{2}m(m-1)}. \tag{14}$$

As already mentioned, the proposed fusion-based metrics were implemented using MATLAB R2020a and its Global Optimization Toolbox. The computer configuration applied in our experiments is summarized in Table 2.

Table 2. Computer configuration applied in our experiments.

Computer model	STRIX Z270H Gaming
Operating system	Windows 10
Memory	15 GB
CPU	Intel(R) Core(TM) i7-7700K CPU 4.20 GHz (8 cores)
GPU	Nvidia GeForce GTX 1080

4.2. Comparison to the State-of-the-Art

In this subsection, the proposed fusion-based metrics are compared to several state-of-the-art FR-IQA whose original source codes were made publicly available by the authors. Moreover, we reimplemented the fusion-based SSIM-CNN [51] method in MATLAB R2020a (available at: https://github.com/Skythianos/SSIM-CNN (accessed on 12 May 2022)). The PLCC, SROCC, and KROCC performance comparisons of the proposed fusion-based FR-IQA metrics with the state-of-the-art are summarized in Tables 3 and 4. Specifically, Table 3 demonstrates the results on LIVE [22] and TID2013 [23], while Table 4 contains the

obtained results for TID2008 [24] and CSIQ [25] databases. The obtained results clearly show that the proposed *LCSA* metrics are able to outperform the state-of-the-art. Specifically, those *LCSA* metrics that were parameter-tuned on database *d* always deliver the highest correlation values, while another *LCSA* not parameter-tuned on database *d* usually provides the second-best results.

Table 3. PLCC, SROCC, and KROCC performance comparison of the proposed fusion-based FR-IQA metrics with the state-of-the-art. The best results are typed in bold, and the second best results are underlined.

	LIVE [22]			TID2013 [23]		
FR-IQA Metric	**PLCC**	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**
2stepQA [74]	0.937	0.932	0.828	0.736	0.733	0.550
CSV [75]	<u>0.967</u>	0.959	0.834	0.852	0.848	0.657
DISTS [76]	0.954	0.954	0.811	0.759	0.711	0.524
ESSIM [77]	0.963	0.962	0.840	0.740	0.797	0.627
FSIM [62]	0.960	0.963	0.833	0.859	0.802	0.629
FSIMc [62]	0.961	0.965	0.836	0.877	0.851	0.667
GSM [63]	0.944	0.955	0.831	0.789	0.787	0.593
IFC [38]	0.927	0.926	0.758	0.554	0.539	0.394
IFS [64]	0.959	0.960	0.825	0.879	0.870	0.679
IW-SSIM [33]	0.952	0.956	0.817	0.832	0.778	0.598
MAD [25]	<u>0.967</u>	0.967	0.842	0.827	0.778	0.600
MS-SSIM [34]	0.941	0.951	0.804	0.794	0.785	0.604
NQM [65]	0.912	0.909	0.741	0.690	0.643	0.474
PSNR	0.872	0.876	0.687	0.616	0.646	0.467
ReSIFT [78]	0.961	0.962	0.838	0.630	0.623	0.471
RFSIM [66]	0.935	0.940	0.782	0.833	0.774	0.595
RVSIM [79]	0.641	0.630	0.495	0.763	0.683	0.520
SFF [67]	0.963	0.965	0.836	0.871	0.851	0.658
SR-SIM [12]	0.955	0.962	0.829	0.859	0.800	0.631
SSIM [29]	0.941	0.951	0.804	0.618	0.616	0.437
SSIM-CNN [51]	0.965	0.963	0.838	0.759	0.752	0.566
SUMMER [80]	<u>0.967</u>	0.959	0.833	0.623	0.622	0.472
VIF [39]	0.941	0.964	0.828	0.774	0.677	0.515
VSI [68]	0.948	0.952	0.805	<u>0.900</u>	<u>0.894</u>	<u>0.677</u>
LCSA1	**0.974**	**0.974**	**0.857**	0.820	0.788	0.607
LCSA2	0.846	0.962	0.828	**0.916**	**0.903**	**0.731**
LCSA3	0.947	0.969	0.843	0.770	0.821	0.647
LCSA4	<u>0.967</u>	<u>0.970</u>	<u>0.844</u>	0.859	0.823	0.649

Table 4. PLCC, SROCC, and KROCC performance comparison of the proposed fusion-based FR-IQA metrics with the state-of-the-art. The best results are typed in bold, and the second best results are underlined.

	TID2008 [24]			CSIQ [25]		
FR-IQA Metric	**PLCC**	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**
2stepQA [74]	0.757	0.769	0.574	0.841	0.849	0.655
CSV [75]	0.852	0.848	0.657	0.933	0.933	0.766
DISTS [76]	0.705	0.668	0.488	0.930	0.930	0.764
ESSIM [77]	0.658	0.876	0.696	0.814	0.933	0.768
FSIM [62]	0.874	0.881	0.695	0.912	0.924	0.757
FSIMc [62]	0.876	0.884	0.699	0.919	0.931	0.769
GSM [63]	0.782	0.781	0.578	0.896	0.911	0.737
IFC [38]	0.575	0.568	0.424	0.837	0.767	0.590
IFS [64]	0.879	0.869	0.678	0.958	0.958	0.817
IW-SSIM [33]	0.842	0.856	0.664	0.804	0.921	0.753
MAD [25]	0.831	0.829	0.639	0.950	0.947	0.797
MS-SSIM [34]	0.838	0.846	0.648	0.899	0.913	0.739
NQM [65]	0.608	0.624	0.461	0.743	0.740	0.564
PSNR	0.447	0.489	0.346	0.853	0.809	0.599
ReSIFT [78]	0.627	0.632	0.484	0.884	0.868	0.695
RFSIM [66]	0.865	0.868	0.678	0.912	0.930	0.765

Table 4. *Cont.*

FR-IQA Metric	TID2008 [24]			CSIQ [25]		
	PLCC	SROCC	KROCC	PLCC	SROCC	KROCC
RVSIM [79]	0.789	0.743	0.566	0.923	0.903	0.728
SFF [67]	0.871	0.851	0.658	0.964	0.960	0.826
SR-SIM [12]	0.859	0.799	0.631	0.925	0.932	0.773
SSIM [29]	0.669	0.675	0.485	0.812	0.812	0.606
SSIM-CNN [51]	0.770	0.737	0.551	0.952	0.946	0.794
SUMMER [80]	0.817	0.823	0.623	0.826	0.830	0.658
VIF [39]	0.808	0.749	0.586	0.928	0.920	0.754
VSI [68]	0.898	0.896	0.709	0.928	0.942	0.785
LCSA1	0.886	0.874	0.685	<u>0.966</u>	0.956	0.819
LCSA2	0.896	0.906	0.727	0.897	0.949	0.800
LCSA3	**0.923**	**0.921**	**0.755**	0.964	<u>0.961</u>	<u>0.827</u>
LCSA4	<u>0.906</u>	<u>0.909</u>	<u>0.737</u>	**0.977**	**0.973**	**0.857**

Table 5 illustrates the direct and weighted average of correlation values measured on LIVE [22], TID2013 [23], TID2008 [24], and CSIQ [25]. From the results of direct averages, it can be clearly seen that the proposed *LCSA*2 and *LCSA*4 provide the best results in two out of three performance indices, while *LCSA*3 is able to produce second best KROCC value. The results of weighted averages are biased towards those FR-IQA measures that perform well on TID2013 [23] since it is the largest database from the applied benchmarks. Similarly, *LCSA*2 is the best-performing method in this respect because it provides the best results for SROCC and KROCC. Further, *LCSA*4 delivers the second best PLCC and KROCC values, while *LCSA*3's performance is equivalent in terms of SROCC and KROCC to those of *LCSA*4.

Table 5. PLCC, SROCC, and KROCC performance comparison of the proposed fusion-based FR-IQA metrics with the state-of-the-art. The best results are typed in bold, the second best results are underlined.

FR-IQA Metric	Direct Average			Weighted Average		
	PLCC	SROCC	KROCC	PLCC	SROCC	KROCC
2stepQA [74]	0.818	0.821	0.652	0.781	0.783	0.605
CSV [75]	0.901	0.897	0.729	0.877	0.873	0.694
DISTS [76]	0.837	0.816	0.647	0.792	0.759	0.582
ESSIM [77]	0.794	0.892	0.733	0.756	0.857	0.691
FSIM [62]	0.901	0.893	0.729	0.883	0.860	0.689
FSIMc [62]	0.908	0.908	0.743	0.893	0.885	0.710
GSM [63]	0.853	0.859	0.685	0.821	0.823	0.638
IFC [38]	0.723	0.700	0.542	0.644	0.625	0.473
IFS [64]	0.919	0.914	0.750	0.900	0.893	0.715
IW-SSIM [33]	0.857	0.878	0.708	0.846	0.840	0.664
MAD [25]	0.894	0.880	0.720	0.862	0.838	0.667
MS-SSIM [34]	0.868	0.874	0.699	0.838	0.839	0.659
NQM [65]	0.738	0.729	0.560	0.703	0.684	0.516
PSNR	0.697	0.705	0.525	0.634	0.654	0.480
ReSIFT [78]	0.776	0.771	0.622	0.705	0.700	0.550
RFSIM [66]	0.886	0.878	0.705	0.865	0.841	0.663
RVSIM [79]	0.779	0.740	0.577	0.777	0.723	0.558
SFF [67]	0.917	0.908	0.745	0.895	0.880	0.703
SR-SIM [12]	0.900	0.873	0.716	0.880	0.838	0.675
SSIM [29]	0.760	0.764	0.583	0.698	0.700	0.518
SSIM-CNN [51]	0.861	0.849	0.687	0.814	0.800	0.626
SUMMER [80]	0.808	0.809	0.647	0.745	0.746	0.582
VIF [39]	0.863	0.828	0.671	0.825	0.765	0.605
VSI [68]	<u>0.919</u>	<u>0.921</u>	0.744	**0.909**	<u>0.908</u>	0.716

Table 5. *Cont.*

	Direct Average			Weighted Average		
FR-IQA Metric	**PLCC**	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**
LCSA1	0.912	0.898	0.742	0.877	0.857	0.688
LCSA2	0.889	**0.930**	**0.772**	0.899	**0.917**	**0.751**
LCSA3	0.901	0.918	<u>0.768</u>	0.859	0.885	<u>0.725</u>
LCSA4	**0.927**	0.919	**0.772**	<u>0.901</u>	0.885	<u>0.725</u>

In the following, we examine the performance of the proposed and the other state-of-the-art methods on the individual distortion types of the applied IQA benchmark databases. The distortion types and their abbreviations used by the databases are summarized in Table 6. Further, Tables 7–10 contain detailed results on the different distortion types of LIVE [22], TID2013 [23], TID2008 [24], and CSIQ [25], respectively. To be more specific, the SROCC values are given for each individual distortion types.

Table 6. Distortion types used in the applied benchmark IQA databases (LIVE [22], TID2013 [23], TID2008 [24], and CSIQ [25]).

Abbreviation	Description	LIVE [22]	TID2013 [23]	TID2008 [24]	CSIQ [25]
AGN	additive Gaussian noise	✓	✓	✓	✓
ANC	additive noise in color components		✓	✓	✓
SCN	spatially correlated noise		✓	✓	
MN	masked noise		✓	✓	
HFN	high-frequency noise		✓	✓	
IN	impulse noise		✓	✓	
QN	quantization noise		✓	✓	
FF	simulated fast fading Rayleigh channel	✓			
GB	Gaussian blur	✓	✓	✓	
GCD	global contrast decrement				✓
DEN	image denoising		✓		
JPEG	JPEG compression noise	✓	✓	✓	✓
JP2K	JPEG2000 compression noise	✓	✓	✓	✓
JGTE	JPEG transmission errors		✓	✓	
J2TE	JPEG2000 transmission errors		✓	✓	
NEPN	non-eccentricity pattern noise		✓	✓	
BLOCK	local block-wise distortions of different intensity		✓	✓	
MS	mean shift		✓	✓	
CC	contrast change		✓	✓	
CCS	change of color saturation		✓		
MGN	multiplicative Gaussian noise		✓		
CN	comfort noise		✓		
LCNI	lossy compression of noisy images		✓		
ICQD	image color quantization with dither		✓		
CA	chromatic aberration		✓		
SSR	sparse sampling and reconstruction		✓		

Table 7. Comparison on LIVE's [22] distortion types. SROCC values are given. The highest values are typed in bold, while the second highest ones are underlined.

	FSIM	FSIMc	IFS	MS-SSIM	SFF	VIF	VSI	LCSA1	LCSA2	LCSA3	LCSA4
AGN	0.965	0.972	**0.988**	0.973	<u>0.986</u>	<u>0.986</u>	0.984	0.976	0.961	0.962	0.965
FF	0.950	0.952	0.940	0.947	0.953	0.965	0.943	<u>0.984</u>	0.978	**0.988**	0.980
GB	0.971	0.971	0.967	0.954	0.975	0.973	0.953	0.978	0.989	**0.997**	<u>0.996</u>
JPEG	0.983	<u>0.984</u>	0.978	0.982	0.979	**0.985**	0.976	0.974	0.973	0.964	0.965
JP2K	<u>0.972</u>	0.970	0.969	0.963	0.967	0.970	0.960	0.952	0.969	0.967	**0.978**
All	0.963	0.965	0.960	0.951	0.965	0.964	0.952	**0.974**	0.962	0.969	<u>0.970</u>

Table 8. Comparison on TID2013's [23] distortion types. SROCC values are given. The highest values are typed in bold, while the second highest ones are underlined.

	FSIM	FSIMc	IFS	MS-SSIM	SFF	VIF	VSI	LCSA1	LCSA2	LCSA3	LCSA4
AGN	0.897	0.910	<u>0.938</u>	0.865	0.907	0.899	**0.946**	0.908	0.932	0.925	0.925
ANC	0.821	0.854	0.854	0.773	0.817	0.830	**0.871**	0.846	0.854	0.853	<u>0.857</u>
SCN	0.875	0.890	0.934	0.854	0.898	0.884	<u>0.937</u>	0.908	**0.940**	0.933	0.915
MN	0.794	0.809	0.796	0.807	<u>0.819</u>	**0.845**	0.770	0.792	0.769	0.811	0.801
HFN	0.898	0.904	<u>0.914</u>	0.860	0.898	0.897	**0.920**	0.904	<u>0.914</u>	0.909	0.903
IN	0.807	0.825	0.839	0.763	0.787	<u>0.854</u>	**0.874**	0.574	0.795	0.790	0.728
QN	0.872	**0.881**	0.834	0.871	0.861	0.785	<u>0.875</u>	0.854	0.886	0.844	0.863
GB	0.955	0.955	0.966	<u>0.967</u>	**0.968**	0.965	0.961	0.954	0.956	0.959	0.970
DEN	0.930	0.933	0.918	0.927	0.909	0.891	**0.948**	0.917	<u>0.937</u>	0.913	<u>0.937</u>
JPEG	0.932	<u>0.934</u>	0.929	0.927	0.927	0.919	**0.954**	0.921	0.930	0.929	0.932
JP2K	0.958	0.959	0.961	0.950	0.957	0.952	**0.971**	0.950	<u>0.965</u>	0.957	0.953
JGTE	0.846	0.861	0.893	0.848	0.883	0.841	**0.922**	0.854	<u>0.891</u>	0.863	0.859
J2TE	0.891	0.892	0.901	0.889	0.871	0.876	**0.923**	0.909	<u>0.916</u>	0.913	<u>0.916</u>
NEPN	0.792	0.794	0.784	0.797	0.767	0.772	0.806	**0.826**	0.815	0.815	<u>0.822</u>
BLOCK	<u>0.549</u>	**0.553**	0.100	0.480	0.179	0.531	0.171	0.452	0.353	0.328	0.185
MS	0.753	0.749	0.658	**0.791**	0.665	0.628	<u>0.770</u>	0.554	0.678	0.455	0.620
CC	0.469	0.468	0.447	0.463	0.469	**0.839**	0.475	0.535	0.448	<u>0.631</u>	0.423
CCS	0.275	**0.836**	0.826	0.410	0.827	0.310	0.810	0.712	<u>0.829</u>	0.813	0.813
MGN	0.847	0.857	0.879	0.779	0.843	0.847	**0.912**	0.875	<u>0.900</u>	0.882	0.875
CN	0.912	0.914	0.904	0.853	0.901	0.895	**0.924**	0.911	<u>0.923</u>	0.904	0.906
LCNI	0.947	0.949	0.943	0.907	0.926	0.920	0.956	0.951	**0.958**	0.945	<u>0.957</u>
ICQD	0.876	0.882	<u>0.901</u>	0.856	0.880	0.841	0.884	0.891	**0.903**	0.891	0.900
CA	0.872	**0.893**	0.886	0.878	0.879	0.885	<u>0.891</u>	0.862	0.873	0.870	0.874
SSR	0.957	0.958	0.956	0.948	0.952	0.935	<u>0.963</u>	0.948	0.957	**0.965**	0.955
All	0.802	0.851	0.870	0.785	0.851	0.677	<u>0.894</u>	0.788	**0.903**	0.821	0.823

Table 9. Comparison on TID2008's [24] distortion types. SROCC values are given. The highest values are typed in bold, while the second highest ones are underlined.

	FSIM	FSIMc	IFS	MS-SSIM	SFF	VIF	VSI	LCSA1	LCSA2	LCSA3	LCSA4
AGN	0.857	0.876	0.917	0.809	0.873	0.880	**0.923**	0.887	<u>0.916</u>	0.906	0.905
ANC	0.853	0.893	0.896	0.805	0.863	0.876	**0.912**	0.887	0.890	0.893	0.889
SCN	0.848	0.871	<u>0.931</u>	0.821	0.894	0.870	0.930	0.894	0.915	**0.936**	0.918
MN	0.802	0.826	0.802	0.811	0.837	**0.868**	0.773	0.782	0.733	<u>0.857</u>	0.817
HFN	0.909	0.916	<u>0.922</u>	0.869	0.912	0.908	**0.925**	0.901	0.909	<u>0.922</u>	0.917
IN	0.745	0.772	0.814	0.691	0.748	**0.833**	<u>0.830</u>	0.396	0.729	0.752	0.618
QN	0.856	**0.873**	0.797	<u>0.859</u>	0.845	0.797	**0.873**	0.825	<u>0.859</u>	0.855	0.854
GB	0.947	0.947	0.960	0.956	<u>0.962</u>	0.954	0.953	0.933	0.944	0.953	**0.963**
DEN	0.960	0.962	0.949	0.958	0.938	0.916	**0.969**	0.936	0.956	<u>0.964</u>	0.963
JPEG	0.928	0.929	0.928	0.932	0.932	0.917	**0.962**	0.921	<u>0.942</u>	0.939	0.937
JP2K	0.977	0.978	0.978	0.970	0.977	0.977	0.985	0.975	**0.991**	<u>0.986</u>	0.977
JGTE	0.871	0.876	0.874	0.868	0.857	0.859	**0.916**	0.886	<u>0.914</u>	0.893	0.904
J2TE	0.854	0.856	0.878	0.861	0.839	0.850	0.894	0.889	0.885	**0.911**	<u>0.901</u>
NEPN	0.749	0.751	0.704	0.738	0.697	0.762	0.770	**0.831**	0.773	<u>0.805</u>	0.796
BLOCK	**0.849**	<u>0.846</u>	0.087	0.755	0.537	0.832	0.630	0.826	0.631	0.742	0.672
MS	**0.672**	0.655	0.522	0.734	0.523	0.510	<u>0.671</u>	0.460	0.383	0.554	0.497
CC	0.648	0.651	0.627	0.638	0.646	**0.819**	0.656	0.630	0.604	<u>0.732</u>	0.577
All	0.881	0.884	0.869	0.846	0.851	0.749	0.896	0.874	0.906	**0.921**	<u>0.909</u>

Table 10. Comparison on CSIQ's [25] distortion types. SROCC values are given. The highest values are typed in bold, while the second highest ones are underlined.

	FSIM	FSIMc	IFS	MS-SSIM	SFF	VIF	VSI	LCSA1	LCSA2	LCSA3	LCSA4
AGN	0.926	0.936	0.959	0.947	0.947	0.958	0.964	0.965	<u>0.971</u>	0.967	**0.976**
ANC	0.923	0.937	0.953	0.933	0.955	0.951	<u>0.964</u>	0.912	0.948	0.962	**0.969**
GB	0.973	0.973	0.962	0.971	0.975	0.975	0.968	**0.983**	0.972	0.971	<u>0.981</u>
GCD	0.942	0.944	0.949	0.953	0.954	0.935	0.950	**0.975**	0.959	<u>0.972</u>	0.963
JPEG	0.965	0.966	0.966	0.963	0.964	0.971	0.962	0.967	**0.983**	<u>0.981</u>	0.979
JP2K	0.968	0.970	<u>0.971</u>	0.968	**0.976**	0.967	0.969	0.956	0.950	0.941	0.950
All	0.924	0.931	0.958	0.913	0.960	0.920	0.942	0.956	0.949	<u>0.961</u>	**0.973**

5. Conclusions

In this study, we presented a novel fusion-based FR-IQA metric using simulated annealing. Specifically, an optimization problem was solved based on the weighted sum of

several FR-IQA metrics by minimizing the root mean squared error between the predicted and ground-truth perceptual quality scores. The evaluation of the proposed fusion-based metrics on four large publicly available and widely accepted IQA benchmark databases empirically corroborated that the proposed metrics are able to produce competitive results compared to the state-of-the-art in terms of various performance indices, such as PLCC, SROCC, and KROCC. Future research could involve other optimization techniques and their combination for improved perceptual quality prediction. Another direction is the generalization of the proposed method for other types of media.

Abbreviations

The following abbreviations are used in this manuscript:

FR-IQA	full-reference image quality assessment
IQA	image quality assessment
KROCC	Kendall's rank order correlation coefficient
MOS	mean opinion score
MSE	mean squared error
NR-IQA	no-reference image quality assessment
PLCC	Pearson's linear correlation coefficient
PSNR	peak signal-to-noise ratio
RMSE	root mean square error
RR-IQA	reduced-reference image quality assessment
SA	simulated annealing
SROCC	Spearman's rank order correlation coefficient
SSIM	structural similarity index

References

1. Chubarau, A.; Akhavan, T.; Yoo, H.; Mantiuk, R.K.; Clark, J. Perceptual image quality assessment for various viewing conditions and display systems. *Electron. Imaging* **2020**, *2020*, 67-1. [CrossRef]
2. Tao, L.; Zhu, C.; Xiang, G.; Li, Y.; Jia, H.; Xie, X. LLCNN: A convolutional neural network for low-light image enhancement. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
3. Rehman, A.; Rostami, M.; Wang, Z.; Brunet, D.; Vrscay, E.R. SSIM-inspired image restoration using sparse representation. *EURASIP J. Adv. Signal Process.* **2012**, *2012*, 1–12. [CrossRef]
4. Setiadi, D.R.I.M. PSNR vs SSIM: Imperceptibility quality assessment for image steganography. *Multimed. Tools Appl.* **2021**, *80*, 8423–8444. [CrossRef]
5. Goyal, B.; Gupta, A.; Dogra, A.; Koundal, D. An adaptive bitonic filtering based edge fusion algorithm for Gaussian denoising. *Int. J. Cogn. Comput. Eng.* **2022**, *3*, 90–97. [CrossRef]
6. Wang, Z. Applications of objective image quality assessment methods [applications corner]. *IEEE Signal Process. Mag.* **2011**, *28*, 137–142. [CrossRef]
7. Kalender, W.A. *Computed Tomography: Fundamentals, System Technology, Image Quality, Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2011.

8.   Kaur, B.; Dogra, A.; Goyal, B. Comparative Analysis of Bilateral Filter and its Variants for Magnetic Resonance Imaging. *Open Neuroimaging J.* **2020**, *13*, 21–29. [CrossRef]

9.   Saupe, D.; Hahn, F.; Hosu, V.; Zingman, I.; Rana, M.; Li, S. Crowd workers proven useful: A comparative study of subjective video quality assessment. In Proceedings of the QoMEX 2016: 8th International Conference on Quality of Multimedia Experience, Lisbon, Portugal, 6–8 June 2016.

10. Lin, H.; Hosu, V.; Saupe, D. KADID-10k: A large-scale artificially distorted IQA database. In Proceedings of the 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany, 5–7 June 2019; pp. 1–3.

11. Ciocca, G.; Corchs, S.; Gasparini, F.; Schettini, R. How to assess image quality within a workflow chain: An overview. *Int. J. Digit. Libr.* **2014**, *15*, 1–25. [CrossRef]

12. Zhang, L.; Li, H. SR-SIM: A fast and high performance IQA index based on spectral residual. In Proceedings of the 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; pp. 1473–1476.

13. Soundararajan, R.; Bovik, A.C. RRED indices: Reduced reference entropic differencing for image quality assessment. *IEEE Trans. Image Process.* **2011**, *21*, 517–526. [CrossRef]

14. Min, X.; Zhai, G.; Gu, K.; Liu, Y.; Yang, X. Blind image quality estimation via distortion aggravation. *IEEE Trans. Broadcast.* **2018**, *64*, 508–517. [CrossRef]

15. Min, X.; Gu, K.; Zhai, G.; Liu, J.; Yang, X.; Chen, C.W. Blind quality assessment based on pseudo-reference image. *IEEE Trans. Multimed.* **2017**, *20*, 2049–2062. [CrossRef]

16. Bouida, A.; Khelifi, M.; Beladgham, M.; Hamlili, F.Z. Monte Carlo Optimization of a Combined Image Quality Assessment for Compressed Images Evaluation. *Trait. Du Signal* **2021**, *38*, 281–289. [CrossRef]

17. Merzougui, N. Multi-measures fusion based on multi-objective genetic programming for full-reference image quality assessment. *arXiv* **2017**, arXiv:1801.06030.

18. Okarma, K. On the usefulness of combined metrics for 3D image quality assessment. In *Image Processing & Communications Challenges 6*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 137–144.

19. Okarma, K.; Lech, P.; Lukin, V.V. Combined Full-Reference Image Quality Metrics for Objective Assessment of Multiply Distorted Images. *Electronics* **2021**, *10*, 2256. [CrossRef]

20. Oszust, M. Full-reference image quality assessment with linear combination of genetically selected quality measures. *PLoS ONE* **2016**, *11*, e0158333. [CrossRef]

21. Soares, S.; Antunes, C.H.; Araújo, R. Comparison of a genetic algorithm and simulated annealing for automatic neural network ensemble development. *Neurocomputing* **2013**, *121*, 498–511. [CrossRef]

22. Sheikh, H.R.; Sabir, M.F.; Bovik, A.C. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.* **2006**, *15*, 3440–3451. [CrossRef]

23. Ponomarenko, N.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Jin, L.; Astola, J.; Vozel, B.; Chehdi, K.; Carli, M.; Battisti, F.; et al. Color image database TID2013: Peculiarities and preliminary results. In Proceedings of the European Workshop on Visual Information Processing (EUVIP), Paris, France, 10–12 June 2013; pp. 106–111.

24. Ponomarenko, N.; Lukin, V.; Zelensky, A.; Egiazarian, K.; Carli, M.; Battisti, F. TID2008-a database for evaluation of full-reference visual quality assessment metrics. *Adv. Mod. Radioelectron.* **2009**, *10*, 30–45.

25. Larson, E.C.; Chandler, D.M. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *J. Electron. Imaging* **2010**, *19*, 011006.

26. Ding, K.; Ma, K.; Wang, S.; Simoncelli, E.P. Comparison of full-reference image quality models for optimization of image processing systems. *Int. J. Comput. Vis.* **2021**, *129*, 1258–1281. [CrossRef]

27. Sara, U.; Akter, M.; Uddin, M.S. Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study. *J. Comput. Commun.* **2019**, *7*, 8–18. [CrossRef]

28. Saupe, D.; Hamzaoui, R.; Hartenstein, H. *Fractal Image Compression: An Introductory Overview*; Universität Wien Fakultät für Informatik: Wien, Austria, 1997.

29. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]

30. Wang, S.; Rehman, A.; Wang, Z.; Ma, S.; Gao, W. SSIM-motivated rate-distortion optimization for video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2011**, *22*, 516–529. [CrossRef]

31. Liu, L.; Wang, Y.; Wu, Y. A wavelet-domain structure similarity for image quality assessment. In Proceedings of the 2009 2nd International Congress on Image and Signal Processing, Tianjin, China, 17–19 October 2009; pp. 1–5.

32. Sampat, M.P.; Wang, Z.; Gupta, S.; Bovik, A.C.; Markey, M.K. Complex wavelet structural similarity: A new image similarity index. *IEEE Trans. Image Process.* **2009**, *18*, 2385–2401. [CrossRef] [PubMed]

33. Wang, Z.; Li, Q. Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Process.* **2010**, *20*, 1185–1198. [CrossRef]

34. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2, pp. 1398–1402.

35. Li, C.; Bovik, A.C. Three-component weighted structural similarity index. In *Proceedings of the Image Quality and System Performance VI*; International Society for Optics and Photonics: San Jose, CA, USA, 2009; Volume 7242, p. 72420Q.

36. Kolaman, A.; Yadid-Pecht, O. Quaternion structural similarity: A new quality index for color images. *IEEE Trans. Image Process.* **2011**, *21*, 1526–1536. [CrossRef]
37. Zhu, R.; Zhou, F.; Xue, J.H. MvSSIM: A quality assessment index for hyperspectral images. *Neurocomputing* **2018**, *272*, 250–257. [CrossRef]
38. Sheikh, H.R.; Bovik, A.C.; De Veciana, G. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* **2005**, *14*, 2117–2128. [CrossRef]
39. Sheikh, H.R.; Bovik, A.C. Image information and visual quality. *IEEE Trans. Image Process.* **2006**, *15*, 430–444. [CrossRef]
40. Liang, Y.; Wang, J.; Wan, X.; Gong, Y.; Zheng, N. Image quality assessment using similar scene as reference. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 3–18.
41. Kim, J.; Lee, S. Deep learning of human visual sensitivity in image quality assessment framework. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1676–1684.
42. Ahn, S.; Choi, Y.; Yoon, K. Deep learning-based distortion sensitivity prediction for full-reference image quality assessment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 344–353.
43. Ding, Y.; Zhao, Y.; Zhao, X. Image quality assessment based on multi-feature extraction and synthesis with support vector regression. *Signal Process. Image Commun.* **2017**, *54*, 81–92. [CrossRef]
44. Tang, Z.; Zheng, Y.; Gu, K.; Liao, K.; Wang, W.; Yu, M. Full-reference image quality assessment by combining features in spatial and frequency domains. *IEEE Trans. Broadcast.* **2018**, *65*, 138–151. [CrossRef]
45. Kovesi, P. Image features from phase congruency. *Videre J. Comput. Vis. Res.* **1999**, *1*, 1–26.
46. Okarma, K. Combined full-reference image quality metric linearly correlated with subjective assessment. In *Proceedings of the International Conference on Artificial Intelligence and Soft Computing*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 539–546.
47. Mansouri, A.; Aznaveh, A.M.; Torkamani-Azar, F.; Jahanshahi, J.A. Image quality assessment using the singular value decomposition theorem. *Opt. Rev.* **2009**, *16*, 49–53. [CrossRef]
48. Oszust, M. Image quality assessment with lasso regression and pairwise score differences. *Multimed. Tools Appl.* **2017**, *76*, 13255–13270. [CrossRef]
49. Yuan, Y.; Guo, Q.; Lu, X. Image quality assessment: A sparse learning way. *Neurocomputing* **2015**, *159*, 227–241. [CrossRef]
50. Lukin, V.V.; Ponomarenko, N.N.; Ieremeiev, O.I.; Egiazarian, K.O.; Astola, J. Combining full-reference image visual quality metrics by neural network. In Proceedings of the Human Vision and Electronic Imaging XX, San Francisco, CA, USA, 17 March 2015; Volume 9394, pp. 172–183.
51. Amirshahi, S.A.; Pedersen, M.; Beghdadi, A. Reviving traditional image quality metrics using CNNs. In Proceedings of the Color and Imaging Conference, Albuquerque, NM, USA, 4–8 November 2018; Volume 2018, pp. 241–246.
52. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
53. Bakurov, I.; Buzzelli, M.; Schettini, R.; Castelli, M.; Vanneschi, L. Structural similarity index (SSIM) revisited: A data-driven approach. *Expert Syst. Appl.* **2022**, *189*, 116087. [CrossRef]
54. Okarma, K. Combined visual quality metric of remote sensing images based on neural network. In *Radioelectronic and Computer Systems*; National Aerospace University: Kharkiv, Ukraine, 2020; pp. 4–15.
55. Okarma, K.; Chlewicki, W.; Kopytek, M.; Marciniak, B.; Lukin, V. Entropy-Based Combined Metric for Automatic Objective Quality Assessment of Stitched Panoramic Images. *Entropy* **2021**, *23*, 1525. [CrossRef]
56. Xu, L.; Lin, W.; Kuo, C.C.J. *Visual Quality Assessment by Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2015.
57. Pedersen, M.; Hardeberg, J.Y. Full-reference image quality metrics: Classification and evaluation. *Found. Trends® Comput. Graph. Vis.* **2012**, *7*, 1–80.
58. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. A comprehensive evaluation of full reference image quality assessment algorithms. In Proceedings of the 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; pp. 1477–1480.
59. Zhai, G.; Min, X. Perceptual image quality assessment: A survey. *Sci. China Inf. Sci.* **2020**, *63*, 1–52. [CrossRef]
60. Min, X.; Ma, K.; Gu, K.; Zhai, G.; Wang, Z.; Lin, W. Unified blind quality assessment of compressed natural, graphic, and screen content images. *IEEE Trans. Image Process.* **2017**, *26*, 5462–5474. [CrossRef]
61. Min, X.; Gu, K.; Zhai, G.; Yang, X.; Zhang, W.; Le Callet, P.; Chen, C.W. Screen content quality assessment: Overview, benchmark, and beyond. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–36. [CrossRef]
62. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [CrossRef] [PubMed]
63. Liu, A.; Lin, W.; Narwaria, M. Image quality assessment based on gradient similarity. *IEEE Trans. Image Process.* **2011**, *21*, 1500–1512. [PubMed]
64. Chang, H.W.; Zhang, Q.W.; Wu, Q.G.; Gan, Y. Perceptual image quality assessment by independent feature detector. *Neurocomputing* **2015**, *151*, 1142–1152. [CrossRef]
65. Damera-Venkata, N.; Kite, T.D.; Geisler, W.S.; Evans, B.L.; Bovik, A.C. Image quality assessment based on a degradation model. *IEEE Trans. Image Process.* **2000**, *9*, 636–650. [CrossRef] [PubMed]

66. Zhang, L.; Zhang, L.; Mou, X. RFSIM: A feature based image quality assessment metric using Riesz transforms. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 321–324.

67. Chang, H.W.; Yang, H.; Gan, Y.; Wang, M.H. Sparse feature fidelity for perceptual image quality assessment. *IEEE Trans. Image Process.* **2013**, *22*, 4007–4018. [CrossRef]

68. Zhang, L.; Shen, Y.; Li, H. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Trans. Image Process.* **2014**, *23*, 4270–4281. [CrossRef]

69. Shi, C.; Lin, Y. Full reference image quality assessment based on visual salience with color appearance and gradient similarity. *IEEE Access* **2020**, *8*, 97310–97320. [CrossRef]

70. Shi, C.; Lin, Y. Image Quality Assessment Based on Three Features Fusion in Three Fusion Steps. *Symmetry* **2022**, *14*, 773. [CrossRef]

71. Wu, J.; Lin, W.; Shi, G. Image quality assessment with degradation on spatial structure. *IEEE Signal Process. Lett.* **2014**, *21*, 437–440. [CrossRef]

72. Van Laarhoven, P.J.; Aarts, E.H. Simulated annealing. In *Simulated Annealing: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 1987; pp. 7–15.

73. Kirkpatrick, S.; Gelatt Jr, C.D.; Vecchi, M.P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680. [CrossRef] [PubMed]

74. Yu, X.; Bampis, C.G.; Gupta, P.; Bovik, A.C. Predicting the quality of images compressed after distortion in two steps. *IEEE Trans. Image Process.* **2019**, *28*, 5757–5770. [CrossRef]

75. Temel, D.; AlRegib, G. CSV: Image quality assessment based on color, structure, and visual system. *Signal Process. Image Commun.* **2016**, *48*, 92–103. [CrossRef]

76. Ding, K.; Ma, K.; Wang, S.; Simoncelli, E.P. Image quality assessment: Unifying structure and texture similarity. *arXiv* **2020**, arXiv:2004.07728.

77. Zhang, X.; Feng, X.; Wang, W.; Xue, W. Edge strength similarity for image quality assessment. *IEEE Signal Process. Lett.* **2013**, *20*, 319–322. [CrossRef]

78. Temel, D.; AlRegib, G. ReSIFT: Reliability-weighted sift-based image quality assessment. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 2047–2051.

79. Yang, G.; Li, D.; Lu, F.; Liao, Y.; Yang, W. RVSIM: A feature similarity method for full-reference image quality assessment. *EURASIP J. Image Video Process.* **2018**, *2018*, 1–15. [CrossRef]

80. Temel, D.; AlRegib, G. Perceptual image quality assessment through spectral analysis of error representations. *Signal Process. Image Commun.* **2019**, *70*, 37–46. [CrossRef]