Article

An Optimization-Based Family of Predictive, Fusion-Based Models for Full-Reference Image Quality Assessment

Domonkos Varga

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

Abstract: Given the reference (distortion-free) image, full-reference image quality assessment (FR-IQA) algorithms seek to assess the perceptual quality of the test image. Over the years, many effective, hand-crafted FR-IQA metrics have been proposed in the literature. In this work, we present a novel framework for FR-IQA that combines multiple metrics and tries to leverage the strength of each by formulating FR-IQA as an optimization problem. Following the idea of other fusion-based metrics, the perceptual quality of a test image is defined as the weighted product of several already existing, hand-crafted FR-IQA metrics. Unlike other methods, the weights are determined in an optimization-based framework and the objective function is defined to maximize the correlation and minimize the root mean square error between the predicted and ground-truth quality scores. The obtained metrics are evaluated on four popular benchmark IQA databases and compared to the state of the art. This comparison has revealed that the compiled fusion-based metrics are able to outperform other competing algorithms, including deep learning-based ones.

Keywords: full-reference image quality assessment; optimization; quality-aware features

1. Introduction

With social media and streaming applications booming, it is required from systems, that are able to quickly transmit a large number of images, to provide the best available user experience [1]. However, various distortions are added to digital images during storage, compression, and transmission. Therefore, the continuous evaluation and monitoring of image quality is of great importance to content providers [2]. As a consequence, objective image quality assessment (IQA) has become a very hot research topic [3], because it tries to devise mathematical models that are able to give perceptual quality estimation consistent with human judgment. The literature usually divides objective IQA into three branches [4,5] based on the availability or unavailability of the reference (distortion-free) images in the quality evaluation process. As the terminology suggests, full-reference (FR) IQA evaluates the quality of distorted images with full access to their reference counterparts, while no-reference (NR) IQA has no access and reduced-reference (RR) IQA has partial access to them.

Because the underlying model of the human visual system (HVS) is extremely complex and its many elements are not fully understood [6], researchers have proposed many FR-IQA algorithms, which take into consideration different aspects of the HVS, over the years. Recently, there have been numerous attempts to increase the performance of FR-IQA by combining several already existing FR-IQA metrics to compile a "super" evaluator. First, Okarma [7] introduced such a fusion-based metric by applying the product and power of MS-SSIM [8], VIF [9], and R-SVD [10]. Later, this idea was developed further into several directions. A line of works utilized optimization or regression techniques to determine optimal weights or exponents in summations or products of already existing FR-IQA metrics. For instance, Oszust [11] determined the optimal weights using a genetic algorithm with a root mean square error (RMSE) objective function which was calculated between predicted and ground-truth scores. Bakurov et al. [12] chose a similar solution, but the

authors revisited the SSIM [13] and MS-SSIM [8] metrics to find optimal parameters in their formulas, using evolutionary and swarm intelligence methods instead of the originally proposed grid search. On the other hand, Okarma [14] used the MATLAB fminsearch function to determine the optimal exponents in a weighted product of traditional FR-IQA metrics. Another line of works utilize machine learning techniques to create fusion-based methods. The illustrative example is the paper of Lukin et al. [15] where the results of traditional FR-IQA metrics were used as a feature vector to train a shallow neural network. Amirshahi et al. [16] compiled a fusion-based metric by comparing the activation maps produced through reference and distorted images of an AlexNet [17] convolutional neural network using traditional image quality metrics.

1.1. Contributions

In this paper, we make the following contributions. We propose a novel framework for FR-IQA that combines multiple metrics and tries to leverage the strength of each by formulating FR-IQA as an optimization problem. Following the idea of other fusion-based metrics, the perceptual quality of a test image is defined as the weighted product of several already existing, hand-crafted FR-IQA metrics. Unlike other methods [7,18,19], the weights in the product are determined in a novel optimization-based framework and the objective function is defined to maximize the correlation strength and minimize the root mean square error between the predicted and ground-truth quality scores.

1.2. Structure of the Paper

To provide a clear and organized presentation of our work, this paper is structured as follows. In Section 2, we discuss the main approaches of FR-IQA and summarize significant methods of the field. Section 3 outlines our proposed method. In Section 4, we present the experimental results and analyze the performance of our method by comparing it to other state-of-the-art methods. We conclude this paper in Section 5 and discuss potential applications and future research directions.

2. Related Work

Taking the mean square error between reference and distorted images is a simple and straightforward FR-IQA metric. However, the provided quality scores do not correlate well with human judgment [20]. Similarly, PSNR [21] is also simple and straightforward but fails to give satisfactory results [22]. Other metrics take the sensitivity of the HVS to structural degradation into consideration, such as the structural similarity index (SSIM) [13]. On the basis of SSIM [13], a large number of FR-IQA metrics has been proposed over the years, such as MS-SSIM [8], CW-SSIM [23], ESSIM [24], GSSIM [25], IW-SSIM [26], and 3-SSIM [27]. In SSIM [13], a comparison between the distorted and reference (distortion-free) images is performed on the basis of three features, i.e., luminance, contrast, and structure. To be more specific, the SSIM between two images (denoted here by **A** and **B**) in an image patch around (x, y) coordinates is given as

$$SSIM(x, y) = (l(x, y))^{\alpha} \times (c(x, y))^{\beta} \times (s(x, y))^{\gamma},$$ (1)

where the luminance component is defined as

$$l(x, y) = \frac{2\mu_{\mathbf{A}}(x, y)\mu_{\mathbf{B}}(x, y) + C_1}{(\mu_{\mathbf{A}}(x, y))^2 + (\mu_{\mathbf{B}}(x, y))^2 + C_1},$$ (2)

the contrast component is given as

$$c(x, y) = \frac{2\sigma_{\mathbf{A}}(x, y)\sigma_{\mathbf{B}}(x, y) + C_2}{(\sigma_{\mathbf{A}}(x, y))^2 + (\sigma_{\mathbf{B}}(x, y))^2 + C_2},$$ (3)

and the structure component is determined as

$$s(x,y) = \frac{\sigma_{\mathbf{AB}}(x,y) + C_3}{\sigma_{\mathbf{A}}(x,y)\sigma_{\mathbf{B}}(x,y) + C_3}. \tag{4}$$

In Equations (1)–(4), $\mu_{\mathbf{A}}(x,y)$ and $\mu_{\mathbf{B}}(x,y)$ denote the average of the pixel values in the image patch around (x,y) in images **A** and **B**, respectively. Similarly, $(\sigma_{\mathbf{A}}(x,y))^2$ and $(\sigma_{\mathbf{B}}(x,y))^2$ stand for the variances. Further, $\sigma_{\mathbf{AB}}(x,y)$ is the covariance calculated between the two images from **A** and **B**. The constants are calculated as $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$, and $C_3 = C_2/2$. Further, L stands for the dynamic range of the pixel values, and for images with 8-bit depth, $L = 255$ is the recommended value. By default, $K_1 = 0.01$ and $K_2 = 0.03$ are also constants whose exact values were chosen by Wang et al. [13] after an ablation study. To give the perceptual quality of the distorted image in possession of the reference image, the arithmetic average of $SSIM(x,y)$ is taken. As already mentioned, a huge number of FR-IQA metrics has been inspired by the original SSIM. For comprehensive overviews on SSIM-motivated methods, the following papers can be recommended [12,28–31]. Here, several representative methods are mentioned in the following. The authors of multi-scale SSIM [8] extended the idea of SSIM into multiple scales. Sampat et al. [23] replaced the components of SSIM by complex wavelet coefficients [32]. In contrast, Zhang et al. [24] defined an edge strength-based image quality metric where the strength of edges was defined in horizontal and diagonal directions using directional derivatives. Chen et al. [25] took a similar approach, but the edge information was characterized by gradient magnitudes. In [26], the authors used the information content measure as a weighting factor in the pooling process of SSIM [13] to obtain improved prediction results. This idea was further improved by Larson et al. [33] where low-level distortions, which are nearly imperceptible, were modeled by local luminance and contrast masking, while high-level distortions were modeled using spatial-frequency components. Kolaman and Yadid-Pecht [34] extended the SSIM metric to colorful images by modeling colors with quaternions. In [35], the authors analyzed different strategies aiming at the usage of visual saliency maps [36] in improving IQA algorithms. A proposal was the weighting of local estimates by local saliency values. In [37], first- and second-order Riesz-transform [38] coefficients were used to create feature maps for the reference and the distorted images which were compared to give an estimation of the perceptual image quality. Similarly, Zhang et al. [39] compared feature maps to quantify image quality, but the authors used phase congruency [40] and gradient magnitude maps.

Recently, the scientific community has paid more and more attention to the deployment of machine and deep learning models in almost all computer vision tasks [41]. The field of image quality assessment has accommodated this trend [3,42]. For instance, Tang et al. [43] extracted spatial and frequency domain features from reference–distorted image pairs and trained a random forest regressor for image quality prediction. In contrast, Bosse et al. [44] devised a convolutional neural network (CNN) architecture which can be trained end-to-end on single images or on image pairs. Similarly, Zhang et al. [45] trained an end-to-end CNN in a patch-wise fashion for FR-IQA and compared the effectiveness of deep features extracted from different pretrained CNNs. As a consequence, it can be used for both NR- and FR-IQA. In [46], the authors proposed a pairwise-learning framework for FR-IQA. Several works extracted deep features via pretrained CNNs from reference–distorted image pairs and compared them to assess the perceptual image quality. For instance, Amirshahi et al. [47] compared the histograms of deep features using a histogram intersection kernel (HIK) [48] at multiple levels. The perceptual quality was obtained by aggregating the similarity scores provided by the HIKs. Later, this approach was further developed in [16] by replacing the HIK in comparing convolutional feature maps with a traditional image similarity metric. In [49], the authors used the error map calculated between the reference and distorted images and the subjective saliencies of the distorted images to train a CNN for perceptual image quality estimation.

Recently, several researchers devised fusion-based FR-IQA methods where the goal is creating a "super-evaluator" using already known FR-IQA metrics to achieve advanced performance. A large number of fusion-based algorithms determine weights for each FR-IQA metric in a summation or in a product of sequence [1]. An illustrative example is the method proposed by Okarma [7]. Namely, the properties of three different FR-IQA metrics were examined thoroughly, and a combined metric was devised based on the metrics' arithmetical product and power. By using mathematical optimization techniques, the parameter values of this fusion-based metric were refined in [14]. Oszust [50] and Yuan et al. [51] also developed this approach further by applying lasso regression and kernel ridge regression, respectively. Oszust [11] determined the weights in a linear combination of traditional FR-IQA metrics by applying a genetic algorithm. In [52], this approach was further developed by using multi-gene genetic programming [53]. The effectiveness of this approach was also demonstrated on screen content images [54]. Simulated annealing was also applied in this framework [55]. Machine learning techniques were also used in creating fusion-based algorithms. An illustrative example is Lukin et al.'s [15] work. Namely, the authors used the outcomes of several FR-IQA metrics as features and trained a neural network on top of them to predict perceptual quality. A similar approach using a neural network was proposed for the quality assessment of remote sensing images [56].

In summary, this section has highlighted the various approaches that have been proposed in the literature for FR-IQA. Although the reviewed studies have contributed significantly to the field, a detailed overview about IQA or FR-IQA is out of the scope of this study. For a general overview about the field of IQA, the PhD dissertations of Jenadeleh [57] and Men [58] can be recommended while Min-juan et al. [59], Phadikar et al. [60], George et al. [61], and Pedersen et al. [30] provide state-of-the-art studies on FR-IQA.

3. Proposed Method

In [7], Okarma took into account the different properties of three different FR-IQA metrics (MS-SSIM [8], VIF [62], and R-SVD [10]) and defined a combined quality metric (CQM):

$$CQM = (MSSSIM)^a \times (VIF)^b \times (RSVD)^c,$$ (5)

where $a = 7$, $b = 0.3$, and $c = -0.15$ values were used because they lead to a near-optimal solution on an IQA benchmark database. Following this basic idea of Okarma [7], a fusion-based metric is defined as the weighted product of n different traditional FR-IQA methods' results:

$$Q_p = \prod_{i=1}^{n} q_i^{\alpha_i},$$ (6)

where q_is are the results of the applied FR-IQA metrics and α_is are the associated weights. Specifically, we chose $n = 18$ and the following metrics were utilized: FSIM [39], FSIMc [39], GSM [63], IFC [9], IW-SSIM [26], MAD [33], MS-SSIM [8], NQM [64], PSNR [21], RFSIM [37], SFF [65], SSIM [13], SR-SIM [66], UQI [67], VIF [62], VSI [68], and VSNR [69]. A summary of the acronyms of the used FR-IQA metrics can be found in Table 1. In the literature, the parameters of an FR-IQA metric are tuned on a smaller subset of images. In the case of a traditional metric, such as SSIM [13] given by Equations (1)–(4), the number of tunable parameters is one or two. As a consequence, appropriate values can be easily found applying for cycles over a search space. In contrast, our fusion-based metric given by Equation (6) contains $n = 18$ parameters and an optimization task is defined to find their exact values. To determine the optimal weights (parameters) in Equation (6), the following optimization problem is defined:

$$\max_{\alpha} \frac{SROCC(\mathbf{Q}_p, \mathbf{S}) + KROCC(\mathbf{Q}_p, \mathbf{S})}{RMSE(F(\mathbf{Q}_p, \boldsymbol{\beta}), \mathbf{S})},$$ (7)

$$\text{subject to } \alpha_i \in \mathbb{R}, n \in \mathbb{N}, \boldsymbol{\beta} \geq 0,$$

where $\mathbf{Q}_p$ and $\mathbf{S}$ are vectors containing the predicted and ground-truth quality scores, respectively. $SROCC(\cdot, \cdot)$ and $KROCC(\cdot, \cdot)$ denote the Spearman's rank-order correlation coefficient and Kendall's rank-order correlation coefficient calculated between two vectors, respectively. Further, $RMSE(\cdot, \cdot)$ is the root mean square error determined between two vectors. Prior to the calculation of the RMSE, a non-linear mapping is applied to the predicted scores following the recommendations of [4]. In this paper, the following non-linear function was applied

$$F(Q_p, \boldsymbol{\beta}) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2(Q_p - \beta_3)}} \right) + \beta_4 Q_p + \beta_5, \tag{8}$$

with the following β parameters, $\beta_1 = 10$, $\beta_2 = 0$, $\beta_3 = mean(Q_p)$, $\beta_4 = 1$, and $\beta_4 = 0.1$, which were also used in the MATLAB implementation of the VSI [68] method.

Because an FR-IQA metric is supposed to provide objective scores which have a high correlation and low RMSE with respect to subjective quality scores collected from human observers, the objective function's—given by Equation (7)—numerator consists of the sum of SROCC and KROCC while the denominator corresponds to the RMSE. Our preliminary investigations revealed that considering only SROCC or KROCC may result in a higher RMSE than those of the state of the art. That is why we decided to divide the sum of SROCC and KROCC by the RMSE.

Table 1. Acronyms of the used FR-IQA metrics applied in our fusion-based method.

Method's Acronym	Full Name
FSIM [39]	feature similarity index
FSIMc [39]	feature similarity index extension
GSM [63]	gradient similarity measure
IFC [9]	information fidelity criterion
IW-SSIM [26]	information content weighted SSIM
MAD [33]	most apparent distortion
MS-SSIM [8]	multi-scale SSIM
NQM [64]	noise quality measure
PSNR [21]	peak signal-to-noise ratio
RFSIM [37]	Riesz-transform-based feature similarity metric
SFF [65]	sparse feature fidelity
SSIM [13]	structural similarity index measure
SR-SIM [66]	spectral residual-based similarity
UQI [67]	universal image quality index
VIF [62]	visual information fidelity
VSI [68]	visual saliency-induced index
VSNR [69]	visual signal-to-noise ratio

Two nature-inspired optimization methods, such as the genetic algorithm [70] (GA) and pattern search [71] (PS), were applied to the problem defined by Equation (7) to determine the optimal weights. Further, the simplex method of Lagarias et al. [72], which is implemented in the *fminsearch* function of MATLAB's Optimization Toolbox, was also used. To improve the efficiency of the fusion, each method was able to execute model selection (which FR-IQA metric to aggregate or not). The main motivation behind the choice of optimization methods was to collect algorithms that are able to give at least approximate solutions for NP (non-deterministic polynomial-time) hard problems. Figures 1 and 2 depict the compilation of the proposed fusion-based FR-IQA metric. Specifically, the fusion was carried out on 20% of the reference images and their corresponding distorted counterparts in our method. In the literature, 20% is a common choice for parameter setting in a derived formula [73,74], but there are also researchers who used 30% [39] or 80% [75]. In total, four fusion strategies were realized with the help of one

optimization method and benchmark database. Further, the fusion strategies were also cross-database tested. Each optimization method was carried out 100 times and the best solution was finally selected. We codenamed our method OFIQA to refer to the fact that the decision fusion was carried out via optimization.
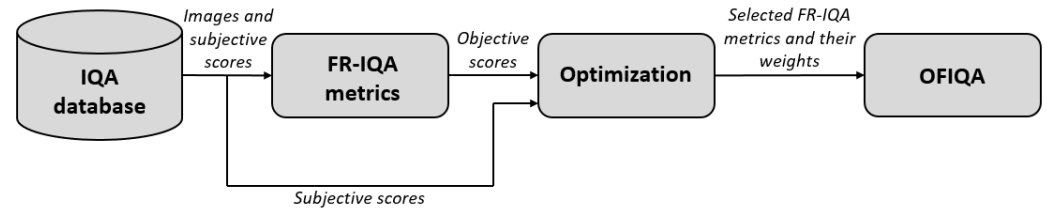
Figure 1. Twenty percent of the reference images with their corresponding distorted counterparts are selected to determine the parameters of the proposed fusion-based metric in the optimization process. The resulting metric is codenamed as OFIQA.
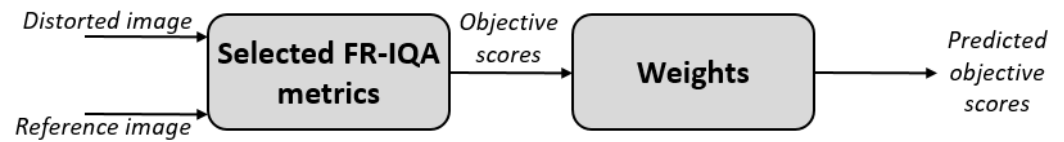
Figure 2. The weighted product of the selected FR-IQA metrics is used to estimate the perceptual quality of a distorted image in the evaluation stage.

In the GA, the population size and the number of generations were set to 100. The best solutions on the four benchmark databases were provided by the following equations:

$$OFIQA_{LIVE}^{GA} = FSIM^{-3.6888} \times GSM^{12.5693} \times IWSSIM^{0.9556} \times IFS^{-1.8159}, \qquad (9)$$

$$OFIQA_{TID2013}^{GA} = VSI^{13.9336} \times FSIMc^{2.2946} \times GSM^{-10.864} \times NQM^{-0.1713} \times \\ \times SRSIM^{2.4651} \times IFS^{0.5139}, \qquad (10)$$

$$OFIQA_{TID2008}^{GA} = VSI^{7.0221} \times FSIM^{0.259} \times FSIMc^{1.0055} \times GSM^{-19.8267} \times \\ \times PSNR^{0.1471} \times VIF^{0.1452} \times SFF^{2.4029}, \qquad (11)$$

$$OFIQA_{CSIQ}^{GA} = FSIMc^{-2.7532} \times MAD^{0.9692} \times MSSSIM^{1.1892} \times SSIM^{1.6561} \times VIF^{-0.75} \times \\ \times IFS^{-3.4013} \times SFF^{2.2901}. \qquad (12)$$

In the case of the PS after 100 runs, the following fusion-based metrics can be obtained:

$$OFIQA_{LIVE}^{PS} = FSIM^{0.6964} \times FSIMc^{-2.6056} \times MAD^{1.0817} \times MSSSIM^{-0.4711} \times \\ \times SSIM^{0.7302} \times UQI^{0.9946}, \qquad (13)$$

$$OFIQA_{TID2013}^{PS} = VSI^{24.1037} \times FSIM^{0.9292} \times GSM^{-19.5555} \times IWSSIM^{2.1053} \times \\ \times MSSSIM^{-6.1562} \times PSNR^{0.4649} \times VIF^{0.5463} \times SFF^{4.3998}, \qquad (14)$$

$$OFIQA_{TID2008}^{PS} = VSI^{23.5097} \times FSIM^{1.2155} \times FSIMc^{0.2494} \times GSM^{-21.8595} \times$$
$$\times IWSSIM^{1.2984} \times MSSSIM^{-1.9792} \times PSNR^{0.5571} \times SSIM^{-1.8374} \times$$
$$\times VIF^{0.5491}, \tag{15}$$

$$OFIQA_{CSIQ}^{PS} = FSIM^{0.3471} \times FSIMc^{0.7575} \times GSM^{-60.4948} \times MAD^{1.8509} \times$$
$$\times NQM^{0.8049} \times PSNR^{0.8181} \times UQI^{0.5142} \times VIF^{-0.1294}. \tag{16}$$

Using the method of Lagarias et al. [72], the following fusion metrics can be obtained:

$$OFIQA_{LIVE}^{fmin} = VSI^{0.442} \times FSIMc^{1.1986} \times GSM^{1.0479} \times IFC^{-0.1531} \times$$
$$\times IWSSIM^{1.8895} \times MAD^{1.6539} \times MSSSIM^{-0.8459} \times NQM^{0.4463} \times$$
$$\times PSNR^{-0.2781} \times RFSIM^{-0.1719} \times SRSIM^{-1.1448} \times SSIM^{0.0811} \times$$
$$\times UQI^{0.0955} \times VIF^{-0.6669} \times VSNR^{0.0765} \times SFF^{-0.1433}. \tag{17}$$

$$OFIQA_{TID2013}^{fmin} = VSI^{0.6577} \times FSIM^{0.1742} \times FSIMc^{0.7197} \times GSM^{0.6092} \times$$
$$\times IFC^{0.3768} \times IWSSIM^{0.8904} \times MAD^{0.4482} \times MSSSIM^{0.9852} \times$$
$$\times NQM^{0.6304} \times PSNR^{0.8077} \times RFSIM^{0.3861} \times SSIM^{0.7985} \times$$
$$\times UQI^{0.3612} \times VIF^{0.0149} \times VSNR^{0.3338} \times IFS^{0.8979} \times$$
$$\times SFF^{0.9113}, \tag{18}$$

$$OFIQA_{TID2008}^{fmin} = VSI^{2.0607} \times FSIM^{0.4211} \times FSIMc^{0.8323} \times GSM^{0.1672} \times$$
$$\times IFC^{0.0322} \times MAD^{0.0753} \times NQM^{0.098} \times PSNR^{0.3727} \times$$
$$\times RFSIM^{0.5523} \times SRSIM^{0.5783} \times SSIM^{-0.2377} \times UQI^{-0.3083} \times$$
$$\times VIF^{0.5273} \times VSNR^{-0.0292} \times IFS^{1.9221} \times SFF^{0.0902}, \tag{19}$$

$$OFIQA_{CSIQ}^{fmin} = VSI^{-0.4328} \times FSIM^{0.4287} \times GSM^{0.0521} \times IFC^{0.0569} \times$$
$$\times IWSSIM^{-1.4847} \times MAD^{0.982} \times MSSSIM^{0.3763} \times NQM^{0.3919} \times$$
$$\times PSNR^{0.2141} \times RFSIM^{-0.1544} \times SRSIM^{0.7473} \times UQI^{0.2373} \times$$
$$\times VIF^{-0.2373} \times VIF^{-0.2243} \times VSNR^{0.1147} \times IFS^{-1.3102} \times SFF^{1.7745}. \tag{20}$$

4. Results

In this section, the experimental numerical results are presented. First, the used benchmark IQA databases are introduced in Section 4.1. Next, Section 4.2 defines the applied evaluation metrics. A parameter study with respect to the applied optimization methods is presented in Section 4.3. Finally, the results of a comparison to the state of the art is given in Section 4.4.

4.1. Databases

For evaluation, four IQA benchmark databases are used, i.e., LIVE (Laboratory for Image and Video Engineering) [4], TID2013 (Tampere Image Database) [76], TID2008 [77], and CSIQ (Categorical Image Quality) [33], which contain a small set of reference images (whose perceptual qualities are considered perfect) and a large set of quality annotated distorted images generated from the reference images using different distortion types at

different distortion levels. The main characteristics of the applied databases are given in Table 2.

Table 2. Applied benchmark IQA databases.

	LIVE [4]	**TID2013 [76]**	**TID2008 [77]**	**CSIQ [33]**
Image resolution	$\sim 768 \times 512$	512×384	512×384	500×500
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

4.2. Evaluation Metrics

In this study, four different performance indices, i.e., root mean square error (RMSE), Pearson's linear correlation coefficient (PLCC), Spearman's rank-order correlation coefficient (SROCC), and Kendall's rank-order correlation coefficient (KROCC), are applied to characterize the performance of the proposed fusion-based metric and other considered state-of-the-art methods in an ablation study and a comparison to the state of the art. The RMSE and PLCC are calculated after a non-linear mapping of the vector of predicted scores. This mapping has already been given by Equation (8). The RMSE is given as

$$RMSE(\mathbf{Q}_p, \mathbf{S}) = \sqrt{\frac{(\mathbf{Q}_p - \mathbf{S})^T (\mathbf{Q}_p - \mathbf{S})}{m}} \tag{21}$$

where $\mathbf{Q}_p$ is the vector of predicted scores after the non-linear mapping, $\mathbf{S}$ is the vector of ground-truth scores, and m denotes the number of samples (in this case images). PLCC is given as

$$PLCC(\mathbf{Q}_p, \mathbf{S}) = \frac{\bar{\mathbf{Q}}_p^T \bar{\mathbf{S}}}{\sqrt{\bar{\mathbf{Q}}_p^T \bar{\mathbf{Q}}_p \bar{\mathbf{S}}^T \bar{\mathbf{S}}}} \tag{22}$$

where $\bar{\mathbf{Q}}_p$ and $\bar{\mathbf{S}}$ are mean-removed vectors. SROCC is defined as

$$SROCC(\mathbf{Q}, \mathbf{S}) = 1 - \frac{6 \times \sum_{i=1}^{m} d_i^2}{m(m^2 - 1)}, \tag{23}$$

where d_i stands for the difference between $\mathbf{Q}$ and $\mathbf{S}$ at the ith entry. KROCC is defined as

$$KROCC(\mathbf{Q}, \mathbf{S}) = \frac{m_c - m_d}{\frac{1}{2} m(m - 1)}, \tag{24}$$

where m_c and m_d are the number of concordant and discordant pairs in the database, respectively.

In Table 3, the details of the computer configuration applied in our experiments are given.

Table 3. 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

4.3. Ablation Study

In this subsection, an ablation study is carried out with respect to the applied optimization method. As already mentioned, a GA [70], a PS [71], and the simplex method of Lagarias et al. [72] (implemented in MATLAB's fminsearch) were considered. Further, the obtained metrics have already been given by Equations (9)–(20). The results in terms of the RMSE and SROCC are summarized in Tables 4 and 5. From these numerical results, it can be clearly seen that the weights obtained by the GA are significantly better than those of the other two methods. Namely, they provide a consistently good performance in terms of the RMSE and SROCC on all the databases, while the other weights fail to provide a good performance in several cases and sometimes give an unacceptably high RMSE or a low SROCC (the underlined values in Table 4).

Table 4. RMSE performance comparison with respect to the applied optimization method applied in the proposed fusion-based FR-IQA metrics. Results are given for LIVE [4], TID2013 [76], TID2008 [77], and CSIQ [33]. The best results for each database are typed in red, the second best results are in green, the third best results are in blue, and the worst results are underlined.

FR-IQA Metric	LIVE [4]	TID2013 [76]	TID2008 [77]	CSIQ [33]
$OFIQA^{GA}_{LIVE}$	7.895	0.593	0.641	0.109
$OFIQA^{GA}_{TID2013}$	8.062	0.526	0.606	0.091
$OFIQA^{GA}_{TID2008}$	7.078	0.570	0.557	0.068
$OFIQA^{GA}_{CSIQ}$	6.918	0.665	0.645	0.067
$OFIQA^{PS}_{LIVE}$	6.860	0.675	1.116	0.165
$OFIQA^{PS}_{TID2013}$	7.503	1.150	0.537	0.069
$OFIQA^{PS}_{TID2008}$	7.662	0.766	0.544	0.081
$OFIQA^{PS}_{CSIQ}$	7.882	0.658	0.672	0.124
$OFIQA^{fmin}_{LIVE}$	6.612	0.642	0.648	0.060
$OFIQA^{fmin}_{TID2013}$	16.195	0.856	0.834	0.833
$OFIQA^{fmin}_{TID2008}$	14.456	0.833	0.631	0.630
$OFIQA^{fmin}_{CSIQ}$	7.167	0.812	0.680	0.104

Table 5. SROCC performance comparison with respect to the applied optimization method applied in the proposed fusion-based FR-IQA metrics. Results are given for LIVE [4], TID2013 [76], TID2008 [77], and CSIQ [33]. The best results for each database are typed in red, the second best results are in green, the third best results are in blue, and the worst results are underlined.

FR-IQA Metric	LIVE [4]	TID2013 [76]	TID2008 [77]	CSIQ [33]
$OFIQA^{GA}_{LIVE}$	0.961	0.863	0.888	0.938
$OFIQA^{GA}_{TID2013}$	0.957	0.890	0.904	0.923
$OFIQA^{GA}_{TID2008}$	0.967	0.825	0.911	0.964
$OFIQA^{GA}_{CSIQ}$	0.972	0.808	0.882	0.965
$OFIQA^{PS}_{LIVE}$	0.968	0.790	0.585	0.807
$OFIQA^{PS}_{TID2013}$	0.965	0.826	0.914	0.962
$OFIQA^{PS}_{TID2008}$	0.963	0.822	0.915	0.944
$OFIQA^{PS}_{CSIQ}$	0.965	0.794	0.861	0.964
$OFIQA^{fmin}_{LIVE}$	0.972	0.864	0.867	0.969
$OFIQA^{fmin}_{TID2013}$	0.781	0.845	0.769	0.496
$OFIQA^{fmin}_{TID2008}$	0.834	0.717	0.888	0.586
$OFIQA^{fmin}_{CSIQ}$	0.974	0.823	0.846	0.955

4.4. Comparison to the State of the Art

In this subsection, the proposed FR-IQA metrics are compared to a set of state-of-the-art methods (2stepQA [78], CSV [79], DISTS [80], ESSIM [24], FSIM [39], FSIMc [39], GSM [63], IFC [9], IFS [81], IW-SSIM [26], MAD [33], MS-SSIM [8], NQM [64], PSNR [21], Re-SIFT [82], RFSIM [37], RVSIM [83], SFF [65], SR-SIM [66], SSIM [13], SUMMER [84], VIF [62], and VSI [68]), whose MATLAB source codes were made available by researchers. Further, we reimplemented the SSIM-CNN method proposed by Amirshahi et al. [16] and made it available for the research community (https://github.com/Skythianos/SSIM-CNN, accessed on 1 January 2023). In addition to this, the results of the recently published GP-SSIM [85] and the results of deep learning-based DeepSIM [86], DeepIQA [44], PieAPP [46], and LPIPS [45] are also included in our comparison based on Bakurov et al.'s [85] study. Our proposed fusion-based FR-IQA metrics have also been exactly given by Equations (9)–(20) and are used in this comparison.

The numerical results measured on different databases are summarized in Tables 6–8. From these results, it can be seen that $OFIQA^{GA}_{TID2013}$ is able to provide the lowest error on TID2013 [76] and the second lowest on TID2008 [77], respectively. On the other hand, $OFIQA^{GA}_{TID2008}$ gives the lowest, second lowest, and third lowest error on TID2008 [77], CSIQ [33], and TID2013 [76], respectively. The RMSE values for each database are summarized in Figures 3–6. If we take a look at the correlation strengths, we can observe the following. On LIVE [4], the deep learning-based DeepSIM provides the highest correlation performance. However, the proposed $OFIQA^{GA}_{CSIQ}$'s results closely follow those of DeepSIM. Namely, the differences between the two methods are 0.01 and 0.02 in terms of PLCC and SROCC, respectively. Further, $OFIQA^{GA}_{TID2008}$ provides the third highest results in terms of PLCC and SROCC on CLIVE [4]. On TID2013 [76], $OFIQA^{GA}_{TID2013}$ gives the highest results in terms of PLCC and KROCC, while $OFIQA^{GA}_{TID2008}$ provides the third highest PLCC. On TID2008 [77], $OFIQA^{GA}_{TID2008}$ has the highest SROCC and KROCC values and it is outperformed by PLCC in the deep learning-based DeepIQA. Table 9 illustrates a summary of the direct and weighted averages of the correlation performance indices on the considered databases. It can be seen that $OFIQA^{GA}_{TID2008}$ has the highest performance in terms of PLCC and KROCC, if we consider the direct averages of the correlation strengths. $OFIQA^{GA}_{TID2008}$ preserves it first in terms of PLCC and gives the second highest performance of KROCC, if we consider the weighted averages. $OFIQA^{GA}_{TID2013}$ is the second best in terms of PLCC/SROCC and the third best in terms of KROCC. In the weighted averages, the VSI [68] is the best in terms of SROCC and the third best in terms of PLCC/KROCC. However, it gives a higher RMSE on the considered IQA databases than $OFIQA^{GA}_{TID2013}$ which is the second best weighted SROCC.
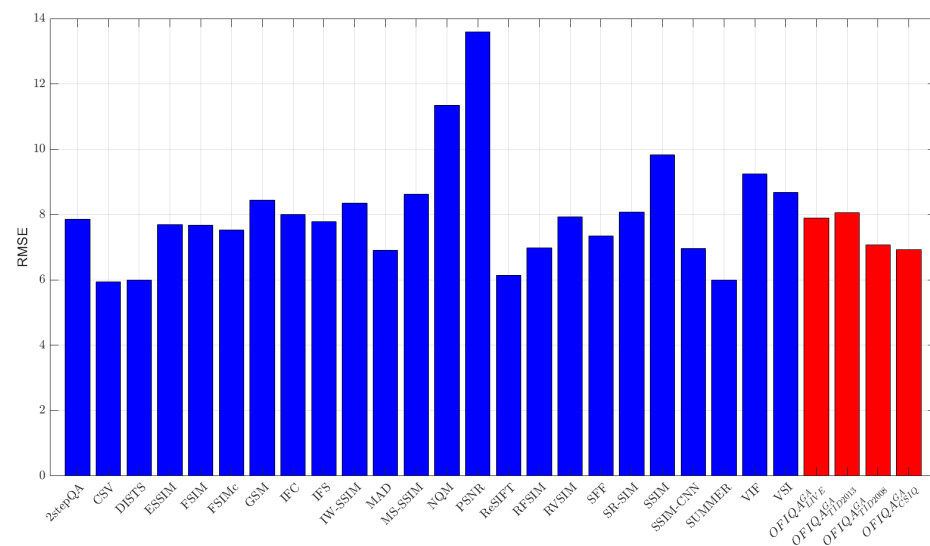

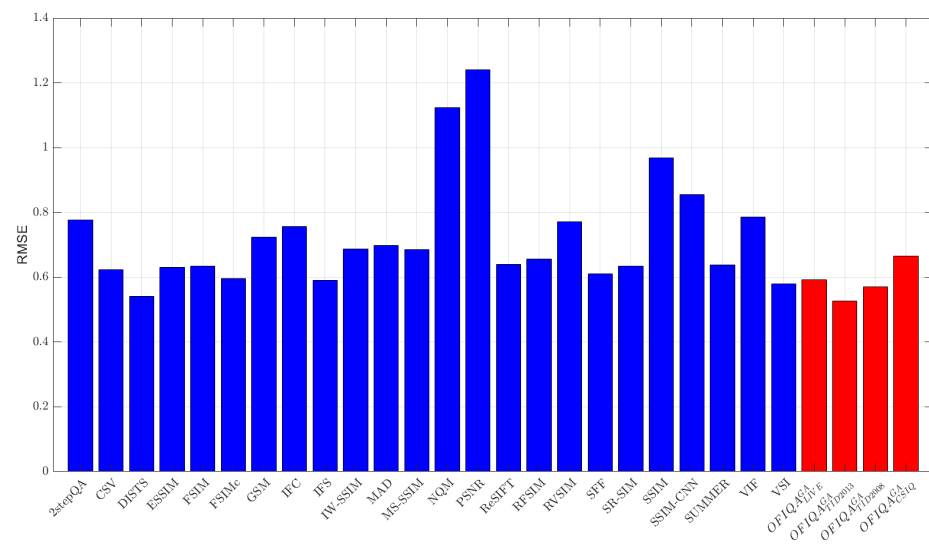**Figure 3.** RMSE measured on LIVE [4].

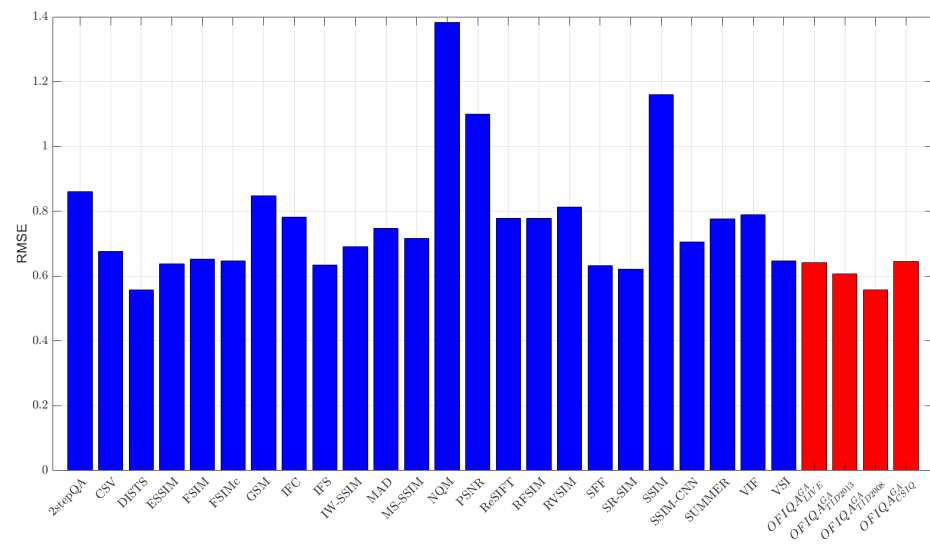**Figure 4.** RMSE measured on TID2013 [76].
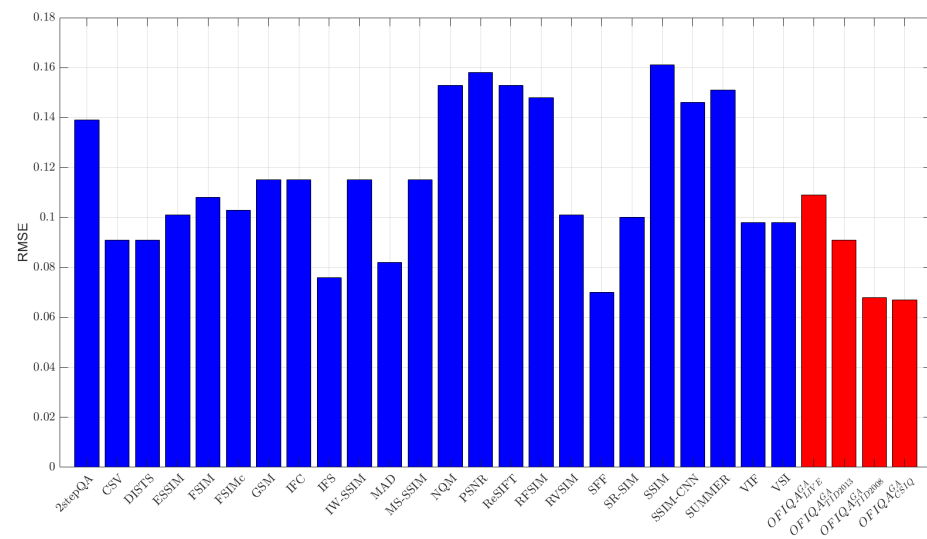

Figure 5. RMSE measured on TID2008 [77].

Figure 6. RMSE measured on CSIQ [33].

Table 6. RMSE performance comparison of the proposed fusion-based FR-IQA metrics with the state of the art on LIVE [4], TID2013 [76], TID2008 [77], and CSIQ [33]. The best results are typed in red, the second best results are in green, and the third best results are in blue.

FR-IQA Metric	LIVE [4]	TID2013 [76]	TID2008 [77]	CSIQ [33]
2stepQA [78]	7.856	0.776	0.861	0.139
CSV [79]	5.945	0.624	0.677	0.091
DISTS [80]	6.005	0.541	0.558	0.091
ESSIM [24]	7.689	0.630	0.638	0.101
FSIM [39]	7.678	0.635	0.653	0.108
FSIMc [39]	7.530	0.596	0.647	0.103
GSM [63]	8.433	0.723	0.847	0.115
IFC [9]	8.001	0.756	0.783	0.115
IFS [81]	7.776	0.591	0.635	0.076
IW-SSIM [26]	8.347	0.688	0.690	0.115
MAD [33]	6.907	0.698	0.747	0.082
MS-SSIM [8]	8.619	0.686	0.717	0.115
NQM [64]	11.347	1.123	1.382	0.153
PSNR [21]	13.596	1.240	1.099	0.158
ReSIFT [82]	6.145	0.640	0.779	0.153
RFSIM [37]	6.989	0.657	0.778	0.148
RVSIM [83]	7.927	0.772	0.813	0.101
SFF [65]	7.346	0.610	0.633	0.070
SR-SIM [66]	8.081	0.635	0.621	0.100
SSIM [13]	9.831	0.968	1.160	0.161
SSIM-CNN [16]	6.967	0.856	0.705	0.146
SUMMER [84]	6.002	0.638	0.777	0.151
VIF [62]	9.240	0.786	0.790	0.098
VSI [68]	8.681	0.580	0.647	0.098
GP-SSIM [85]	-	-	-	-
DeepSIM [86]	-	-	-	-
DeepIQA [44]	-	-	-	-
PieAPP [46]	-	-	-	-
LPIPS [45]	-	-	-	-
$OFIQA_{LIVE}^{GA}$	7.895	0.593	0.641	0.109
$OFIQA_{TID2013}^{GA}$	8.062	0.526	0.606	0.091
$OFIQA_{TID2008}^{GA}$	7.078	0.570	0.557	0.068
$OFIQA_{CSIQ}^{GA}$	6.918	0.665	0.645	0.067

Table 7. PLCC, SROCC, and KROCC performance comparison of the proposed fusion-based FR-IQA metrics with the state of the art on LIVE [4] and TID2013 [76]. The best results are typed in red, the second best results are in green, and the third best results are in blue.

FR-IQA Metric	LIVE [4]			TID2013 [76]		
	PLCC	SROCC	KROCC	PLCC	SROCC	KROCC
2stepQA [78]	0.937	0.932	0.828	0.736	0.733	0.550
CSV [79]	0.967	0.959	0.834	0.852	0.848	0.657
DISTS [80]	0.954	0.954	0.811	0.759	0.711	0.524
ESSIM [24]	0.963	0.962	0.840	0.740	0.797	0.627
FSIM [39]	0.960	0.963	0.833	0.859	0.802	0.629
FSIMc [39]	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

Table 7. *Cont.*

FR-IQA Metric	LIVE [4]			TID2013 [76]		
	PLCC	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**
IFC [9]	0.927	0.926	0.758	0.554	0.539	0.394
IFS [81]	0.959	0.960	0.825	0.879	0.870	0.679
IW-SSIM [26]	0.952	0.956	0.817	0.832	0.778	0.598
MAD [33]	0.967	0.967	0.842	0.827	0.778	0.600
MS-SSIM [8]	0.941	0.951	0.804	0.794	0.785	0.604
NQM [64]	0.912	0.909	0.741	0.690	0.643	0.474
PSNR [21]	0.872	0.876	0.687	0.616	0.646	0.467
ReSIFT [82]	0.961	0.962	0.838	0.630	0.623	0.471
RFSIM [37]	0.935	0.940	0.782	0.833	0.774	0.595
RVSIM [83]	0.641	0.630	0.495	0.763	0.683	0.520
SFF [65]	0.963	0.965	0.836	0.871	0.851	0.658
SR-SIM [66]	0.955	0.962	0.829	0.859	0.800	0.631
SSIM [13]	0.941	0.951	0.804	0.618	0.616	0.437
SSIM-CNN [16]	0.965	0.963	0.838	0.759	0.752	0.566
SUMMER [84]	0.967	0.959	0.833	0.623	0.622	0.472
VIF [62]	0.941	0.964	0.828	0.774	0.677	0.515
VSI [68]	0.948	0.952	0.805	0.900	0.894	0.677
GP-SSIM [85]	0.908	0.918	-	0.846	0.808	-
DeepSIM [86]	0.968	0.974	-	0.872	0.846	-
DeepIQA [44]	0.940	0.947	-	0.834	0.831	-
PieAPP [46]	0.908	0.919	-	0.859	0.876	-
LPIPS [45]	0.932	0.934	-	0.749	0.670	-
$OFIQA_{LIVE}^{GA}$	0.957	0.961	0.828	0.878	0.863	0.672
$OFIQA_{TID2013}^{GA}$	0.956	0.957	0.814	0.906	0.890	0.713
$OFIQA_{TID2008}^{GA}$	0.966	0.967	0.839	0.888	0.825	0.651
$OFIQA_{CSIQ}^{GA}$	0.967	0.972	0.854	0.844	0.808	0.634

Table 8. PLCC, SROCC, and KROCC performance comparison of the proposed fusion-based FR-IQA metrics with the state of the art on TID2008 [77] and CSIQ [33]. The best results are typed in red, the second best results are in green, and the third best results are in blue.

FR-IQA Metric	TID2008 [77]			CSIQ [33]		
	PLCC	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**
2stepQA [78]	0.757	0.769	0.574	0.841	0.849	0.655
CSV [79]	0.852	0.848	0.657	0.933	0.933	0.766
DISTS [80]	0.705	0.668	0.488	0.930	0.930	0.764
ESSIM [24]	0.658	0.876	0.696	0.814	0.933	0.768
FSIM [39]	0.874	0.881	0.695	0.912	0.924	0.757
FSIMc [39]	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 [9]	0.575	0.568	0.424	0.837	0.767	0.590
IFS [81]	0.879	0.869	0.678	0.958	0.958	0.817
IW-SSIM [26]	0.842	0.856	0.664	0.804	0.921	0.753
MAD [33]	0.831	0.829	0.639	0.950	0.947	0.797
MS-SSIM [8]	0.838	0.846	0.648	0.899	0.913	0.739
NQM [64]	0.608	0.624	0.461	0.743	0.740	0.564
PSNR [21]	0.447	0.489	0.346	0.853	0.809	0.599
ReSIFT [82]	0.627	0.632	0.484	0.884	0.868	0.695
RFSIM [37]	0.865	0.868	0.678	0.912	0.930	0.765

Table 8. *Cont.*

	TID2008 [77]			CSIQ [33]		
FR-IQA Metric	**PLCC**	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**
RVSIM [83]	0.789	0.743	0.566	0.923	0.903	0.728
SFF [65]	0.871	0.851	0.658	0.964	0.960	0.826
SR-SIM [66]	0.859	0.799	0.631	0.925	0.932	0.773
SSIM [13]	0.669	0.675	0.485	0.812	0.812	0.606
SSIM-CNN [16]	0.770	0.737	0.551	0.952	0.946	0.794
SUMMER [84]	0.817	0.823	0.623	0.826	0.830	0.658
VIF [62]	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
GP-SSIM [85]	0.859	0.892	-	0.928	0.953	-
DeepSIM [86]	0.876	0.887	-	0.919	0.919	-
DeepIQA [44]	0.917	0.908	-	0.901	0.909	-
PieAPP [46]	0.610	0.788	-	0.877	0.892	-
LPIPS [45]	0.772	0.731	-	0.896	0.876	-
$OFIQA_{LIVE}^{GA}$	0.879	0.888	0.700	0.910	0.938	0.786
$OFIQA_{TID2013}^{GA}$	0.892	0.904	0.722	0.938	0.923	0.754
$OFIQA_{TID2008}^{GA}$	0.910	0.911	0.738	0.966	0.964	0.833
$OFIQA_{CSIQ}^{GA}$	0.877	0.882	0.693	0.967	0.965	0.835

Table 9. 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 red, the second best results are in green, and the third best results are in blue.

	Direct Average			Weighted Average		
FR-IQA Metric	**PLCC**	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**
2stepQA [78]	0.818	0.821	0.652	0.781	0.783	0.605
CSV [79]	0.901	0.897	0.729	0.877	0.873	0.694
DISTS [80]	0.837	0.816	0.647	0.792	0.759	0.582
ESSIM [24]	0.794	0.892	0.733	0.756	0.857	0.691
FSIM [39]	0.901	0.893	0.729	0.883	0.860	0.689
FSIMc [39]	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 [9]	0.723	0.700	0.542	0.644	0.625	0.473
IFS [81]	0.919	0.914	0.750	0.900	0.893	0.715
IW-SSIM [26]	0.857	0.878	0.708	0.846	0.840	0.664
MAD [33]	0.894	0.880	0.720	0.862	0.838	0.667
MS-SSIM [8]	0.868	0.874	0.699	0.838	0.839	0.659
NQM [64]	0.738	0.729	0.560	0.703	0.684	0.516
PSNR [21]	0.697	0.705	0.525	0.634	0.654	0.480
ReSIFT [82]	0.776	0.771	0.622	0.705	0.700	0.550
RFSIM [37]	0.886	0.878	0.705	0.865	0.841	0.663
RVSIM [83]	0.779	0.740	0.577	0.777	0.723	0.558
SFF [65]	0.917	0.908	0.745	0.895	0.880	0.703
SR-SIM [66]	0.900	0.873	0.716	0.880	0.838	0.675
SSIM [13]	0.760	0.764	0.583	0.698	0.700	0.518
SSIM-CNN [16]	0.861	0.849	0.687	0.814	0.800	0.626
SUMMER [84]	0.808	0.809	0.647	0.745	0.746	0.582
VIF [62]	0.863	0.828	0.671	0.825	0.765	0.605
VSI [68]	0.919	0.921	0.744	0.909	0.908	0.716
GP-SSIM [85]	0.885	0.893	-	0.868	0.864	-
DeepSIM [86]	0.909	0.907	-	0.891	0.883	-
DeepIQA [44]	0.898	0.899	-	0.878	0.877	-
PieAPP [46]	0.814	0.869	-	0.801	0.860	-
LPIPS [45]	0.837	0.803	-	0.798	0.747	-
$OFIQA_{LIVE}^{GA}$	0.906	0.913	0.747	0.892	0.892	0.714
$OFIQA_{TID2013}^{GA}$	0.923	0.919	0.751	0.913	0.906	0.733
$OFIQA_{TID2008}^{GA}$	0.933	0.917	0.765	0.914	0.884	0.722
$OFIQA_{CSIQ}^{GA}$	0.914	0.907	0.754	0.885	0.869	0.704

5. Conclusions

This paper proposed a novel decision-fusion framework based on optimization for FR-IQA. First, the fusion-based metric was defined as a weighted product of n different traditional FR-IQA measures following Okarma's [7] ideas. Next, an optimization problem was specified using SROCC, KROCC, and the RMSE between the predicted and the ground-truth quality scores in the objective function to maximize the correlation strength and minimize the error. Then, several optimization techniques were applied to determine the weights in the weighted product of quality measures. To obtain a fusion-based metric with its parameters, 20% of the reference images and their distorted counterparts were used. Our analysis revealed that a GA is a suitable choice to solve the defined optimization problem. The experimental results and a comparison to the state of the art on four large, widely accepted benchmark databases, LIVE [4], TID2013 [76], TID2008 [77], and CSIQ [33], uncovered that the FR-IQA metrics coming from our optimization-based framework are able to outperform other traditional and deep learning-based state-of-the-art algorithms. Future research may involve the usage of other objective functions or multi-objective optimization.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	convolutional neural network
CPU	central processing unit
CQM	combined quality metric
CSIQ	categorical image quality
FR	full reference
FR-IQA	full-reference image quality assessment
GA	genetic algorithm
GPU	graphics processing unit
HIK	histogram intersection kernel
HVS	human visual system
IQA	image quality assessment
KROCC	Kendall's rank-order correlation coefficient
LIVE	laboratory for image and video engineering
NR	no reference
PLCC	Pearson's linear correlation coefficient
PS	pattern search
RMSE	root mean square error
RR	reduced reference
SROCC	Spearman's rank-order correlation coefficient
SSIM	structural similarity index
TID	Tampere image database

References

1. 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]
2. Chami, Z.A.; Jaoude, C.A.; Chbeir, R.; Barhamgi, M.; Alraja, M.N. A No-Reference and Full-Reference image quality assessment and enhancement framework in real-time. *Multimed. Tools Appl.* **2022**, *81*, 32491–32517 . [CrossRef]
3. 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.
4. 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] [PubMed]
5. 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.
6. Asadi, H.; Mohamed, S.; Lim, C.P.; Nahavandi, S. A review on otolith models in human perception. *Behav. Brain Res.* **2016**, *309*, 67–76. [CrossRef] [PubMed]
7. 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, Zakopane, Poland, 13–17 June 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 539–546.
8. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2, pp. 1398–1402.
9. 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]
10. 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]
11. Oszust, M. Full-reference image quality assessment with linear combination of genetically selected quality measures. *PLoS ONE* **2016**, *11*, e0158333. [CrossRef] [PubMed]
12. 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]
13. 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]
14. 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]
15. 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, SPIE, San Francisco, CA, USA, 8–12 February 2015; Volume 9394, pp. 172–183.
16. Amirshahi, S.A.; Pedersen, M.; Beghdadi, A. Reviving traditional image quality metrics using CNNs. In Proceedings of the Color and Imaging Conference. Society for Imaging Science and Technology, Vancouver, BC, Canada, 12–16 November 2018; Volume 2018, pp. 241–246.
17. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
18. Okarma, K. Combined image similarity index. *Opt. Rev.* **2012**, *19*, 349–354. [CrossRef]
19. Okarma, K. Extended hybrid image similarity–combined full-reference image quality metric linearly correlated with subjective scores. *Elektron. Elektrotechnika* **2013**, *19*, 129–132. [CrossRef]
20. Girod, B. What's wrong with mean-squared error? In *Digital Images and Human Vision*; MIT Press: Cambridge, MA, USA, 1993; pp. 207–220.
21. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
22. Lin, W.; Kuo, C.C.J. Perceptual visual quality metrics: A survey. *J. Vis. Commun. Image Represent.* **2011**, *22*, 297–312. [CrossRef]
23. 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]
24. Zhang, X.; Feng, X.; Wang, W.; Xue, W. Edge strength similarity for image quality assessment. *IEEE Signal Process. Lett.* **2013**, *20*, 319–322. [CrossRef]
25. Chen, G.H.; Yang, C.L.; Xie, S.L. Gradient-based structural similarity for image quality assessment. In Proceedings of the 2006 International Conference on Image Processing, Atlanta, GA, USA, 8–11 October 2006; pp. 2929–2932.
26. Wang, Z.; Li, Q. Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Process.* **2010**, *20*, 1185–1198. [CrossRef]
27. Li, C.; Bovik, A.C. Content-weighted video quality assessment using a three-component image model. *J. Electron. Imaging* **2010**, *19*, 011003.

28. Rouse, D.M.; Hemami, S.S. Understanding and simplifying the structural similarity metric. In Proceedings of the 2008 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 1188–1191.

29. Nilsson, J.; Akenine-Möller, T. Understanding ssim. *arXiv* **2020**, arXiv:2006.13846 .

30. Pedersen, M.; Hardeberg, J.Y. Full-reference image quality metrics: Classification and evaluation. *Found. Trends® Comput. Graph. Vis.* **2012**, *7*, 1–80.

31. 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.

32. Kingsbury, N. The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement. In Proceedings of the 9th European Signal Processing Conference (EUSIPCO 1998), Rhodes, Greece, 8–11 September 1998; pp. 1–4.

33. 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.

34. Kolaman, A.; Yadid-Pecht, O. Quaternion structural similarity: A new quality index for color images. *IEEE Trans. Image Process.* **2011**, *21*, 1526–1536. [CrossRef]

35. Lukin, V.; Bataeva, E.; Abramov, S. Saliency map in image visual quality assessment and processing. *Radioelectron. Comput. Syst.* **2023**, *1*, 112–121. [CrossRef]

36. Gu, K.; Zhai, G.; Lin, W.; Yang, X.; Zhang, W. Visual saliency detection with free energy theory. *IEEE Signal Process. Lett.* **2015**, *22*, 1552–1555. [CrossRef]

37. 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.

38. Martin, G.; Iwaniec, T. Riesz transforms and related singular integrals. *J. Reine Angew. Math. (Crelles J.)* **1996**, *1996*, 25–58.

39. 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]

40. Kovesi, P. Image features from phase congruency. *Videre J. Comput. Vis. Res.* **1999**, *1*, 1–26.

41. Götz-Hahn, F.; Hosu, V.; Lin, H.; Saupe, D. KonVid-150k: A Dataset for No-Reference Video Quality Assessment of Videos in-the-Wild. *IEEE Access* **2021**, *9*, 72139–72160. [CrossRef]

42. Hosu, V.; Hahn, F.; Jenadeleh, M.; Lin, H.; Men, H.; Szirányi, T.; Li, S.; Saupe, D. The Konstanz natural video database (KoNViD-1k). In Proceedings of the 2017 Ninth international conference on quality of multimedia experience (QoMEX), Erfurt, Germany, 31 May–2 June 2017; pp. 1–6.

43. 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]

44. Bosse, S.; Maniry, D.; Müller, K.R.; Wiegand, T.; Samek, W. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Trans. Image Process.* **2017**, *27*, 206–219. [CrossRef]

45. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 586–595.

46. Prashnani, E.; Cai, H.; Mostofi, Y.; Sen, P. Pieapp: Perceptual image-error assessment through pairwise preference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1808–1817.

47. Amirshahi, S.A.; Pedersen, M.; Yu, S.X. Image quality assessment by comparing CNN features between images. *J. Imaging Sci. Technol.* **2016**, *60*, 60410-1–60410-10. [CrossRef]

48. Barla, A.; Franceschi, E.; Odone, F.; Verri, A. Image kernels. In Proceedings of the International Workshop on Support Vector Machines, Niagara Falls, ON, Canada, 10 August 2002; Springer: Berlin, Germany, 2002; pp. 83–96.

49. Ai, D.; Liu, Y.; Yang, Y.; Lu, M.; Liu, Y.; Ling, N. A Full-Reference Image Quality Assessment Method with Saliency and Error Feature Fusion. In Proceedings of the 2022 IEEE International Symposium on Circuits and Systems (ISCAS), Austin, TX, USA, 28 May–1 June 2022; pp. 3165–3169.

50. Oszust, M. Image quality assessment with lasso regression and pairwise score differences. *Multimed. Tools Appl.* **2017**, *76*, 13255–13270. [CrossRef]

51. Yuan, Y.; Guo, Q.; Lu, X. Image quality assessment: A sparse learning way. *Neurocomputing* **2015**, *159*, 227–241. [CrossRef]

52. Merzougui, N.; Djerou, L. Multi-gene Genetic Programming based Predictive Models for Full-reference Image Quality Assessment. *J. Imaging Sci. Technol.* **2021**, *65*, 60409-1–60409-13. [CrossRef]

53. Koza, J.R. Genetic programming as a means for programming computers by natural selection. *Stat. Comput.* **1994**, *4*, 87–112. [CrossRef]

54. Merzougui, N.; Djerou, L. Genetic Programming for Screen Content Image Quality Assessment. In Proceedings of the Artificial Intelligence: Theories and Applications: First International Conference, ICAITA 2022, Mascara, Algeria, 7–8 November 2022; Revised Selected Papers; Springer: Berlin, Germany, 2023; pp. 52–64.

55. Varga, D. Full-Reference Image Quality Assessment Based on an Optimal Linear Combination of Quality Measures Selected by Simulated Annealing. *J. Imaging* **2022**, *8*, 224. [CrossRef] [PubMed]

56. Ieremeiev, O.; Lukin, V.; Okarma, K.; Egiazarian, K. Full-reference quality metric based on neural network to assess the visual quality of remote sensing images. *Remote Sens.* **2020**, *12*, 2349. [CrossRef]

57.	Jenadeleh, M. Blind Image and Video Quality Assessment. Ph.D. Thesis, University of Konstanz, Konstanz, Germany, 2018.
58.	Men, H. Boosting for Visual Quality Assessment with Applications for Frame Interpolation Methods. Ph.D. Thesis, University of Konstanz, Konstanz, Germany, 2022.
59.	Gao, M.-J.; Dang, H.-S.; Wei, L.-L.; Liu, G.-J.; Zhang, X.-D. Review and Prospect of Full Reference Image Quality Assessment. *Acta Electonica Sin.* **2021**, *49*, 2261.
60.	Phadikar, B.S.; Maity, G.K.; Phadikar, A. Full reference image quality assessment: A survey. In *Industry Interactive Innovations in Science, Engineering and Technology, Proceedings of the International Conference, I3SET 2016, Singapore, 10–12 June 2018*; Springer: Singapore, 2018; pp. 197–208.
61.	George, A.; Livingston, S.J. A survey on full reference image quality assessment algorithms. *Int. J. Res. Eng. Technol.* **2013**, *2*, 303–307.
62.	Sheikh, H.R.; Bovik, A.C. Image information and visual quality. *IEEE Trans. Image Process.* **2006**, *15*, 430–444. [CrossRef]
63.	Liu, A.; Lin, W.; Narwaria, M. Image quality assessment based on gradient similarity. *IEEE Trans. Image Process.* **2011**, *21*, 1500–1512.
64.	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]
65.	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] [PubMed]
66.	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.
67.	Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [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.	Chandler, D.M.; Hemami, S.S. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Trans. Image Process.* **2007**, *16*, 2284–2298. [CrossRef]
70.	Goldberg, D. *GeneticAtgorithms in Search, Optimization & Machine Learning*; Addison-Wesley: Boston, MA, USA, 1989.
71.	Audet, C.; Dennis Jr, J.E. Analysis of generalized pattern searches. *SIAM J. Optim.* **2002**, *13*, 889–903. [CrossRef]
72.	Lagarias, J.C.; Reeds, J.A.; Wright, M.H.; Wright, P.E. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM J. Optim.* **1998**, *9*, 112–147. [CrossRef]
73.	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]
74.	Shi, C.; Lin, Y. Image Quality Assessment Based on Three Features Fusion in Three Fusion Steps. *Symmetry* **2022**, *14*, 773. [CrossRef]
75.	Wu, J.; Lin, W.; Shi, G. Image quality assessment with degradation on spatial structure. *IEEE Signal Process. Lett.* **2014**, *21*, 437–440. [CrossRef]
76.	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.
77.	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.
78.	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]
79.	Temel, D.; AlRegib, G. CSV: Image quality assessment based on color, structure, and visual system. *Signal Process. Image Commun.* **2016**, *48*, 92–103. [CrossRef]
80.	Ding, K.; Ma, K.; Wang, S.; Simoncelli, E.P. Image quality assessment: Unifying structure and texture similarity. *arXiv* **2020**, arXiv:2004.07728.
81.	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]
82.	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.
83.	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]
84.	Temel, D.; AlRegib, G. Perceptual image quality assessment through spectral analysis of error representations. *Signal Process. Image Commun.* **2019**, *70*, 37–46. [CrossRef]

85.   Bakurov, I.; Buzzelli, M.; Schettini, R.; Castelli, M.; Vanneschi, L. Full-Reference Image Quality Expression via Genetic Programming. *IEEE Trans. Image Process.* **2023**, *32*, 1458–1473. [CrossRef]

86.   Gao, F.; Wang, Y.; Li, P.; Tan, M.; Yu, J.; Zhu, Y. Deepsim: Deep similarity for image quality assessment. *Neurocomputing* **2017**, *257*, 104–114. [CrossRef]