

Article

New Combined Metric for Full-Reference Image Quality Assessment

Mariusz Frackiewicz , Łukasz Machalica and Henryk Palus 

Department of Data Science and Engineering, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland; lukaszmachalica00@gmail.com (Ł.M.); henryk.palus@polsl.pl (H.P.)

* Correspondence: mariusz.frackiewicz@polsl.pl; Tel.: +48-32-2371066

Abstract: In recent years, many new metrics highly correlated with the Mean Opinion Score (MOS) have been proposed for assessing image quality through Full-Reference Image Quality Assessment (FR-IQA) methods, such as MDSI, HPSI, and GMSD. Eight of these selected metrics, which compare reference and distorted images in a symmetrical manner, are briefly described in this article, and their performance is evaluated using correlation criteria (PLCC, SROCC, and KROCC), as well as RMSE. The aim of this paper is to develop a new, efficient quality index based on a combination of several high-performance metrics already utilized in the field of Image Quality Assessment (IQA). The study was conducted on four benchmark image databases (TID2008, TID2013, KADID-10k, and PIPAL) and identified the three best-performing metrics for each database. The paper introduces a New Combined Metric (NCM), which is a weighted sum of three component metrics, and demonstrates its superiority over each of its component metrics across all the examined databases. An optimization method for determining the weights of the NCM is also presented. Additionally, an alternative version of the combined metric, based on the fastest metrics and employing symmetric calculations for pairs of compared images, is discussed. This version also demonstrates strong performance.

Keywords: image quality assessment; quality metrics; combined metrics; image databases; mean opinion scores



Citation: Frackiewicz, M.; Machalica, Ł.; Palus, H. New Combined Metric for Full-Reference Image Quality Assessment. *Symmetry* **2024**, *16*, 1622. <https://doi.org/10.3390/sym16121622>

Academic Editor: Jie Yang

Received: 14 November 2024

Revised: 2 December 2024

Accepted: 5 December 2024

Published: 7 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Every day, a huge number of digital cameras generate a vast stream of images. Due to the multitude of applications of imaging devices in areas such as vision-based quality control of components in manufacturing processes, security monitoring, object detection systems in automotive applications, and analysis of diagnostic images in medicine, there has been a strong increase in the demand for Image Quality Assessment (IQA) methods.

Image quality can be assessed either subjectively or objectively. Subjective methods rely on the perceptual evaluation of image quality by human observers, which means that conducting these assessments incurs significant financial costs and requires a large number of participants. In contrast, objective methods utilize mathematical models to determine the values of various metrics related to image quality. Among Image Quality Assessment (IQA) methods, the most advanced are those that perform a symmetrical comparison between distorted images and their originals, referred to as Full-Reference IQA (FR-IQA) techniques. The scores generated by each IQA metric can be evaluated against subjective assessments, such as the Mean Opinion Score (MOS) or the Difference Mean Opinion Score (DMOS), derived from human viewers. For FR-IQA methods, the correlation coefficients obtained from comparisons with MOS indicate the effectiveness of the metric: the higher the coefficient, the more closely the metric aligns with human perception. For many years, efforts have been made in the field of FR-IQA to improve and refine existing quality metrics. Significant importance is attached to attempts to combine one metric with other quality measures to increase the correlation of the resulting quality index with MOS. Meanwhile, the rapid development of machine-learning and deep-learning techniques provides new methods for image quality assessment. The FR-IQA problem can be understood as a

challenge in developing mathematical models that can perceptually assess the image quality in alignment with human judgment.

Over the years, numerous metrics for FR-IQA have been proposed that take various aspects of the Human Visual System (HVS) into account. Recently, attempts have been made to enhance the effectiveness of FR-IQA by combining existing metrics to create a “super” index. Theoretical foundations for such metric fusion can be found in Liu’s work [1], where it is applied to very old and classical metrics such as PSNR, VSNR, SSIM, and VIF. In the paper by Okarma [2], the properties of three FR-IQA metrics (MS-SSIM, VIF, and R-SVD) were analyzed, and a combined quality metric based on their product was proposed. It is named the Combined Quality Metric (CQM), and its three correlation coefficients increased in relation to the correlation coefficients with MOS of individual product multipliers.

Later, this concept was further developed using optimization or regression techniques to determine the optimal weights or exponents in the products of existing FR-IQA indices [3,4]. The Combined Image Similarity Index (CISI) proposed by Okarma in [3] employs metrics similar to those used in the CQM. However, instead of the R-SVD metric, it utilizes the FSIMc metric. The CISI index demonstrates a higher correlation with Mean Opinion Score (MOS) compared to the CQM index. In the 2013 article by Okarma [4], a new EHIS metric is introduced, which is based on the product of four multipliers: two familiar from CQM (MS-SSIM and VIF) and two new ones (WFSIMc and RFSIM). This approach improves the correlation with MOS over the combined CISI metric.

Another metrics fusion strategy was proposed by the author of [5], who presented several versions of a linear combination (weighted sum) based on metrics selected from a dozen FR-IQA metrics. He referred to these new combined metrics as Linearly Combined Similarity Measures (LCSIM). The use of linear combination metrics requires determining the weighting factors for the FR-IQA metrics, which is achieved by solving the RMSE error minimization task using a genetic algorithm.

Another line of work in creating methods based on the fusion of metrics utilizes machine-learning techniques. One example can be seen in [6], where the results of six traditional FR-IQA metrics (FSIMc, PSNR-HMA, PSNR-HVS, SFF, SR-SIM, and VIF) were used as a feature vector for training and testing a four-layer neural network. The output results produced by the neural network demonstrate a significant improvement over those achieved by the input metrics. Currently, deep neural networks, particularly CNNs, can learn the best combinations of metrics to optimize image quality assessment [7,8].

Combining metrics leads to increased computational complexity, which can be an issue in the context of real-time applications. Nevertheless, the fusion of quality metrics for FR-IQA is a promising research direction that could significantly improve the correlation of objective quality assessments with MOS.

In this paper, we consider a new combined metric for FR-IQA, based on metrics that are highly correlated with MOS and well-known from the literature. The structure of this paper is as follows: following the introduction, Section 2 provides an overview of eight relatively new and promising FR-IQA metrics. Section 3 presents their linear combination, along with the experimental results. Finally, Section 4 concludes the paper.

2. Overview of Highly Correlated FR-IQA Metrics

2.1. Feature SIMilarity Index FSIMc (Color Version)

In [9], the Feature SIMilarity Index (FSIM) was introduced as a quality assessment metric for grayscale images, along with its color version, FSIMc. The local quality of the assessed image, f_2 , when symmetrically compared to the reference image, f_1 , is represented by two low-level similarity maps derived from phase congruences (PC_1 , PC_2) [10]:

$$S_{PC}(x, y) = \frac{2PC_1(x, y) \cdot PC_2(x, y) + T_1}{PC_1^2(x, y) + PC_2^2(x, y) + T_1} \quad (1)$$

and gradient magnitudes (G_1, G_2):

$$S_G(x, y) = \frac{2G_1(x, y) \cdot G_2(x, y) + T_2}{G_1^2(x, y) + G_2^2(x, y) + T_2}, \quad (2)$$

where G_1 and G_2 are Scharr gradient operators and T_1 and T_2 are positive constants introduced to enhance the stability of the formulas. Phase Congruence (PC) quantifies the presence and intensity of local features, including edges, corners, and textures. It is derived from the analysis of phase information in the frequency domain of the image (specifically, the phase of its Fourier transform). As an additional weighting factor for the similarity maps, the following PC_m is used:

$$PC_m(x, y) = \max(PC_1(x, y), PC_2(x, y)). \quad (3)$$

The final expression for the proposed image quality index is given by:

$$FSIM = \frac{\sum_{(x,y) \in \Omega} S_{PC}(x, y) \cdot S_G(x, y) \cdot PC_m(x, y)}{\sum_{(x,y) \in \Omega} PC_m(x, y)}. \quad (4)$$

FSIMc, the color extension of FSIM, incorporates the chrominance components I and Q . The calculation of the index values starts by decomposing the compared images, f_1 and f_2 , into their YIQ color components, where Y represents luminance and I and Q represent chrominance. Similar to the earlier defined similarity maps, additional similarity maps for the I and Q components were introduced:

$$S_I(x, y) = \frac{2I_1(x, y) \cdot I_2(x, y) + T_3}{I_1^2(x, y) + I_2^2(x, y) + T_3}, \quad (5)$$

$$S_Q(x, y) = \frac{2Q_1(x, y) \cdot Q_2(x, y) + T_4}{Q_1^2(x, y) + Q_2^2(x, y) + T_4}, \quad (6)$$

where T_3 and T_4 are positive constants. The overall chrominance similarity map is given by:

$$S_C(x, y) = S_I(x, y) \cdot S_Q(x, y). \quad (7)$$

The inclusion of chromatic components in the FSIM index results in the following version of the formula for color images:

$$FSIMc = \frac{\sum_{(x,y) \in \Omega} S_{PC}(x, y) \cdot S_G(x, y) \cdot [S_C(x, y)]^\lambda \cdot PC_m(x, y)}{\sum_{(x,y) \in \Omega} PC_m(x, y)}, \quad (8)$$

where the positive value of the λ exponent highlights the significance of chrominance in the color image quality assessment process. For subsequent studies utilizing FSIMc, the following parameter values were employed, as specified in [9]: $T_1 = 0.85$, $T_2 = 160$, $T_3 = T_4 = 200$, and $\lambda = 0.03$.

2.2. Mean Deviation Similarity Index MDSI

Many IQA metrics work as follows: they determine local distortions in the images, build similarity maps, and implement a pooling strategy based on the mean, weighted mean, standard deviation, etc. An example of this approach to IQA index modeling is the Mean Deviation Similarity Index (MDSI), described in [11]. The calculation of MDSI starts with converting the RGB color space components of the input images to a luminance component:

$$L = 0.2989R + 0.5870G + 0.1140B \quad (9)$$

and two chromaticity components:

$$\begin{bmatrix} H \\ M \end{bmatrix} = \begin{bmatrix} 0.30 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (10)$$

This index is derived from the calculation of gradient similarity (GS) for structural distortions and chromaticity similarity (CS) for color distortions. The local structural similarity map is typically computed using gradient values. Traditionally, structural similarity maps are obtained by calculating the gradient values separately for the original and distorted images. The traditional approach for the MDSI metric has been improved by integrating the gradient value map, which combines the luminance channel values from both images:

$$f = 0.5(L_r + L_d), \quad (11)$$

where f represents the fused luminance image, r is the reference image, and d refers to the distorted image. The formulas for the proposed structural similarity are given below:

$$GS_{rf}(x) = \frac{2G_r(x)G_f(x) + C_2}{G_r^2(x) + G_f^2(x) + C_2}, \quad (12)$$

$$GS_{df}(x) = \frac{2G_d(x)G_f(x) + C_2}{G_d^2(x) + G_f^2(x) + C_2}, \quad (13)$$

$$\widehat{GS}(x) = GS(x) + [GS_{df}(x) - GS_{rf}(x)]. \quad (14)$$

The gradient magnitude is calculated using the simple Prewitt operator. Additionally, the authors of the MDSI index have adjusted the method for evaluating local chromaticity similarity. In contrast to the previously discussed IQA metrics, such as FSIM or VSI, which assess chromaticity separately for the two chrominance components, this approach combines them in a different way. In the case of MDSI, it was suggested to calculate the color similarity for both chrominance components simultaneously, using the following formula:

$$\widehat{CS}(x) = \frac{2(H_r(x)H_d(x) + M_r(x)M_d(x)) + C_3}{H_r^2(x) + H_d^2(x) + M_r(x)^2 + M_d(x)^2 + C_3}, \quad (15)$$

where C_3 is a constant introduced for numerical stability. The joint color similarity map, $\widehat{CS}(x)$, is then combined with the $\widehat{GS}(x)$ map using a weighted mean:

$$\widehat{GCS}(x) = \alpha \widehat{GS}(x) + (1 - \alpha) \widehat{CS}(x), \quad (16)$$

where α determines the relative importance of the similarity maps $\widehat{GS}(x)$ and $\widehat{CS}(x)$. The final step involves converting the resulting \widehat{GCS} map into an MDSI score through a pooling strategy based on a specific deviation method:

$$MDSI = \left[\frac{1}{N} \sum_{i=1}^N |\widehat{GCS}_i|^{1/4} - \left(\frac{1}{N} \sum_{i=1}^N |\widehat{GCS}_i|^{1/4} \right) \right]^{1/4}. \quad (17)$$

The original article [11] provides suggestions for selecting various parameters that influence the performance of the MDSI index.

2.3. Haar Wavelet Perceptual Similarity Index HPSI

The Haar Wavelet-based Perceptual Similarity Index (HPSI) [12] is a relatively novel and computationally efficient similarity metric for FR-IQA. HPSI uses coefficients obtained

from Haar wavelet decomposition for assessing local similarities between two images. The one-dimensional Haar filters are given by:

$$h_1^{1D} = \frac{1}{\sqrt{2}} \cdot [1, 1], \quad (18)$$

$$g_1^{1D} = \frac{1}{\sqrt{2}} \cdot [-1, 1], \quad (19)$$

where h_1^{1D} represents the low-pass scaling filter and g_1^{1D} refers to the corresponding high-pass wavelet filter. For any scale, $j \in \mathbb{N}$, two-dimensional Haar filters can be constructed as follows:

$$g_j^{(1)} = g_1^{1D} \otimes h_1^{1D}, \quad (20)$$

$$g_j^{(2)} = h_1^{1D} \otimes g_1^{1D}, \quad (21)$$

where the symbol \otimes denotes the outer product, and the one-dimensional filters h_j^{1D} and g_j^{1D} for $j > 1$ are defined as:

$$g_j^{1D} = h_1^{1D} * (g_{j-1}^{1D})_{\uparrow 2}, \quad (22)$$

$$h_j^{1D} = h_1^{1D} * (h_{j-1}^{1D})_{\uparrow 2}, \quad (23)$$

where $\uparrow 2$ is the dyadic upsampling operator and $*$ denotes the one-dimensional convolution operator. To effectively predict the perceptual similarity perceived by human viewers, it may be beneficial to apply an additional nonlinear mapping to the local similarities derived from the high-frequency Haar wavelet filter responses. This nonlinearity is represented by a logistic function, defined with a parameter $\alpha > 0$, as follows:

$$l_\alpha(x) = \frac{1}{1 + e^{-\alpha x}}. \quad (24)$$

For two grayscale images, f_1 and f_2 , the local similarity measure employed to calculate the HPSI is derived from the first two steps of the two-dimensional discrete Haar wavelet transform, as expressed by the following formula:

$$HS_{f_1, f_2}^{(k)}[x] = l_\alpha\left(\frac{1}{2} \sum_{j=1}^2 S(|(g_j^{(k)} * f_1)[x]|, |(g_j^{(k)} * f_2)[x]|, C)\right), \quad (25)$$

where $C > 0$, $k \in \{1, 2\}$ selects either horizontal or vertical Haar wavelet filters, S denotes the similarity measure, and $*$ is the two-dimensional convolution operator. Similar to FSIMc, HPSI also applies a specific weighting map, which is derived here from the response of a single low-frequency Haar wavelet filter:

$$W_f^{(k)}[x] = |(g_3^{(k)} * f)[x]|, \quad (26)$$

where $k \in \{1, 2\}$ again differentiates between horizontal and vertical filters. The final expression for the HPSI for grayscale images f_1 and f_2 is provided as a weighted average of the local similarity map, $HS_{f_1, f_2}^{(k)}$:

$$HPSI_{f_1, f_2} = l_\alpha^{-1}\left(\frac{\sum_x \sum_{k=1}^2 HS_{f_1, f_2}^{(k)}[x] \cdot W_{f_1, f_2}^{(k)}[x]}{\sum_x \sum_{k=1}^2 W_{f_1, f_2}^{(k)}[x]}\right)^2, \quad (27)$$

where:

$$W_{f_1, f_2}^{(k)}[x] = \max(W_{f_1}^{(k)}[x], W_{f_2}^{(k)}[x]). \quad (28)$$

HPSI can be extended for color images in the YIQ color space using a third local similarity map based on the chrominance components I and Q . This map, $HS_{f_1, f_2}^{(3)}$, is defined as:

$$HS_{f_1, f_2}^{(3)}[x] = l_\alpha \left(\frac{1}{2} (S(|(m * f_1^I)[x]|, |(m * f_2^I)[x]|, C) + S(|(m * f_1^Q)[x]|, |(m * f_2^Q)[x]|, C)) \right), \quad (29)$$

where m is a 2×2 mean filter and then:

$$W_{f_1^Y, f_2^Y}^{(3)}[x] = \frac{1}{2} (W_{f_1^Y, f_2^Y}^{(1)}[x] + W_{f_1^Y, f_2^Y}^{(2)}[x]). \quad (30)$$

The final form of the HPSI for color images is defined as follows:

$$HPSI_{c_{f_1, f_2}} = l_\alpha^{-1} \left(\frac{\sum_x \sum_{k=1}^3 HS_{f_1, f_2}^{(k)}[x] \cdot W_{f_1^Y, f_2^Y}^{(k)}[x]}{\sum_x \sum_{k=1}^3 W_{f_1^Y, f_2^Y}^{(k)}[x]} \right)^2. \quad (31)$$

The fast computation time of the HPSI may explain its high usefulness in various tasks.

2.4. Visual Saliency with Color Appearance and Gradient Similarity Index VCGS

The VCGS index [13] uses color space CIELAB and combines three feature similarity maps: visual saliency with color appearance similarity map, S_{VC} , gradient similarity map, S_G , and chrominance similarity map, S_C . The first of these maps is calculated using a formula based on visual saliency with color appearance (VC) for both images:

$$S_{VC} = \frac{2VC_1 \cdot VC_2 + K_{VC}}{VC_1^2 + VC_2^2 + K_{VC}}, \quad (32)$$

where K_{VC} is a small constant that controls the numerical stability of the formula. The gradient similarity map using the Scharr operator applied to the L component is calculated according to the formula:

$$S_G = \frac{2G_1 \cdot G_2 + K_G}{G_1^2 + G_2^2 + K_G}, \quad (33)$$

where K_G is a small constant that controls the numerical stability of the formula. The third map measures the similarity of the a^* and b^* chrominance components in the CIELAB color space and is given by the formula:

$$S_C = \frac{2a_1 \cdot a_2 + K_C}{a_1^2 + a_2^2 + K_C} \cdot \frac{2b_1 \cdot b_2 + K_C}{b_1^2 + b_2^2 + K_C}, \quad (34)$$

where K_C is a small constant that controls the numerical stability of the formula. The final form of the VCGS metric is given by the following formula:

$$VCGS = \frac{\sum_{\Omega} S_{VC} \cdot (S_G)^\alpha \cdot (S_C)^\lambda \cdot VC_m}{\sum_{\Omega} VC_m}, \quad (35)$$

where Ω is the spatial domain, $VC_m = \max(VC_1, VC_2)$ is used to weight the relevance of two maps in overall similarity, and α and λ represent the relative importance of the similarity maps depending on where they occur.

2.5. SuperPixel SIMilarity Index SPSIM

Superpixel-based SIMilarity (SPSIM) [14] utilizes superpixel segmentation for feature extraction. Superpixels are clusters of neighboring pixels that share similar characteristics, such as color, intensity, or structure. This pixel grouping results in a mosaic consisting of a

significantly smaller number of superpixels, which facilitates faster subsequent processing. A key advantage of using superpixel-based segmentation over other oversegmentation algorithms is the ability to predefine the number of generated superpixels. Additionally, superpixel segmentation improves the distinction of perceptually significant regions in the image. Among the various superpixel generation algorithms, we can distinguish between graph-based, gradient-based, clustering-based, and watershed-based methods, among others [15]. The shape and size of superpixels depend on the applied algorithm, with each pixel belonging to exactly one superpixel. These algorithms control the number and properties of the superpixels, such as compactness and minimum size. One of the most popular and efficient algorithms for superpixel segmentation is the k-means-based Simple Linear Iterative Clustering (SLIC) algorithm [16]. This algorithm is notable for producing superpixels with a consistent shape and size. A key benefit of SLIC is that segmentation only requires specifying the desired number of superpixels in the output image. Consequently, the SLIC algorithm is used in the SPSIM quality index discussed in this paper. For each superpixel, the algorithm calculates the mean CIELAB color values and the Local Binary Pattern (LBP) features. Superpixels are initially generated on the reference image and then applied to both the reference and distorted images.

The SPSIM index calculation algorithm relies on pixel gradient similarity and luminance-chrominance superpixel similarity. The YUV color space, rather than RGB, is utilized for SPSIM computation, where Y represents luminance and U and V denote chrominance components. If s_i is used to represent a superpixel containing pixel i , the following formulas can be written for luminance L_i and luminance similarity $M_L(i)$:

$$L_i = \frac{1}{|s_i|} \sum_{j \in s_i} Y(j), \quad M_L(i) = \frac{2L_r(i)L_d(i) + T_1}{L_r^2(i) + L_d^2(i) + T_1}, \quad (36)$$

where $Y(j)$ represents the luminance of pixel j and $L_r(i)$ and $L_d(i)$ denote the average luminance values for superpixel s_i in the reference and distorted images, respectively. T_1 is a positive constant introduced to prevent instability in the equation. Similar expressions can be formulated for both the U and V chrominance components:

$$U_i = \frac{1}{|s_i|} \sum_{j \in s_i} U(j), \quad M_U(i) = \frac{2U_r(i)U_d(i) + T_1}{U_r^2(i) + U_d^2(i) + T_1}, \quad (37)$$

$$V_i = \frac{1}{|s_i|} \sum_{j \in s_i} V(j), \quad M_V(i) = \frac{2V_r(i)V_d(i) + T_1}{V_r^2(i) + V_d^2(i) + T_1}. \quad (38)$$

The chrominance similarity, M_C , can then be calculated as shown below:

$$M_C(i) = M_U(i)M_V(i). \quad (39)$$

The gradient similarity, M_G , is described by the following formula:

$$M_G(i) = \frac{2G_r(i)G_d(i) + T_2}{G_r^2(i) + G_d^2(i) + T_2}, \quad (40)$$

where the gradient magnitude, G , is composed of two components calculated using a simple Prewitt operator, and T_1 and T_2 are constants selected by the authors of the algorithm to account for contrast-related errors. Further information on the determination of T_1 and T_2 can be found in [14]. The formula for calculating the similarity of superpixel i in both images is as follows:

$$M(i) = M_G(i)[M_L(i)]^\alpha e^{\beta(M_C(i)-1)}, \quad (41)$$

where the parameters α and β represent the weights for the luminance and chrominance components, respectively. Finally, the SPSIM index is calculated as a weighted sum of $M(i)$

and the corresponding weights, which are determined based on the texture complexity (TC), described by the standard deviation (*std*) and kurtosis (*Kurt*) of the superpixels:

$$TC_r(i) = \frac{std(S_r(i))}{Kurt[S_r(i)] + 3}, TC_d(i) = \frac{std(S_d(i))}{Kurt[S_d(i)] + 3}, \quad (42)$$

$$w(i) = \exp(0.05 \cdot \text{abs}(TC_d(i) - TC_r(i))), \quad (43)$$

$$SPSIM = \frac{\sum_{i=1}^N M(i)w(i)}{\sum_{i=1}^N w(i)}, \quad (44)$$

where $S_r(i)$ and $S_d(i)$ are, respectively, the superpixels in the reference and distorted images that contain the i -th pixel.

2.6. Local Global Variation Index LGV and Saliency Weighted Local Global Variation Index SWLGV

Varga [17] introduced new quality indices that utilize both gradients in the image and Grünwald–Letnikov fractional derivatives. While gradients capture local variations within the image, fractional derivatives describe global variations, represented by the flowing global similarity map:

$$S_G(x, y) = \frac{2 \cdot {}^{GL}D^\alpha R(x, y) \cdot {}^{GL}D^\alpha D(x, y) + c_1}{({}^{GL}D^\alpha R(x, y))^2 + ({}^{GL}D^\alpha D(x, y))^2 + c_1}, \quad (45)$$

where $R(x, y)$ is the reference image, $D(x, y)$ is the distorted image, ${}^{GL}D^\alpha$ is the α -order Grünwald–Letnikov fractional derivative, and c_1 is a constant number that provides numerical stability. The order of the fractional derivative was set to $\alpha = 0.6$. The 3×3 Scharr operator was used to compute the gradients for the local gradient map, $S_L(x, y)$:

$$S_L(x, y) = \frac{2 \cdot G_R(x, y) \cdot G_D(x, y) + c_2}{G_R^2(x, y) + G_D^2(x, y) + c_2}, \quad (46)$$

where c_2 is a constant that ensures numerical stability. The similarity map between the two compared images was calculated using the two previously defined gradient maps and the exponential coefficient $\lambda = 0.7$:

$$S(x, y) = (S_G(x, y))^\lambda \cdot (S_L(x, y))^{1-\lambda}. \quad (47)$$

Finally, the similarity map obtained is fused with the saliency map. This index is referred to as the Local Global Variation (LGV):

$$LGV = \frac{1}{M \cdot N} \sum_{x=1}^M \sum_{y=1}^N S(x, y), \quad (48)$$

where $M \cdot N$ is the resolution of images.

The SWLGV index, in contrast to LGV, also incorporates the mechanism of visual saliency. It emphasizes the differences between the reference and distorted images in the most distinctive regions. By labeling the maps of the distinguishing regions as $SM_R(x, y)$ for the reference image and $SM_D(x, y)$ for the distorted image, we can create a formula for the image pair:

$$SM(x, y) = \max(SM_R(x, y), SM_D(x, y)), \quad (49)$$

where $SM_R(x, y)$ and $SM_D(x, y)$ are visual saliency maps built as proposed in [18]. The SWLGV index is defined as the weighted average of $S(x, y)$ and $SM(x, y)$, where $SM(x, y)$ represents the weights:

$$SWLGV = \frac{\sum_{i=1}^M \sum_{j=1}^N SM(x, y) \cdot S(x, y)}{\sum_{i=1}^M \sum_{j=1}^N SM(x, y)}. \quad (50)$$

2.7. Gradient Magnitude Similarity Deviation Index GMSD

GMSD [19] is a relatively simple metric, which is based on a gradient similarity map and uses a 3×3 Prewitt filter. The magnitudes of the gradients of images r and d at position i , denoted by $m_r(i)$ and $m_d(i)$, are calculated as follows:

$$m_r(i) = \sqrt{(r \otimes h_x)^2(i) + (r \otimes h_y)^2(i)}, \quad (51)$$

$$m_d(i) = \sqrt{(d \otimes h_x)^2(i) + (d \otimes h_y)^2(i)}, \quad (52)$$

where \otimes denotes a convolution operation. The magnitude gradient similarity map, $GMS(i)$, is then calculated as follows:

$$GMS(i) = \frac{2m_r(i)m_d(i) + c}{m_r^2(i) + m_d^2(i) + c}, \quad (53)$$

where c is a constant number that provides numerical stability. The formulas above demonstrate a symmetrical approach to both referenced and distorted images. The average gradient value from the $GMS(i)$ map was then determined as:

$$GMSM = \frac{1}{N} \sum_{i=1}^N GMS(i), \quad (54)$$

where N is a number of pixels in image. Finally, the GMSD index is defined by the formula:

$$GMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (GMS(i) - GMSM)^2}. \quad (55)$$

2.8. Evaluation Criteria for IQA

Individual IQA metrics are commonly compared with the subjective ratings of specific images. To assess the linearity, monotonicity, and accuracy of these predictions, four criteria are used: the Pearson Linear Correlation Coefficient (PLCC), the Spearman Rank Order Correlation Coefficient (SROCC), the Kendall Rank Order Correlation Coefficient (KROCC), and the Root Mean Squared Error (RMSE). The formulas for these comparisons are provided below:

$$PLCC = \frac{\sum_{i=1}^N (p_i - \bar{p})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^N (p_i - \bar{p})^2 (s_i - \bar{s})^2}}, \quad (56)$$

where p_i and s_i represent the raw values of the subjective and objective measures, respectively, and \bar{p} and \bar{s} are the mean values of the subjective and objective measures.

Spearman Rank Order Correlation Coefficient is given by the formula:

$$SROCC = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)}, \quad (57)$$

where d_i represents the difference between the ranks of both measures for the i -th observation and N is the total number of observations.

Kendall Rank Order Correlation Coefficient (KROCC) is provided by the formula:

$$KROCC = \frac{N_c - N_d}{0.5(N-1)N} \quad (58)$$

where N_c and N_d denote the counts of concordant and discordant pairs.

Root Mean Squared Error (RMSE) is given by the following equation:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - s_i)^2} \quad (59)$$

where p_i and s_i are defined as above.

The above correlation coefficients are useful tools for objectively assessing the agreement between IQA computational models and subjective MOS assessments. However, they capture specific aspects of this relationship, such as linearity in the case of PLCC or monotonicity in the cases of SROCC and KROCC. SROCC and KROCC are suitable for scenarios where the relationship between variables is nonlinear, with KROCC offering high robustness to small changes in the data. RMSE, on the other hand, is a measure of error primarily used to evaluate the accuracy of a model's predictions. Unlike PLCC, SROCC, and KROCC, it is not a measure of correlation, and its role is fundamentally different. RMSE quantifies the average distance between predicted and actual values. It is sensitive to the magnitude of errors because the differences are squared before averaging, meaning that larger deviations have a disproportionate impact on the final value. High RMSE values indicate poor agreement between predicted and actual values, but RMSE does not provide information about the type of relationship (e.g., whether it is monotonic, linear, or otherwise).

As recommended in [20], a nonlinear mapping was applied to calculate PLCC and RMSE. This process involves the use of a fitting function, usually a logistic function with five beta parameters, $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$, to better represent the relationship between predicted performance, x , and MOS.

$$p(x, \beta) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(x - \beta_3))} \right) + \beta_4 x + \beta_5 \quad (60)$$

3. The New Combined Metric (NCM) and Its Experimental Research

The New Combined Metric (NCM):

$$NCM = \alpha \cdot M_1 + \beta \cdot M_2 + \gamma \cdot M_3, \quad (61)$$

where M_1 , M_2 , and M_3 are the selected FR-IQA metrics for given dataset and α , β , and γ are the optimized weights.

3.1. Selected IQA Databases

Four benchmark databases, TID2008 [21], TID2013 [22], KADID-10k [23], and PIPAL [24], were chosen for the research. These databases are distinguished by a large set of reference images, diverse distortion types, and varying levels of their presence in the images. For each image in the databases, Mean Opinion Scores (MOSs) are experimentally gathered by collecting assessments from multiple human observers.

The TID2008 image database consists of 1700 distorted images, generated using 17 different distortion types, each applied at four levels to 25 reference images (Figure 1). MOS was provided based on the work of 838 human observers and 256,428 comparisons. All images have a resolution of 512×384 pixels.

The TID2013 image database is an updated and expanded version of TID2008. It retains the same set of reference images (Figure 1), but the number of distortion types has been increased to 24, and the distortion levels have been raised to five. The database includes 3000 distorted digital images. Additionally, the size of the research group from

which the average subjective ratings were derived has been enlarged. MOS ratings were collected from 524,340 comparisons made by 971 observers. The image resolution remains unchanged.



Figure 1. Reference images of the TID2008 and TID2013 databases [21].

Online crowdsourcing for image assessment has enabled the creation of larger databases. One such large database, KADID-10k (Konstanz Artificially Distorted Image Quality Database) [23], contains 10,125 digital images with subjective quality scores (MOSs). It was developed and published by 2209 crowd workers. This database includes a limited selection of reference images (81) (Figure 2), a restricted number of artificial distortion types (25), and five levels for each distortion type. Recently, KADID-10k has become widely used for deep-learning models for image quality assessment [25]. The artificial distortions present in the KADID-10k database include spatial distortions, noise, blurs, and more. The image resolution remains unchanged.

PIPAL is a large IQA dataset, first introduced in 2020 by [24], that increased the number of reference images to 250. In fact, these are 288×288 fragments from images in the DIV2K and Flickr2K high-resolution image collections (Figure 3), with distortion types increased to 40 and distorted images increased to 29,000, and it contains 1,130,000 human ratings. In this image database, the Elo rating system was used to assign the Mean Opinion Scores (MOSs). Currently, the PIPAL dataset is used in many challenges as a benchmark for IQA algorithms.

The key information regarding the selected IQA benchmark databases is presented in Table 1.

Table 1. Comparison of the selected IQA databases.

Database	Year	No. of Ref.	No. of Dist.	Environment	No. of Dist. Images
TID2008	2008	25	17	lab	1700
TID2013	2013	25	24	lab	3000
KADID-10k	2019	81	25	crowdsourcing	10,125
PIPAL	2020	250	40	crowdsourcing	29,000

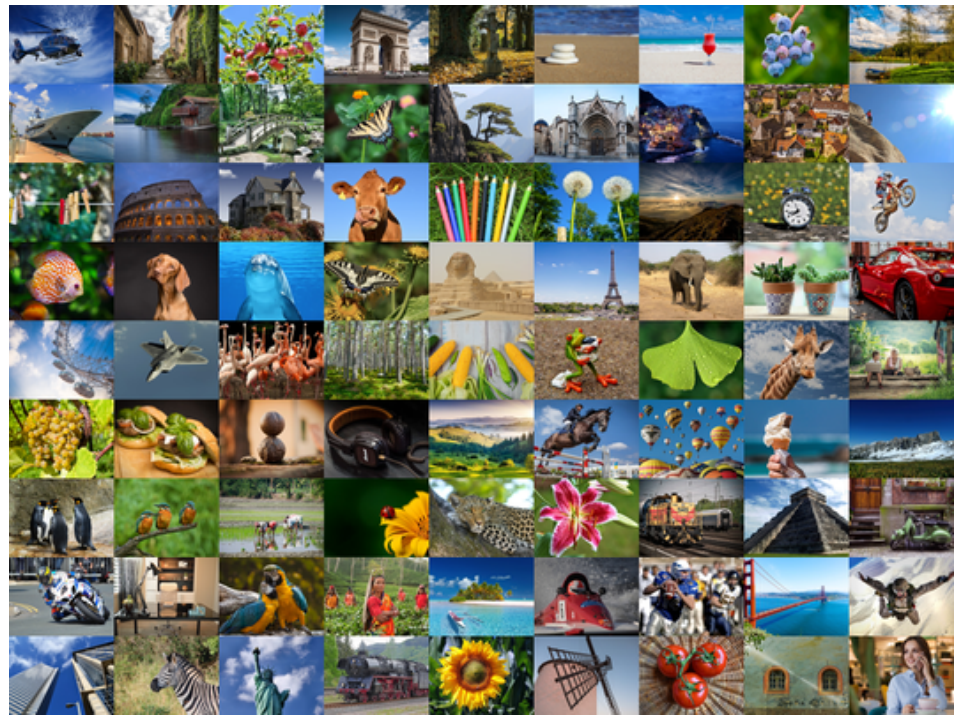


Figure 2. Reference images of the KADID-10k database [23].

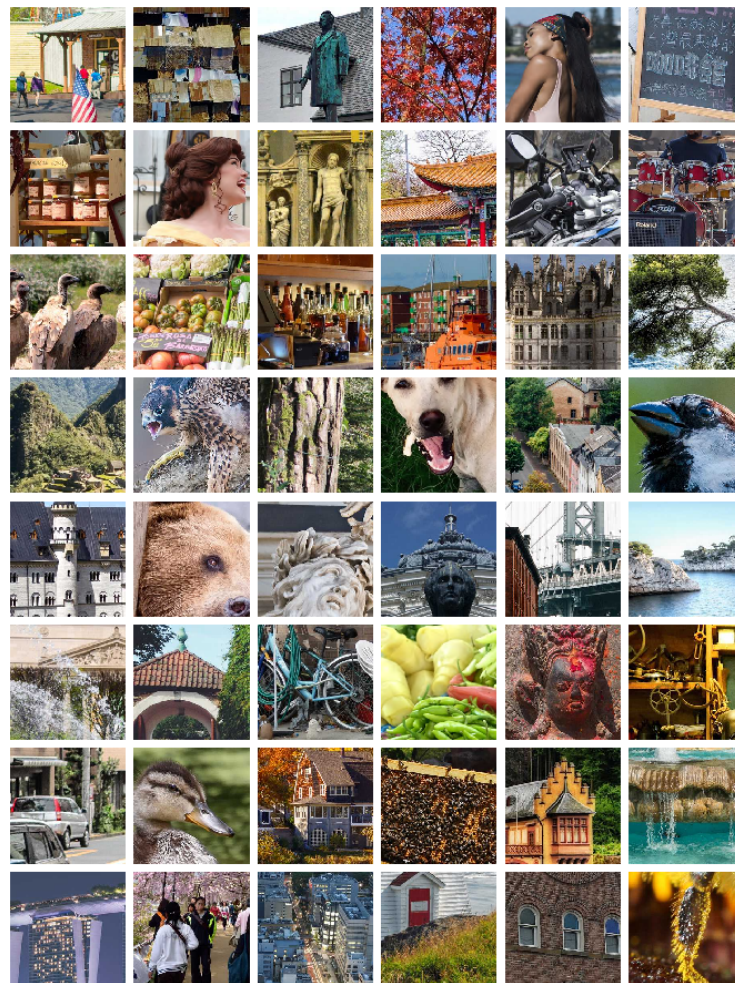


Figure 3. Examples of reference images from the PIPAL database [24].

3.2. Experimental Tests

The experimental study began with the determination of the PLCC, SROCC, and KROCC correlation coefficients, and RMSE values for eight selected highly correlated metrics. The results of these tests for the four study datasets are included in Table 2. The three highest correlation coefficients and the three lowest RMSE values are shown in different colors (the best results in red, the second results in green, and the third results in blue).

The three correlation coefficients and the RMSE value were aggregated into one score. A point scale with values from 1 to 8 was adopted, where the highest points were awarded to the highest correlation coefficients and the lowest RMSE values. This ranking is shown in Table 3, where the point values for each dataset are also summarized in the columns. The number of points determined the three component metrics for each dataset. The three highest scores are highlighted in bold.

The three metrics selected from the table served as the components M_1 , M_2 , and M_3 for the linear combination that determines the New Combined Metric, as defined in Formula (61).

Determining the NCM value requires calculating the α , β , and γ weights present in this formula. These weights are optimized in the Matlab environment using the *fmincon* function. In the optimization task, the PLCC linear correlation coefficient is maximized. The obtained values of the weights for each dataset are given in Table 4. Based on these weights and component metrics, the values of the combined NCM metric were determined. The results are shown in Table 5. The results for the three component metrics are shown in red, while green is used for the best score achieved by the combined NCM metric in each case.

Table 2. Values of correlation coefficients and RMSE for FR-IQA metrics.

Database	Metric	FSIMc [9]	MDSI [11]	HPSI [12]	VCGS [13]	SPSIM [14]	LGV [17]	SWLGV [17]	GMSD [19]
TID2008	PLCC	0.876	0.916	0.907	0.878	0.893	0.865	0.874	0.879
	SROCC	0.884	0.921	0.910	0.897	0.910	0.881	0.889	0.891
	KROCC	0.699	0.751	0.737	0.717	0.730	0.696	0.711	0.709
	RMSE	0.647	0.538	0.566	0.643	0.605	0.674	0.652	0.640
TID2013	PLCC	0.877	0.909	0.893	0.900	0.909	0.778	0.797	0.855
	SROCC	0.851	0.890	0.873	0.893	0.904	0.807	0.807	0.804
	KROCC	0.667	0.712	0.692	0.717	0.725	0.638	0.641	0.634
	RMSE	0.596	0.518	0.557	0.541	0.517	0.779	0.749	0.642
KADID-10k	PLCC	0.851	0.864	0.885	0.868	0.874	0.815	0.835	0.805
	SROCC	0.854	0.885	0.885	0.871	0.874	0.820	0.840	0.847
	KROCC	0.665	0.702	0.699	0.683	0.687	0.630	0.655	0.664
	RMSE	0.568	0.544	0.505	0.538	0.525	0.627	0.595	0.643
PIPAL	PLCC	0.615	0.598	0.641	0.554	0.578	0.529	0.543	0.629
	SROCC	0.589	0.585	0.589	0.534	0.562	0.519	0.536	0.583
	KROCC	0.416	0.408	0.417	0.370	0.391	0.359	0.372	0.414
	RMSE	0.104	0.106	0.101	0.110	0.108	0.112	0.111	0.103

In order to visualize the good quality of the proposed NCM index, scatter plots of the proposed eight metrics and the NCM for the tested bases are shown in Figures 4–7. The scatter plots and their fitted curves show that the proposed combined NCM metric closely matches the MOS estimates for each of the databases.

A study of computation times for the considered FR-IQA metrics was also conducted. The average computation times for each of the databases are shown in Table 6. The three fastest metrics are highlighted in bold. The NCM computation time, which is not

included in Table 6, is approximately equal to the sum of the computation times of its component metrics.

Table 3. Ranking of FR-IQA metrics.

Database	Metric	FSIMc [9]	MDSI [11]	HPSI [12]	VCGS [13]	SPSIM [14]	LGV [17]	SWLGV [17]	GMSD [19]
TID2008	PLCC	3	8	7	4	6	1	2	5
	SROCC	2	8	6	5	7	1	3	4
	KROCC	2	8	7	5	6	1	4	3
	RMSE	3	8	7	4	6	1	2	5
	TOTAL	10	32	27	18	25	4	11	17
TID2013	PLCC	4	7	5	6	8	1	2	3
	SROCC	4	6	5	7	8	3	2	1
	KROCC	4	6	5	7	8	2	3	1
	RMSE	4	7	5	6	8	1	2	3
	TOTAL	16	26	20	26	32	7	9	8
KADID-10k	PLCC	4	5	8	6	7	2	3	1
	SROCC	4	8	7	5	6	1	2	3
	KROCC	4	8	7	5	6	1	2	3
	RMSE	4	5	8	6	7	2	3	1
	TOTAL	16	26	30	22	26	6	10	8
PIPAL	PLCC	6	5	8	3	4	1	2	7
	SROCC	8	6	7	2	4	1	3	5
	KROCC	7	5	8	2	4	1	3	6
	RMSE	6	5	8	3	4	1	2	7
	TOTAL	27	21	31	10	16	4	10	25

Table 4. Optimized weight values for the NCM metric.

Criterion: Database	Weight	Three Best Value	Metrics	Three Fast Value	Metrics
TID2008	α	0.578	MDSI [11]	0.618	MDSI [11]
	β	0.285	HPSI [12]	0.372	HPSI [12]
	γ	0.136	SPSIM [14]	0.010	GMSD [19]
TID2013	α	0.459	MDSI [11]	0.680	MDSI [11]
	β	0.086	VCGS [13]	0.310	HPSI [12]
	γ	0.455	SPSIM [14]	0.010	GMSD [19]
KADID-10k	α	0.386	MDSI [11]	0.342	MDSI [11]
	β	0.404	HPSI [12]	0.486	HPSI [12]
	γ	0.210	SPSIM [14]	0.172	GMSD [19]
PIPAL	α	0.320	FSIMc [9]	0.317	MDSI [11]
	β	0.507	HPSI [12]	0.584	HPSI [12]
	γ	0.174	GMSD [19]	0.099	GMSD [19]

Table 5. Results of NCM metric for the three best component metrics.

Db.	Met.	FSIMc [9]	MDSI [11]	HPSI [12]	VCGS [13]	SPSIM [14]	LGV [17]	SWLGV [17]	GMSD [19]	NCM M_1, M_2, M_3
TID2008	PLCC	0.876	0.916	0.907	0.878	0.893	0.865	0.874	0.879	0.924 SPSIM [14]
	SROCC	0.884	0.921	0.910	0.897	0.910	0.881	0.889	0.891	0.924 MDSI [11]
	KROCC	0.699	0.751	0.737	0.717	0.730	0.696	0.711	0.709	0.759 HPSI [12]
	RMSE	0.647	0.538	0.566	0.643	0.605	0.674	0.652	0.640	0.514
TID2013	PLCC	0.877	0.909	0.893	0.900	0.909	0.778	0.797	0.855	0.922 SPSIM [14]
	SROCC	0.851	0.890	0.873	0.893	0.904	0.807	0.807	0.804	0.906 MDSI [11]
	KROCC	0.667	0.712	0.692	0.717	0.725	0.638	0.641	0.634	0.732 VCGS [13]
	RMSE	0.596	0.518	0.557	0.541	0.517	0.779	0.749	0.642	0.481
KADID-10k	PLCC	0.851	0.864	0.885	0.868	0.874	0.815	0.835	0.805	0.896 SPSIM [14]
	SROCC	0.854	0.885	0.885	0.871	0.874	0.820	0.840	0.847	0.897 MDSI [11]
	KROCC	0.665	0.702	0.699	0.683	0.687	0.630	0.655	0.664	0.717 HPSI [12]
	RMSE	0.568	0.544	0.505	0.538	0.525	0.627	0.595	0.643	0.480
PIPAL	PLCC	0.615	0.598	0.641	0.554	0.577	0.529	0.543	0.629	0.653 GMSD [19]
	SROCC	0.589	0.585	0.589	0.534	0.562	0.519	0.536	0.583	0.608 FSIMc [9]
	KROCC	0.416	0.408	0.417	0.370	0.391	0.359	0.372	0.414	0.432 HPSI [12]
	RMSE	0.104	0.106	0.101	0.110	0.108	0.112	0.111	0.103	0.100

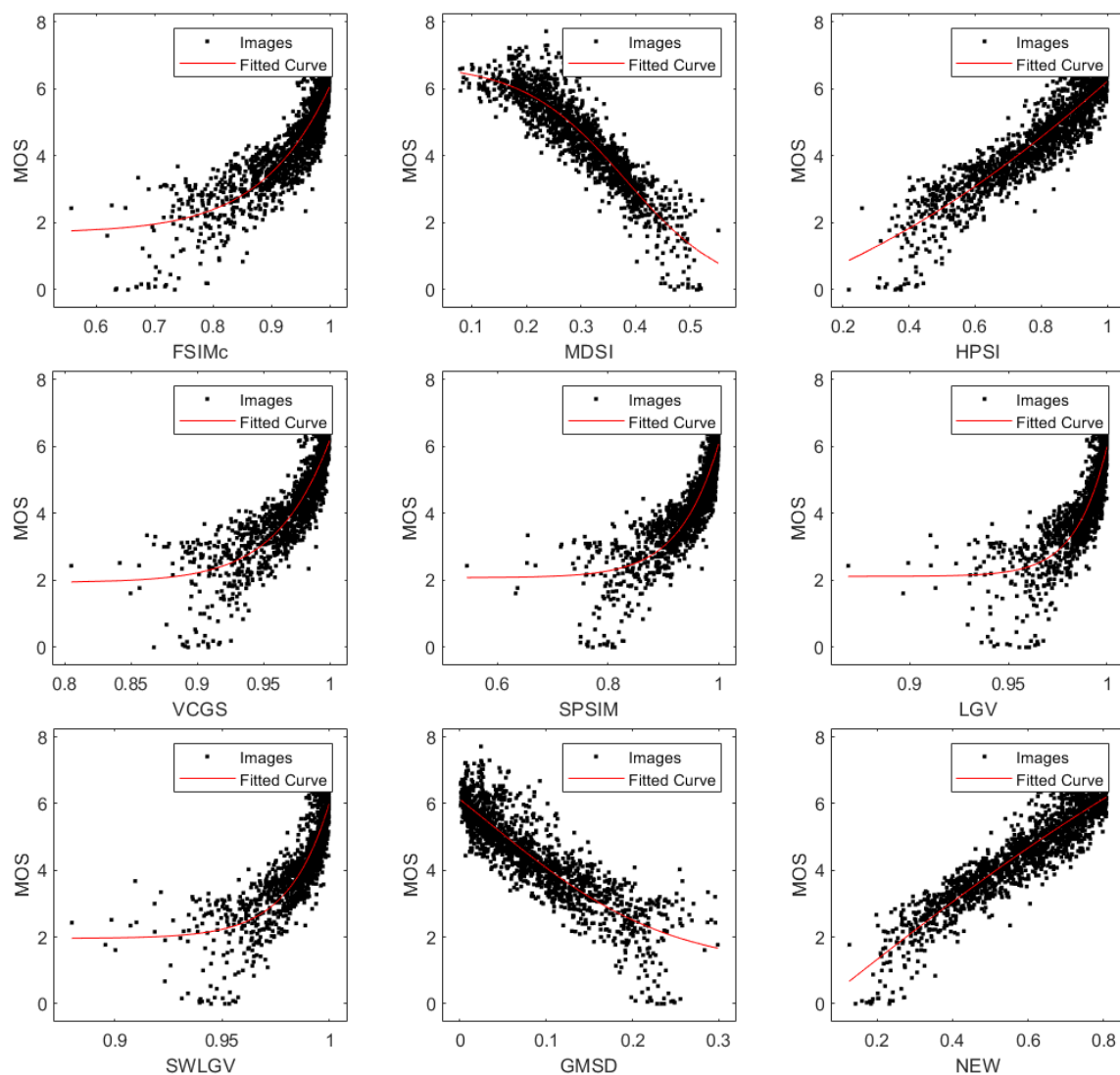


Figure 4. Scatter plots of subjective MOS against IQA metrics obtained from the TID2008 database.

Table 6. Computation times (s) for IQA metrics.

Database	FSIMc [9]	MSDI [11]	HPSI [12]	VCGS [13]	SPSIM [14]	LGV [17]	SWLGV [17]	GMSD [19]
TID2008	0.089	0.014	0.028	0.212	0.128	0.435	5.986	0.019
TID2013	0.104	0.016	0.037	0.214	0.113	0.339	6.130	0.019
KADID-10k	0.136	0.026	0.046	0.299	0.143	0.451	5.890	0.028
PIPAL	0.172	0.011	0.012	0.091	0.176	0.619	2.628	0.009

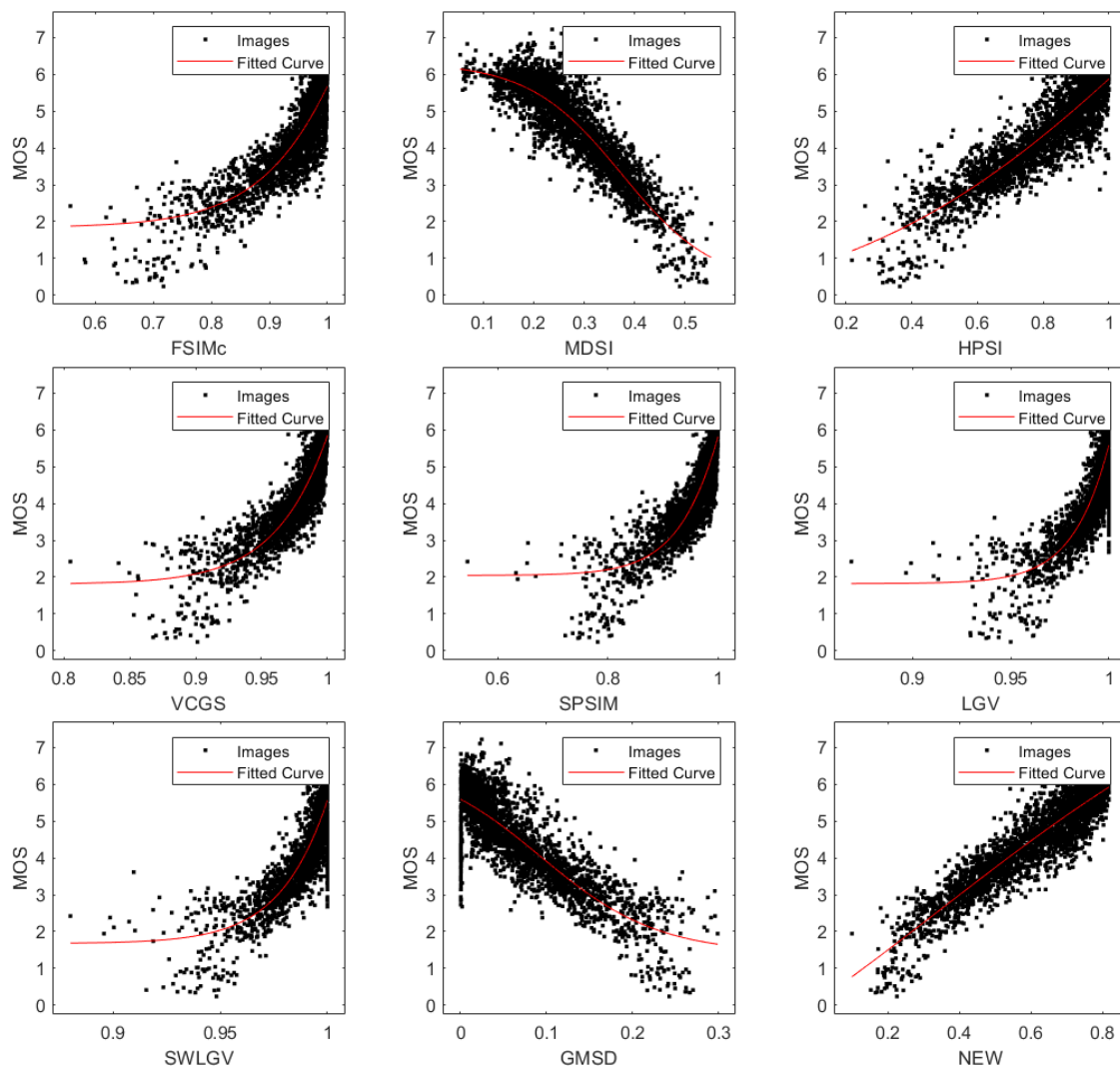


Figure 5. Scatter plots of subjective MOS against IQA metrics obtained from the TID2013 database.

For the three fastest FR-IQA metrics highlighted in red (see Table 7), a linear combination was formed by redetermining the optimal α , β , and γ weights (see Table 5) from the perspective of PLCC maximization. The resulting NCM metric using the three fastest metrics achieved the best performance, as marked in green in Table 7.

The study was conducted in the MATLAB R2024a programming environment on a computer with the specifications provided in Table 8.

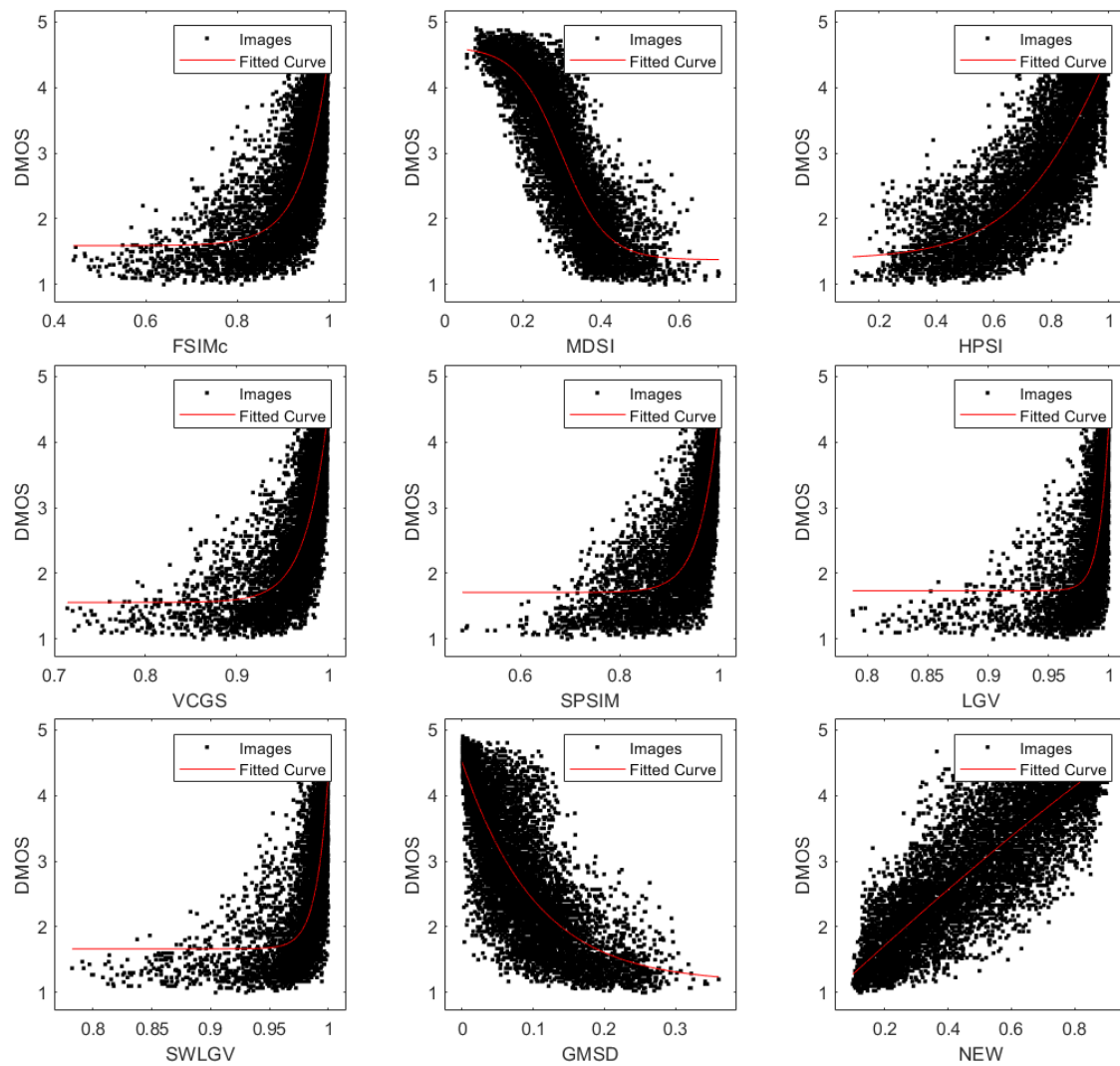


Figure 6. Scatter plots of subjective MOS against IQA metrics obtained from the KADID-10k database.

Table 7. Results of the NCM metric for the three fastest component metrics.

Db.	Met.	FSIMc [9]	MDSI [11]	HPSI [12]	VCGS [13]	SPSIM [14]	LGV [17]	SWLGV [17]	GMSD [19]	NCM M_1, M_2, M_3
TID2008	PLCC	0.876	0.916	0.907	0.878	0.893	0.865	0.874	0.879	GMSD
	SROCC	0.884	0.921	0.910	0.897	0.910	0.881	0.889	0.891	MDSI
	KROCC	0.699	0.751	0.737	0.717	0.730	0.696	0.711	0.709	HPSI
	RMSE	0.647	0.538	0.566	0.643	0.605	0.674	0.652	0.640	
TID2013	PLCC	0.877	0.909	0.893	0.900	0.909	0.778	0.797	0.855	GMSD
	SROCC	0.851	0.890	0.873	0.893	0.904	0.807	0.807	0.804	MDSI
	KROCC	0.667	0.712	0.692	0.717	0.725	0.638	0.641	0.634	HPSI
	RMSE	0.596	0.518	0.557	0.541	0.517	0.779	0.749	0.642	
KADID-10k	PLCC	0.851	0.864	0.885	0.868	0.874	0.815	0.835	0.805	GMSD
	SROCC	0.854	0.885	0.885	0.871	0.874	0.820	0.840	0.847	MDSI
	KROCC	0.665	0.702	0.699	0.683	0.687	0.630	0.655	0.664	HPSI
	RMSE	0.568	0.544	0.505	0.538	0.525	0.627	0.595	0.643	
PIPAL	PLCC	0.615	0.598	0.641	0.554	0.577	0.529	0.543	0.629	GMSD
	SROCC	0.589	0.585	0.589	0.534	0.562	0.519	0.536	0.583	MDSI
	KROCC	0.416	0.408	0.417	0.370	0.391	0.359	0.372	0.414	HPSI
	RMSE	0.104	0.106	0.101	0.110	0.108	0.112	0.111	0.103	

Table 8. Parameters of desktop computer used for experiments.

Processor	Intel(R) Core(TM) i5-7400 CPU @ 3.00 GHz (4 cores)
RAM	32 GB
OS	Windows 10
Env.	Matlab 2024a

The best results obtained using the NCM were additionally compared with those of other combined metrics presented in the literature [26,27]. The comparison is highlighted in bold in Table 9. The authors of [26] used combined metrics (MFMOGP3, MFMOGP4) based on the additive combination of component metrics, ranging from 8 to 10 metrics. In [27], combined metrics (OFIQA) based on the product form, involving between 4 and 17 factor-metrics, were proposed. For Table 9, we selected the best results from both of the above-mentioned works. For the TID2008 database, the results are comparable, while for the TID2013 database the proposed NCM index achieves the highest correlation coefficients. We conducted the comparison on the TID2008 and TID2013 databases, as both older works on combined metrics did not consider newer databases (KADID-10k, PIPAL).

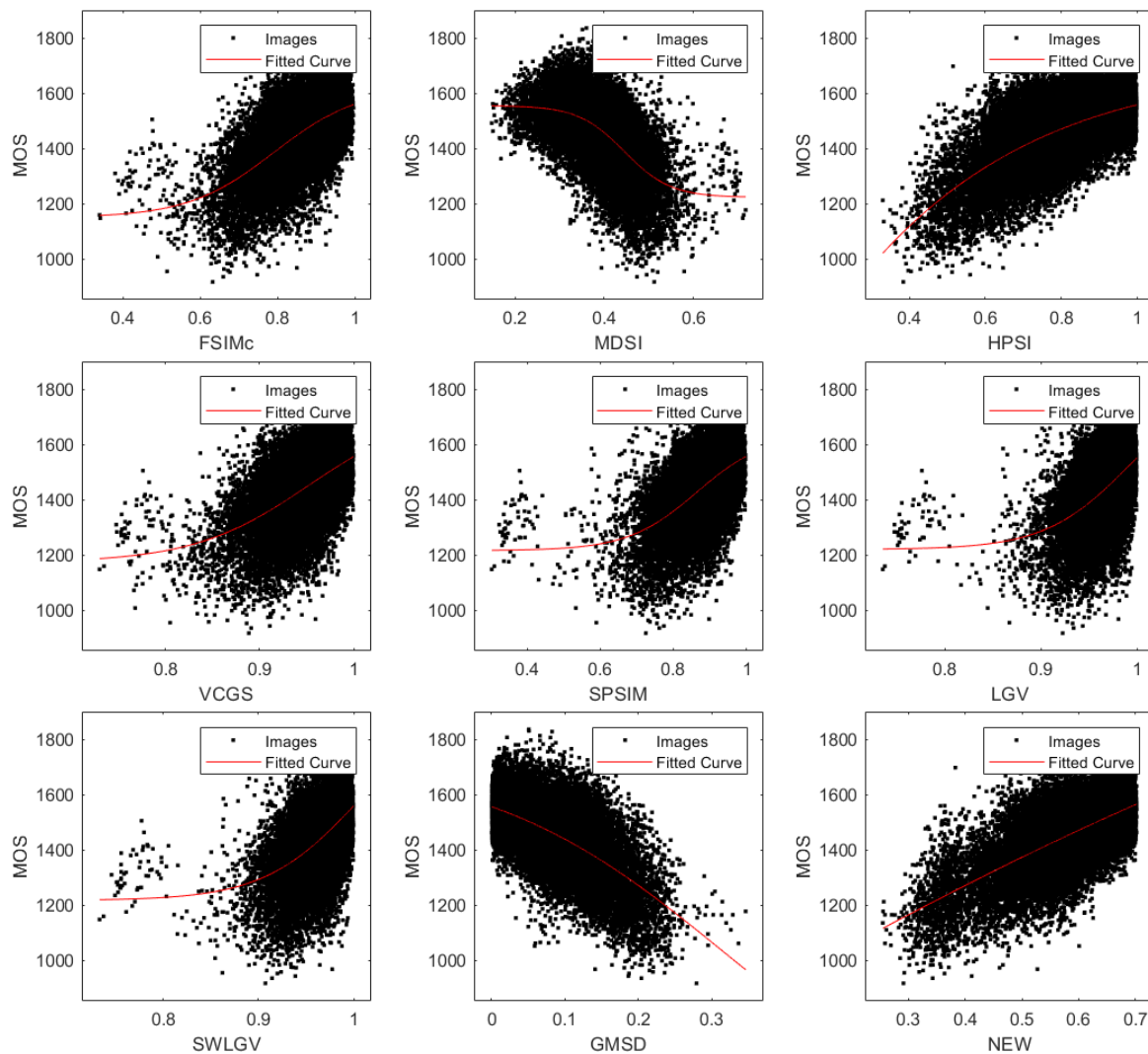
**Figure 7.** Scatter plots of subjective MOS against IQA metrics obtained from the PIPAL database.

Table 9. Comparison of the proposed NCM metric with other combined metrics.

Database	Metric	MFMOGP3 [26]	MFMOGP4 [26]	OFIQA [27]	NCM
TID2008	PLCC	0.925	0.902	0.910	0.922
	SROCC	0.923	0.911	0.915	0.923
	KROCC	0.757	0.727	0.738	0.757
	RMSE	0.511	0.580	0.557	0.519
TID2013	PLCC	0.883	0.914	0.906	0.922
	SROCC	0.868	0.902	0.890	0.923
	KROCC	0.688	0.725	0.713	0.757
	RMSE	0.581	0.503	0.526	0.519

4. Conclusions

From the existing literature on FR-IQA metrics, it is evident that there is no single metric that significantly outperforms the others. Therefore, the idea of creating a linear or nonlinear combination of several top metrics has emerged. The proposed approach opted for an additive combination of the three metrics with the highest correlation coefficients and the lowest RMSE. The resulting combined NCM metric was based on component metrics that depended on the selected database. NCM achieved the best results among all tested metrics across all tested databases. In addition, a case was examined where the three fastest metrics, i.e., MDSI, HPSI, and GMSD, were selected as components. The combined metric obtained in this case also achieved the best results compared to all the tested metrics. Potential extensions of the proposed approach include replacing the linear combination of metrics to their nonlinear combination, exploring alternative methods for optimizing weight selection, and more.

Author Contributions: Conceptualization, M.F. and H.P.; methodology, M.F.; software, L.M.; validation, M.F. and H.P.; investigation, M.F. and L.M.; resources, M.F.; data curation, M.F.; writing—original draft preparation, M.F.; writing—review and editing, M.F.; visualization, M.F.; supervision, H.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Polish Ministry for Science and Education under internal grant 02/070/BK_24/0055 for the Department of Data Science and Engineering, Silesian University of Technology, Gliwice, Poland.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IQA	Image Quality Assessment
FR-IQA	Full-Reference Image Quality Assessment
MOS	Mean Opinion Score
FSIMc	Feature SIMilarity (color version)
MDSI	Mean Deviation Similarity Index
HPSI	Haar wavelet Perceptual Similarity Index
VCGS	Visual saliency with Color appearance and Gradient Similarity
SPSIM	SuperPixel SIMilarity
LGV	Local Global Variation
SWLGV	Saliency Weighted Local Global Variation
GMSD	Gradient Magnitude Similarity Deviation
NCM	New Combined Metric
PSNR	Peak Signal-to-Noise Ratio
VSNR	Visual Signal-to-Noise Ratio

VIF	Visual Information Fidelity index
MS-SSIM	Multi Scale Structural SIMilarity index
PLCC	Pearson Linear Correlation Coefficient
SROCC	Spearman Rank Order Correlation Coefficient
KROCC	Kendall Rank Order Correlation Coefficient
RMSE	Root Mean Squared Error
SVD	Singular Value Decomposition
HVS	Human Visual System
DMOS	Differential Mean Opinion Score
TID	Tampere Image Database
KADID-10k	Konstanz Artificially Distorted Image quality Database
PIPAL	Perceptual Image Processing ALgorithms database
WFSIMc	Weighted FSIM (color version) index
RFSIM	Riesz-transform-based Feature SIMilarity index
CQM	Combined Quality Metric
CISI	Combined Image Similarity Index
LCSIM	Linearly Combined Similarity Measures

References

- Liu, M.; Yang, X. A new image quality approach based on decision fusion. In Proceedings of the 2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery, Jinan, China, 18–20 October 2008; Volume 4, pp. 10–14.
- Okarma, K. Combined full-reference image quality metric linearly correlated with subjective assessment. In Proceedings of the Artificial Intelligence and Soft Computing: 10th International Conference, ICAISC 2010, Zakopane, Poland, 13–17 June 2010; Part I 10; Springer: Berlin/Heidelberg, Germany, 2010; pp. 539–546.
- Okarma, K. Combined image similarity index. *Opt. Rev.* **2012**, *19*, 349–354. [\[CrossRef\]](#)
- Okarma, K. Extended hybrid image similarity–combined full-reference image quality metric linearly correlated with subjective scores. *Elektron. Ir Elektrotech.* **2013**, *19*, 129–132. [\[CrossRef\]](#)
- Oszust, M. Full-reference image quality assessment with linear combination of genetically selected quality measures. *PLoS ONE* **2016**, *11*, e0158333. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lukin, V.; Ponomarenko, N.; Ieremeiev, O.; Egiazarian, K.; 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, 9–12 February 2015; Volume 9394, pp. 172–183.
- 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. [\[CrossRef\]](#) [\[PubMed\]](#)
- Varga, D. A combined full-reference image quality assessment method based on convolutional activation maps. *Algorithms* **2020**, *13*, 313. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Kovesi, P. Image features from phase congruency. *Videre J. Comput. Vis. Res.* **1999**, *1*, 1–26.
- Nafchi, H.Z.; Shahkolaei, A.; Hedjam, R.; Cheriet, M. Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator. *IEEE Access* **2016**, *4*, 5579–5590. [\[CrossRef\]](#)
- Reisenhofer, R.; Bosse, S.; Kutyniok, G.; Wiegand, T. A Haar wavelet-based perceptual similarity index for image quality assessment. *Signal Process. Image Commun.* **2018**, *61*, 33–43. [\[CrossRef\]](#)
- Shi, C.; Lin, Y. Full Reference Image Quality Assessment Based on Visual Saliency with Color Appearance and Gradient Similarity. *IEEE Access* **2020**, *8*, 97310–97320. [\[CrossRef\]](#)
- Sun, W.; Liao, Q.; Xue, J.H.; Zhou, F. SPSIM: A superpixel-based similarity index for full-reference image quality assessment. *IEEE Trans. Image Process.* **2018**, *27*, 4232–4244. [\[CrossRef\]](#) [\[PubMed\]](#)
- Stutz, D.; Hermans, A.; Leibe, B. Superpixels: An evaluation of the state-of-the-art. *Comput. Vis. Image Underst.* **2018**, *166*, 1–27. [\[CrossRef\]](#)
- Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [\[CrossRef\]](#) [\[PubMed\]](#)
- Varga, D. Full-Reference Image Quality Assessment Based on Grünwald–Letnikov Derivative, Image Gradients, and Visual Saliency. *Electronics* **2022**, *11*, 559. [\[CrossRef\]](#)
- Imamoglu, N.; Lin, W.; Fang, Y. A saliency detection model using low-level features based on wavelet transform. *IEEE Trans. Multimed.* **2012**, *15*, 96–105. [\[CrossRef\]](#)
- Xue, W.; Zhang, L.; Mou, X.; Bovik, A.C. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Trans. Image Process.* **2013**, *23*, 684–695. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)

21. 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.
22. Ponomarenko, N.; Jin, L.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Astola, J.; Vozel, B.; Chehdi, K.; Carli, M.; Battisti, P.; et al. Image database TID2013: Peculiarities, results and perspectives. *Signal Process. Image Commun.* **2015**, *30*, 57–77. [[CrossRef](#)]
23. 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.
24. Gu, J.; Cai, H.; Chen, H.; Ye, X.; Ren, J.; Dong, C. PIPAL: A large-scale image quality assessment dataset for perceptual image restoration. In *European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2020; pp. 633–651.
25. Varga, D. Composition-preserving deep approach to full-reference image quality assessment. *Signal Image Video Process.* **2020**, *14*, 1265–1272. [[CrossRef](#)]
26. Merzougui, N.; Djerou, L. Multi-measures fusion based on multi-objective genetic programming, for full-reference image quality assessment. *arXiv* **2017**, arXiv:1801.06030.
27. Varga, D. An optimization-based family of predictive, fusion-based models for full-reference image quality assessment. *J. Imaging* **2023**, *9*, 116. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.