

Article

Combined No-Reference Image Quality Metrics for Visual Quality Assessment Optimized for Remote Sensing Images

Andrii Rubel ¹, Oleg Ieremeiev ¹, Vladimir Lukin ¹, Jarosław Fastowicz ² and Krzysztof Okarma ^{2,*}

¹ Department of Information and Communication Technologies, National Aerospace University, 61070 Kharkiv, Ukraine; a.rubel@khai.edu (A.R.); o.ieremeiev@khai.edu (O.I.); lukin@ai.kharkov.com (V.L.)
² Department of Signal Processing and Multimedia Engineering, West Pomeranian University of Technology in Szczecin, 70-313 Szczecin, Poland; jfastowicz@zut.edu.pl
* Correspondence: okarma@zut.edu.pl

Abstract: No-reference image quality assessment is one of the most demanding areas of image analysis for many applications where the results of the analysis should be strongly correlated with the quality of an input image and the corresponding reference image is unavailable. One of the examples might be remote sensing since the transmission of such obtained images often requires the use of lossy compression and they are often distorted, e.g., by the presence of noise and blur. Since the practical usefulness of acquired and/or preprocessed images is directly related to their quality, there is a need for the development of reliable and adequate no-reference metrics that do not need any reference images. As the performance and universality of many existing metrics are quite limited, one of the possible solutions is the design and application of combined metrics. Several possible approaches to their composition have been previously proposed and successfully used for full-reference metrics. In the paper, three possible approaches to the development and optimization of no-reference combined metrics are investigated and verified for the dataset of images containing distortions typical for remote sensing. The proposed approach leads to good results, significantly improving the correlation of the obtained results with subjective quality scores.

Keywords: image quality assessment; visual quality metrics; neural networks; combined metrics



Citation: Rubel, A.; Ieremeiev, O.; Lukin, V.; Fastowicz, J.; Okarma, K. Combined No-Reference Image Quality Metrics for Visual Quality Assessment Optimized for Remote Sensing Images. *Appl. Sci.* **2022**, *12*, 1986. <https://doi.org/10.3390/app12041986>

Academic Editor: Francesco Bianconi

Received: 22 December 2021

Accepted: 11 February 2022

Published: 14 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Modern remote sensing (RS) systems produce an enormous number of images that are later used for many valuable applications such as ecological monitoring, agriculture, and urban planning [1–3]. It is often supposed that all RS images are of high quality. However, this is not so due to many reasons. Noise and other distortions can be present in acquired data due to the RS sensor principle of operation such as speckle in radar images [4], wave absorption in specific bands as in junk channels of hyperspectral data [5], bad conditions of imaging [6], and imaging or communication system failures [7].

Then, an important task is to estimate an image quality for an original (acquired) image or a processed (e.g., filtered) image [6,8–11]. There are practical situations when full-reference metrics [12,13] can be employed for this purpose [8]. This happens when there is an image that can be considered as “pristine” (reference) and, after processing (e.g., lossy compression), one has the corresponding distorted image that should be compared to the reference one using a certain quality metric where both traditional (such as Mean Square Error—MSE or Peak Signal-to-Noise Ratio—PSNR) or visual quality (e.g., Structural Similarity—SSIM [14,15]) metrics can be applied.

Nevertheless, one can often meet situations where full-reference metrics cannot be used since reference images are unavailable. Then, there are two options: either to try predicting the values of trustworthy full-reference metrics or apply no-reference or reduced-reference ones. Examples of the first option are given in the papers [16,17], where trained neural networks (NNs) are applied to characterize the quality of images, including the

quality of synthetic aperture radar (SAR) images, by predicting the values of different image quality metrics. Such an approach is based on the known fact that full-reference metrics characterize the quality of distorted images better than no-reference ones where accuracy is usually described in terms of conventional or rank correlation of a metric and Mean Opinion Score (MOS) for databases of distorted images [12]. However, the values of full-reference metrics are anyway predicted with errors and the already proposed solutions are not universal.

Because of this, an alternative option is often used, i.e., to apply no-reference (NR) metrics [18–23]. NR metrics designed earlier (e.g., [19,20]) have mainly addressed particular types of distortions such as blur, noise, and blockiness caused by lossy JPEG compression with application to optical images. Being applied to a wide set of possible distortions such as in the database TID2013, such NR metrics fail to provide good results [24]. Even for the best metrics, such as IL-NIQE [25] and NIQE [26], the values of Spearman Rank Order Correlation Coefficient (SROCC) between the metric and Mean Opinion Score values do not exceed 0.5 for the database TID2013 [27], although SROCC values are larger for simpler databases as LIVE [28].

On the contrary, with the development of modern technologies and machine learning-based approaches, image quality assessment (IQA) methods have become more complicated and powerful. Support vector machines, neural networks, and other approaches [18,22–24] have been employed to improve the performance of NR metrics used for IQA for conventional (grayscale and color) images as well as RS images [17,22,23].

The authors of [18] have designed an efficient classifier of distortion types based on support vector machine (SVM) and a set of input features. They have tested it for several available databases. However, they have considered a limited set of three distortion types. Analysis of the 21 NR metrics for Quickbird RS images has been carried out in [22], where the edge intensity metric has been recommended for practical use. However, the study has been performed for only three types of distortions (average filtering, Gaussian white noise, and linear motion degradation). In [23], multiple statistics features sensitive to image quality have been extracted from hyperspectral images and used for the IQA. A particular application of hyperspectral data restoration and super-resolution has been studied.

One way of improvement of metrics performance has been found expedient recently—to design and optimize the so-called combined metrics [11,24,29,30]. The term “combined” means that several “elementary” (quite simple) metrics are calculated sequentially or in parallel, and then united (combined) in a certain manner. In this way, particular drawbacks and weaknesses of elementary metrics can be eliminated while their positive features can be sufficiently enforced with a synergetic effect. Combining or aggregation can be performed in different ways: using “robust” processing with the preliminary fitting of metric to MOS dependencies [24], by simple calculation of a weighted sum or product of elementary metrics [29,30], or employing a trained NN. Although some promising results have been obtained in [24], the analysis for particular types of distortions typical for RS images has not been presented. The appropriateness of such an approach has also been confirmed for video quality assessment by the successful development of the Video Multi-method Assessment Fusion (VMAF) metric by Netflix, validated by RealNetworks [31].

In all cases, one needs certain databases or their parts to carry out training or to employ optimization of the combined metrics as well as to verify the results. Unfortunately, there are no special commonly accepted databases for RS images with MOS determined for them. There are several reasons behind this. First, there are RS images of different nature (optical, infrared, and radar) and characteristics (noise type and intensity, number of channels, and spatial resolution). Second, while it is possible (although not easy) to attract quite many people to evaluate the quality of standard color images, it is considerably more difficult to find enough qualified experts to assess the quality of RS images.

In [11], a slightly “artificial” way out has been found since a subset of images from the database TID2013 [27] has been selected, having distortions of type and intensity inherent for three-channel RS images. Using these selected images (subsets Noise and Actual in

TID2013), we have trained the NNs that used different numbers of elementary full-reference metrics as NN inputs. As the result, we have obtained the SROCC of about 0.965 for the best structures and configurations of the NN-based metrics for the subset Noise + Actual (NA). Keeping in mind these results, we have decided to carry out a similar attempt and study for NR IQA metrics. Thus, the main novelty of this paper consists in the design of general-purpose combined NR metrics using NN and weighted sum/product-based approaches of elementary metric aggregation with application to distortions typical for three-channel RS images.

The paper structure is as follows: in Section 2 the TID2013 database, NR metrics, earlier results, and the applied methodology are discussed, Section 3 focuses on the presentation of experimental results with their analysis. A brief discussion is given in Section 4.

2. Materials and Methods

2.1. Overview of the TID2013 Dataset and Earlier Results in the Combined Metric Design

To design and verify a new metric, one needs a database or several databases that contain images of a needed type corrupted by distortions with type and intensity inherent for images under interest. Besides, these databases should provide MOS or Differential Mean Opinion Score (DMOS) values for all images. In this sense, there are some problems in the design and verification of visual quality metrics for RS images. First, there is a very limited number of available databases of RS images, and they have a limited number of distortion types and/or are intended for other purposes [32,33]. Second, remote sensing images of a certain type can be intended for a particular purpose, therefore IQA results can be in improper agreement with this purpose and criteria used. For example, it is currently not fully clear how traditional criteria used in lossy compression relate to criteria characterizing object detection, classification, and segmentation of compressed images. Third, even if RS images have three components, there are several variants of their visualization as RGB color images. If the number of components is larger than three, there are numerous ways to represent RS data in pseudo-colors and then a correspondence between the assessed quality of visualized images and real values of RS data is not clear.

Because of this, similarly to [11], the analysis is restricted to three-channel RS images assuming that they are represented as RGB color images. Consideration of more complex cases of more channels (components) of RS images is out of the scope of this paper.

Recall that the synthesis and analysis of the full-reference and no-reference quality metrics for RS images using the database TID2013 has already been carried out [11]. The reasons for this may be summarized as follows:

- The TID2013 dataset contains 25 reference images and images with 24 types and 5 levels of distortions where many distortion types take place in the practice of remote sensing;
- These types of distortions are concentrated in two subsets, namely Noise and Actual, that can be processed and analyzed separately (see the details below);
- MOS values have been obtained for all distorted images in TID2013 including the aforementioned subsets where a larger number of participants (volunteers) have been attracted to experiments (this is important since MOS has to be estimated accurately enough to minimize the negative influence of possible inaccuracy on the SROCC calculation and comparison of metrics performance).

The whole TID2013 dataset contains the following types of distortions: Additive Gaussian noise (#1), Additive noise in color components (#2), Spatially correlated noise (#3), Masked noise (#4), High-frequency noise (#5), Impulse noise (#6), Quantization noise (#7), Gaussian blur (#8), Image denoising (#9), JPEG compression (#10), JPEG2000 compression (#11), JPEG transmission errors (#12), JPEG2000 transmission errors (#13), Non-eccentricity pattern noise (#14), Local block-wise distortions of different intensity (#15), Mean shift (intensity shift) (#16), Contrast change (#17), Change of color saturation (#18), Multiplicative Gaussian noise (#19), Comfort noise (#20), Lossy compression of noisy images (#21), Image

color quantization with dither (#22), Chromatic aberrations (#23), and Sparse sampling and reconstruction (#24).

The distortions ##14, 15, 16, 18, 22, and 23 are atypical for RS applications, other types can be met in multichannel imaging. In this sense, two subsets attract special attention. The subset Noise deals with images having the following types of distortions: Additive Gaussian noise (#1), Additive noise in color components (#2), Spatially correlated noise (#3), Masked noise (#4), High-frequency noise (#5), Impulse noise (#6), Quantization noise (#7), Gaussian blur (#8), Image denoising (#9), Multiplicative Gaussian noise (#19), and Lossy compression of noisy images (#21). Another subset (Actual) contains images with aforementioned distortions ##1, 3, 4, 5, 6, 8, 9, 19, and 21, as well as #10 (JPEG compression) and #11 (JPEG2000 compression). Some examples of when these types of distortion might happen in remote sensing are discussed in our earlier paper [11]. Thus, being united, the two subsets can represent a variety of distortions typical for remote sensing.

It is worth noting here that some other types of distortions may also be present in RS images. For example, there may be multiple distortions such as blur + noise or blur + blockiness. However, currently, the analysis is restricted by the availability of the databases and appropriate subsets.

Previous experiments with the design of the combined FR and NR metrics [11,24,29,30] show the following:

- SROCC for a given metric, elementary or combined, depends not only on a metric but also on the database whereas for databases with a smaller number and/or more typical distortions SROCC is usually considerably higher (because of this, the metric performance for the old LIVE database is commonly much better than for the other, more complicated, databases);
- SROCC values for metrics for such subsets of TID2013 as Noise or Actual are usually larger than for the entire database;
- Combined metrics, especially NN-based, can provide better results than the best elementary metric; meanwhile, for a given number of elementary metrics N , it does not mean that the best N elementary metrics have to be combined to produce the best NN-based metrics; the elementary metrics being complementary to each other constitute the best solutions.

Here it is worth recalling that there are three main types of combining elementary metrics:

- $CM = R(EMfMOS(n), n = 1, \dots, N)$, where R means a robust operator (e.g., a sample median or α -trimmed mean) applied to a set of elementary metrics after fitting to MOS (EMfMOS), N is the number of elementary metrics which is usually quite small (e.g., equal to 5); the advantage is that the method is simple; the drawback is that metric to MOS fitting is needed;
- $CM = WSUM(EM(n), n = 1, \dots, N)$ or $CM = WPROD(EM(n), n = 1, \dots, N)$ where $WSUM$ and $WPROD$ denote the weighted sum or weighted product of N elementary metrics with weights optimized to provide the highest correlation with subjective scores for a given dataset; the advantages are that good results may be obtained for a relatively small number of elementary metrics and their combining is very simple and fitting (in general) is not needed; the drawback is that the obtained combined metrics are usually less efficient than the NN-based ones;
- $CM = NN(EM(n), n = 1, \dots, N)$ where $NN()$ means that a neural network is applied to a set of input parameters where elementary metrics (without fitting) serve as inputs; the advantage is that the obtained metrics are usually the most efficient; the drawback is that there are several questions to be answered at the stage of metric design.

Because of the mentioned advantages and drawbacks, this paper concentrates on the two latter approaches to the design of combined metrics. In both cases, the provided performance depends upon many factors. The main two of them are the number of elementary metrics and the choice of the used elementary metrics. A general tendency is that a larger number of elementary metrics leads to better performance but only to a certain

limit. For the combined metrics presented as a weighted sum or product [29,30], it is often enough to have less than 8–10 metrics since a further increase of N does not lead to sufficient benefits in performance, making the combined metric more complicated. The same holds for NN-based metrics [11] where usually it is enough to apply 20–30 elementary metrics instead of 40–50 elementary metrics used as inputs. In this case, the Lasso method [34] allows restricting a set of elementary metrics to be employed [11] with simplifying the design and the final structure of the obtained NN-based combined metric.

An important role is also played by the optimization and training methods. For example, in [24] only a very limited number of weights was allowed in the design of product (multiplicative) type of combined metrics. However, an optimization procedure that should allow avoiding to get into local extrema is important as well. It is worth noting that for the NN-based combined metrics, it is practically impossible to use the convolutional NNs which are very popular nowadays [35]. The main reason is a limited number of distorted images to be used in training and verification. Additionally, a division of images into smaller patches and data augmentation cannot be conducted due to the availability of MOS values obtained only for whole images. Thus, simpler structures of the NNs have to be applied. However, even in this case, there is a wide variety of possible variants.

2.2. Brief Overview of the No-Reference Image Quality Assessment Methods

The no-reference objective IQA metrics are particularly interesting for many applications where “pristine” reference images without any distortions are unavailable. However, due to the lack of possibility of comparisons of the distorted images with them, each NR method should not only determine the amount and/or type of distortions in the way as highly correlated with their subjective perception as possible. Typically, such methods should also be able to detect the presence of some types of distortions without the knowledge of the original reference image. These requirements cause much lower universality of the NR metrics in comparison to the full-reference IQA methods as well as their significantly lower correlation with subjective quality evaluation results available in the IQA datasets, such as LIVE or TID2013. Therefore, many various general-purpose or more specialized NR metrics have been proposed by various researchers, and some of them are briefly presented below.

Some of the most widely known NR metrics for natural images, currently implemented i.a. in MATLAB environment are Naturalness Image Quality Evaluator (NIQE) [26], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [36], and Perception-based Image Quality Evaluator (PIQE) [37]. The first one is based on the measurement of the distances between the image features utilizing the natural scene statistics (NSS) for the assessed image and the same features obtained from an image database during model training with the use of a multivariate Gaussian model. Its extension, known as IL-NIQE [25], utilizes such models calculated for each image patch and the overall quality score is obtained by average pooling. Another, relatively well-known NR metric based on the NSS is the Blind Image Quality Index (BIQI) [38], being a two-step framework composed of 5 sub-indexes sensitive to distortion types present in the LIVE database, utilizing wavelet decomposition, generalized Gaussian distribution (GGD) and a classifier based on the support vector machine (SVM). The BLIINDS-2 [39] metric utilizes the NSS model of the discrete cosine transform (DCT) coefficients a simple Bayesian inference approach for quality prediction, whereas Distortion Identification-based Image Verity and Integrity Evaluation (DIIVINE) [40] does not use any distortion-specific models to extend its universality.

An interesting general-purpose NR objective metric, known as the COdebook Representation for No-reference Image Assessment (CORNIA), has been proposed by Ye et al. [41] utilizing unsupervised feature learning. Instead of the handcrafted features, raw-image-patches extracted from a set of unlabeled images have been used as local descriptors. Xue et al. [42] have proposed a quality-aware clustering (QAC) method to learn a set of cen-

troids for each quality level without the necessity of using the images scored by a human in learning.

Some examples of more specialized metrics, e.g., for sharpness evaluation might be ARISM [43] based on the autoregressive model parameters, Cumulative Probability of Blur Detection (CPBDM) [20], JNBM [19] based on the idea of just noticeable blur (JNB), or the blur metric proposed by Cr ete-Roffet et al. [44]. Some other examples may be the wavelet-based Fast Image SHarpness (FISH) [45] metric, as well as Perceptual Sharpness Index (PSI) [46]. A specialized metric designed to measure blocking effects and relative blur is known as WNJE [47], and an exemplary metric designed for blur assessment based on discrete orthogonal moments is known as BIBLE [48].

The use of High Order Statistics Aggregation (HOSA) has been proposed by Xu et al. [49], whereas Min et al. proposed the BPRI metric [50] based on the use of pseudo-reference image, consisting of three sub-indexes: PSS blockiness measure, LSSs sharpness measure, and LSSn noisiness measure. An interesting hybrid NR metric, called SISBLIM, designed also for evaluation of multiply distorted images, has been proposed by Gu et al. [51] in four versions (SISBLIM_SM, SISBLIM_SFB, SISBLIM_WM, and SISBLIM_WFB).

Since the topic of the paper is related to the methods of an efficient combination of existing metrics, a more detailed description of some other NR metrics used in experiments may be found in respective papers referred for each metric in Table 1, presenting their performance.

2.3. The Design of the No-Reference Combined Metrics for RS Images

Improvement of no-reference metrics is a demanding topic of research. Robust and stable results still pose a challenge and are required for many tasks. One of the most effective approaches to increase their performance is the combination of elementary metrics. The design of the NR combined metrics that would reflect the distortions met in RS images is based on the "Noise and Actual" (NA) subset from the TID2013 database, concerning 13 selected types of distortions listed in Section 2.1. This subset contains 1625 images out of 3000 present in the whole database (it is still more than included in some other commonly used datasets, such as LIVE or CSIQ [52]). Considering the combination based on the weighted product, the generalized formula of the combined metric may be presented as:

$$CM = \prod_{i=1}^N Q_i^{w_i}, \quad (1)$$

assuming the elementary metrics Q_i . The application of the second investigated approach utilizing the weighted sum expressed as:

$$CM^+ = \sum_{i=1}^N (a_i Q_i^{w_i}), \quad (2)$$

with additional weights a_i , increasing the flexibility of the designed combined metric, makes it possible to increase its correlation with subjective scores. Such an approach to metrics' combination, applied previously for full-reference metrics, has led to encouraging results [30]. In both cases, the values of all parameters (weights) for the selected set of elementary metrics are obtained as the results of optimization using the direct search method based on the Nelder–Mead simplex. For this purpose, MATLAB *fminsearch* function may be effectively used.

The experiments have been started by the calculation of the SROCC values of elementary metrics for the NA subset. As shown in Table 1, the best accuracy of visual quality estimation for the NA subset is provided by IL-NIQE [25], which value reaches 0.72 by SROCC. Due to the relatively small number of possible combinations, each pair of metrics has been subject to both types of combinations, according to Equations (1) and (2) with optimization of parameters using the SROCC as the criterion. The further choice of

elementary metrics for the combination using three or more of them has been based on the correlation of their subset: the five best combinations of two metrics have been selected as the basis and the third metric has been combined with them (all metrics have been checked and weights have been optimized). Then five best combinations of three metrics have been selected and each of the other elementary metrics has been checked as the potential fourth element, and so on. The reason for the choice of the five best combinations for each N is caused by the observation in previous experiments with full-reference metrics applied to multiply distorted images [30]. In some cases, the second or third “best” combination of N elementary metrics has led to better results in combination with the $(N + 1)$ -th metric.

Table 1. The absolute SROCC values obtained for the NR elementary metrics (the three best results are marked with bold font).

Metric	SROCC		Metric	SROCC	
	NA Subset	Whole TID2013		NA Subset	Whole TID2013
ARISM [43]	0.0736	0.1449	IL-NIQE [25]	0.7221	0.4921
ARISM _C [43]	0.0520	0.1380	JNBM [19]	0.1632	0.1412
BIBLE [48]	0.3331	0.2812	LPC_SI [53]	0.3196	0.3233
BIQI [38]	0.4999	0.4050	LPSI [54]	0.4695	0.3949
BLIINDS-2 [39]	0.4834	0.3946	MLV [55]	0.1755	0.2013
Blur Metric [44]	0.0393	0.0075	MSGF_PR [56]	0.3085	0.2437
BPRI [50]	0.2849	0.2289	NIQE [26]	0.4451	0.3132
PSS [50]	0.3334	0.0218	PIQE [37]	0.2473	0.1364
LSS _s [50]	0.0880	0.1138	NIQMC [57]	0.1141	0.1129
LSS _n [50]	0.1936	0.1675	NJQA [58]	0.0233	0.0997
BRISQUE [36]	0.4952	0.3673	NMC [59]	0.0485	0.0542
C-DIIVINE [60]	0.4773	0.3734	NR_PWN [21]	0.0259	0.0163
CORNIA [41]	0.4735	0.4352	OG-IQA [61]	0.3874	0.2761
CPBDM [20]	0.0948	0.1115	PSI [46]	0.1677	0.0940
DESIQUE [62]	0.1236	0.0691	QAC [42]	0.5034	0.3723
DIIVINE [40]	0.4456	0.3438	SDQI [63]	0.2254	0.2244
dipIQ [64]	0.2532	0.1395	DIQU [65]	0.2950	0.2401
FISH [45]	0.0219	0.0524	TCLT [66]	0.3151	0.2331
FISH_BB [45]	0.1551	0.1450	GM_LOG [67]	0.1556	0.1089
SISBLIM_SM [51]	0.4787	0.3178	HOSA [49]	0.5690	0.4705
SISBLIM_SFB [51]	0.4276	0.3363	SMETRIC [68]	0.0157	0.0969
SISBLIM_WM [51]	0.3992	0.2392	SSEQ [69]	0.4659	0.3410
SISBLIM_WFB [51]	0.3789	0.2929	WNJE [47]	0.3967	0.3018

An important element in the design of the combined metrics is also the calculation time. To verify the possible limitations some experiments have been conducted calculating some elementary metrics using the Intel i7 10gen laptop processor. Although the execution time of many metrics is below 1 s, it is worth mentioning some of the “slowest” algorithms, namely BLIINDS2 (nearly 16 s), NJQA (over 8 s), IL-NIQE (over 6 s), ARISM/ARISM_C (about 5 s), OG-IQA (over 4 s), or CORNIA (nearly 3 s). Nevertheless, since the processing time may vary, the precise measurement results have not been presented, limiting this thread to the indication of the most computationally demanding metrics verified using an exemplary computer for comparison purposes. The graphical illustration of the differences is presented in Figure 1 in logarithmic scale with respect to the fastest metric.

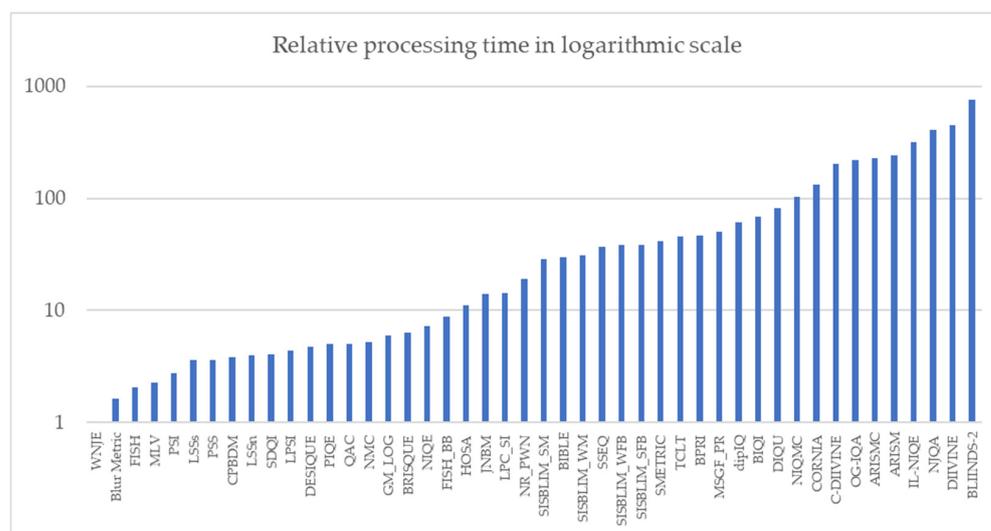


Figure 1. Illustration of the relative processing time of elementary metrics listed in Table 1 with respect to the fastest one (WNJE) presented in the logarithmic scale.

2.4. The Design of the Combined Metrics Based on Neural Networks

One of the main results obtained in earlier papers [11,24,70] concerning the use of neural networks is the confirmation of their high efficiency for the optimization of full-reference combined metrics. As the result of multiple parameters optimization, combined metrics based on neural networks provide several benefits. For a given image database they usually outperform any single elementary metric, and it is enough to use a simple NN model with several layers to achieve that result. Therefore, this stage is characterized by high computational performance, and the duration of computations is determined primarily by the slowest metric (in the case of parallel computations). In this paper, an approach tested previously on full-reference metrics has been applied to no-reference metrics and verified experimentally for the NA subset and the whole TID2013 database. Some preliminary results may also be found in [24], although only general-purpose metrics have been designed, demonstrating quite limited universality, given the results obtained for different considered datasets.

Since the use of more complex deep learning models [71,72] may not always be applicable for different image processing tasks, due to its limitations on the structure and size of the trained model and its computational complexity, the experiments are focused on the application of simpler NN models. Taking into account the distortions characteristic for RS images, some additional NR metrics, not considered previously in [24], have been included in calculations, increasing the possibilities of their choice and optimization.

Since the accuracy values of even the best NR metric (IL-NIQE) are quite low, a remarkable improvement may be expected using the combined NN metric, potentially better than using both approaches considered above (weighted product and weighted sum). Nevertheless, it would be difficult to directly compare the obtained combined metric with the results presented in [24], since the target datasets for the NN differ significantly. However, based on the results for 11 metrics and the fact that the NA subset has been included as a part of previous experiments, the expected improvement should reach the SROCC values at least 0.75–0.8 for a comparable number of metrics.

When designing and training a neural network, it is necessary to take into account some conditions that significantly affect its efficiency:

Input and output data of the neural network.

- The values of the visual quality metrics are used as inputs. The NN result should be an indicator corresponding to the human visual system (HVS). Therefore, the target data are the MOS estimates of the test image databases and the accuracy of the

approximation of the NN result to the MOS estimates determines its effectiveness. For the training and validation phases, the datasets are randomly subdivided into 70% and 30% subsets, respectively;

- The number of the applied metrics.
- One of the design goals is to keep computational complexity possibly low, so a smaller number of metrics is preferable. To analyze the influence of the number of input metrics on the overall efficiency, neural networks with different numbers up to over 40 metrics are considered;
- The choice of the applied metrics.
- Based on the research results for the reference metrics of visual quality in [11], the regression analysis approach using the Lasso method is quite effective. This method allows us to reduce the complexity of the network by assigning zero weights for non-essential features (elementary metrics in our case). By using different thresholds, we can define the “cut-off” levels for each of the possible metrics, thereby determining their importance in the resulting composite metric. It should be noted that a simple linear model is used here, and the selection of results will be approximate. Therefore, it is recommended to validate and carry out single replacements as needed;
- Type of the neural network.
- As the results in [72] have shown, the use of more complex modifications with non-linear dependencies between layers as compared to feed-forward networks leads to longer training but may lead to a noteworthy increase in the accuracy of the designed metric. At the same time, the results obtained for the best types—cascade and Elman networks—gave comparable efficiency. Therefore, in the paper, the feed-forward and cascade-forward neural networks are analyzed.

The Lasso algorithm has an important feature that should be taken into account during the calculations. It performs a sequential optimization of the weights, so its result depends on the starting point of reference. In the context of neural networks, this means that the order of the metrics in the input data is essential. Therefore, the training procedure was carried out in two stages.

At the first stage, 100 sequences have been randomly generated from the considered metrics. For the initial configuration, from 5 input metrics, neural networks have been calculated assuming the use of both types of neural networks and a different number of layers and neurons (20 training rounds with a randomly subdivided test set). As shown in previously obtained results [11], the efficiency is enhanced with an increase in the number of metrics, and for the maximum size (all metrics) the influence of Lasso is minimized. Therefore, the minimum set of input data is the most informative. It should be noted that the presence of only 3 input data vectors (elementary metrics) does not allow to ensure the optimal choice of metrics ($SROCC < 0.7$), and more accurate results have been obtained when searching for 5 metrics ($SROCC > 0.8$). Depending on the configuration and the type of the neural network (feed-forward or cascade-forward), the difference in the SROCC value can reach 0.04 for a smaller number of inputs. In this case, the advantages over the elementary metrics can be provided by both of them, however, more often by the cascade-forward networks, used also in the resulting network.

At the second stage, for each type of neural network, several configurations are considered to select the optimal structure, similarly to [11]: from 1 to 5 hidden layers, with an equal or evenly decreasing number of neurons in each layer. Similarly, it has been previously determined that the appropriate selection of suitable NN activation functions eliminates the need for preliminary data linearization. The use of a higher number of neurons does not provide significantly better results. They are very similar for each number of layers and mostly dependent on the used datasets, therefore the decision to use 5 layers has been made. As activation function in hidden layers, we used tansig that allows normalizing the infinite range of values to the limited range $[-1;1]$ and the linear one for the last layer. After Lasso’s calculations, the combinations of metrics shown in Table 2 are considered (NNZ means the number of non-zero values). Another advantage

of the application of Lasso is that the use of even the smallest thresholds eliminates the weakest metrics.

Table 2. The combinations of metrics considered in the NN-based experiments (with the lists of metrics presented only when less than 10 metrics have been applied).

##	Pre-Processing Rule	Number of Metrics	List of Metrics
1	All metrics	42	-
2	Using Lasso (NNZ > 20)	37	-
3	Using Lasso (NNZ > 30)	35	-
4	Using Lasso (NNZ > 40)	32	-
5	Using Lasso (NNZ > 50)	27	-
6	Using Lasso (NNZ > 60)	19	-
7	Using Lasso (NNZ > 63)	15	-
8	Using Lasso (NNZ > 68)	10	CORNIA, HOSA, LPSI, MSGF, TCLT, IL-NIQE, QAC, SISBLIM_SM, SISBLIM_SFB, DESIQUE,
9	Using Lasso (NNZ > 70)	8	CORNIA, QAC, HOSA, LPSI, TCLT, SISBLIM_SFB, NIQMC, IL-NIQE
10	Using Lasso (NNZ > 75)	5	CORNIA, QAC, HOSA, LPSI, IL-NIQE
11	Using Lasso (NNZ > 80)	3	CORNIA, LPSI, IL-NIQE
12	Custom set 1	5	CORNIA, LPSI, HOSA, QAC, IL-NIQE
13	Custom set 2	5	SISBLIM_SM, HOSA, QAC, IL-NIQE, BRISQUE

Since there are some other databases for which it is possible to check the performance of the optimized combined metrics, such an analysis has been performed for the combination of 5 and 10 elementary metrics. To verify the universality of the proposed approach, it has been checked for widely known LIVE database [28] (for all types of distortions), for several subsets of distortions that can be formed for TID2013, as well as for several subsets for the KADID-10k database [73] that contain distortions typical for the RS applications. For comparison, the elementary metrics used as inputs of the combined NN metric have been taken.

3. Results

The results of the optimization of the pairs of elementary metrics confirm the validity of the proposed approach since a significant increase of the SROCC may be observed for the best combinations in comparison to IL-NIQE shown in Table 1. The best five combinations for the NA subset are presented in Table 3 together with the absolute values of the KROCC and PLCC values, although Spearman's correlation has been assumed as the objective function during the optimization. Similar results obtained for the whole TID2013 database are shown in Table 4. Due to very low values of the Pearson's correlation (below 0.1) for some metrics, their precise values have not been presented. The results provided in Tables 3 and 4 illustrate some interesting properties of the combined metrics. Firstly, all five "best" combinations for the NA subset contain the IL-NIQE metric, however, the situation for the whole TID2013 database is a bit different, particularly for the CM_2^+ family of metrics. Secondly, the elementary metrics combined with IL-NIQE are not the "best" among the elementary metrics listed in Table 1. It confirms that the combination of various types of metrics leads to better performance the use of those based on similar assumptions.

Starting from the combinations listed in Tables 3 and 4, the respective CM_3 and CM_3^+ families of the combined metrics have been optimized, as well as similar combinations of more elementary metrics. The obtained "best" combinations for the NA subset are presented in Table 5. As it may be observed for the combinations of five elementary metrics three of them are the same for both types of combined metrics whereas two of them are different. After a further increase of the number of elementary metrics ($N > 5$) there are no significant improvements in the performance, however, it leads to noticeably higher computational complexity. Table 6 illustrates the best results obtained using the combinations of 3 to 5 elementary metrics using formulas (1) and (2), for the whole TID2013 database.

Table 3. Performance of the five best combinations of two NR elementary metrics ($N = 2$) for the NA subset assuming the use of the weighted product (CM_2) and weighted sum (CM_2^+).

Metric	SROCC	KROCC	PLCC
CM_2 (ARISM, IL-NIQE)	0.7528	0.5540	0.7165
CM_2 (WNJE, IL-NIQE)	0.7526	0.5580	0.6868
CM_2 (ARISMC, IL-NIQE)	0.7489	0.5491	0.7131
CM_2 (LPSI, IL-NIQE)	0.7345	0.5379	<0.1
CM_2 (HOSA, IL-NIQE)	0.7332	0.5364	0.6974
CM_2^+ (ARISMC, IL-NIQE)	0.7682	0.5690	0.7378
CM_2^+ (ARISM, IL-NIQE)	0.7678	0.5691	0.7388
CM_2^+ (WNJE, IL-NIQE)	0.7542	0.5600	0.7655
CM_2^+ (CORNIA, IL-NIQE)	0.7502	0.5541	0.6090
CM_2^+ (HOSA, IL-NIQE)	0.7433	0.5456	0.5156

Table 4. Performance of the five best combinations of two NR elementary metrics ($N = 2$) for the whole TID2013 dataset assuming the use of the weighted product (CM_2) and weighted sum (CM_2^+).

Metric	SROCC	KROCC	PLCC
CM_2 (LPCSI, IL-NIQE)	0.5352	0.3783	0.0553
CM_2 (HOSA, IL-NIQE)	0.5255	0.3732	0.4040
CM_2 (TCLT, IL-NIQE)	0.5225	0.3690	0.1829
CM_2 (WNJE, IL-NIQE)	0.5191	0.3728	0.3092
CM_2 (HOSA, LPCSI)	0.5154	0.3615	<0.1
CM_2^+ (PSI, CORNIA)	0.5631	0.3978	0.4922
CM_2^+ (PSI, IL-NIQE)	0.5571	0.3955	<0.1
CM_2^+ (HOSA, IL-NIQE)	0.5425	0.3877	<0.1
CM_2^+ (SISBLIM_WM, IL-NIQE)	0.5384	0.3804	<0.1
CM_2^+ (CORNIA, Blur Metric)	0.5372	0.3801	<0.1

Table 5. Performance of the best combinations of NR elementary metrics ($N > 2$) for the NA subset assuming the use of the weighted product (CM) and weighted sum (CM^+).

Metric	SROCC	KROCC	PLCC
CM_3 (ARISM, IL-NIQE, BIBLE)	0.7839	0.5856	0.2993
CM_4 (ARISM, IL-NIQE, BIBLE, WNJE)	0.8056	0.6129	0.2769
CM_5 (ARISMC, IL-NIQE, BIBLE, WNJE, GM_LOG)	0.8132	0.6243	0.5847
CM_3^+ (ARISMC, IL-NIQE, CORNIA)	0.8089	0.6122	0.7024
CM_4^+ (ARISMC, IL-NIQE, CORNIA, WNJE)	0.8315	0.6406	0.7922
CM_5^+ (ARISMC, IL-NIQE, CORNIA, WNJE, QAC)	0.8406	0.6522	0.8320

Table 6. Performance of the best combinations of NR elementary metrics ($N > 2$) for the whole TID2013 database assuming the use of the weighted product (CM) and weighted sum (CM^+).

Metric	SROCC	KROCC	PLCC
CM_3 (IL-NIQE, HOSA, TCLT)	0.5765	0.4140	<0.1
CM_4 (IL-NIQE, LPCSI, TCLT, WNJE)	0.5995	0.4305	<0.1
CM_5 (IL-NIQE, HOSA, TCLT, LPCSI, WNJE)	0.6235	0.4521	<0.1
CM_3^+ (PSI, LPCSI, CORNIA)	0.5967	0.4252	<0.1
CM_4^+ (PSI, LPCSI, IL-NIQE, BIBLE)	0.6274	0.4492	0.1301
CM_5^+ (HOSA, LPCSI, IL-NIQE, OG_IQA, WNJE)	0.6442	0.4648	0.5052

The results obtained for neural networks with 5, 10, and 15 inputs selected by the Lasso algorithm are shown in Table 7. In its description 1 L–4 L denote the number of hidden layers, whereas “equal” and “half” stand for the number of neurons. The number of neurons in the first hidden layer was equal to the number of elementary metrics. There were two variants for determining the number(s) of neurons in the next hidden layers. The first option was to set the same number(s) of neurons whilst the second was to reduce the number of neurons approximately twice in each next hidden layer.

Table 7. Results of the cascade NN obtained for the NA subset using the 5, 10, and 15 input metrics (1 L–4 L denote the number of hidden layers, “equal” and “half” are the number of neurons; the best SROCC and PLCC correlations are marked with bold fonts).

##	NN Neurons	5 Metrics		10 Metrics		15 Metrics	
		SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
1	1 L	0.8173	0.8483	0.8768	0.8889	0.8991	0.9106
2	2 L, equal	0.8114	0.8409	0.8813	0.8906	0.8996	0.9122
3	3 L, equal	0.8171	0.8475	0.8811	0.8932	0.9190	0.9268
4	4 L, equal	0.8199	0.8512	0.8730	0.8856	0.9118	0.9221
5	2 L, half	0.8208	0.8527	0.8787	0.8932	0.9060	0.9146
6	3 L, half	0.8155	0.8411	0.8797	0.8946	0.9057	0.9168
7	4 L, half	0.8107	0.8362	0.8739	0.8914	0.9051	0.9155

A graphical illustration of the obtained performance for some of the “best” combinations is presented on the scatter plots shown in Figure 2 and discussed in Section 4.

The dependence of the accuracy of the neural network (considered as the SROCC values calculated for the NA subset) on the number of input metrics is shown in Figure 3. For different configurations of networks, fairly similar results have been obtained. For this task, the most critical factors are the list of the used metrics, their number, whereas the specific network configuration is less crucial. In terms of the totality of all stages of networks training, it can be noted that a single-layer network does not always provide the necessary computing capacity, and the use of 4 hidden layers does not lead to a visible advantage. Hence, a stable result with a low complexity cascade network is ensured when 2–3 hidden layers are used and, basically, with an equal number of neurons in the layers.

The results of the verification of the universality of the proposed NN-based metric are presented in Table 8. The results obtained for the whole LIVE database are not the best. The NN-based metric is better than some elementary metrics and worse than other ones, however, the obtained SROCC for this database is considered large enough. It should be noted here that some elementary metrics have been obtained just for a limited set of distortion types present in the LIVE database and this explains the aforementioned observations. For different subsets of the TID2013, the results for the designed NN-based metric are either the best or close to the best.

For the subsets of the KADID-10k database, the data for 5 and 10 metrics have been considered separately. This database contains the following groups of distortions [73]: blurs (##1–3), color distortions (##4–8), compression (##9–10), noise (##11–15), brightness change (##16–18), spatial distortions (##19–23), over-sharpening (#24), and contrast change (#25).

For 5 metrics, the results for the NN-based metrics are close to the best elementary metric whereas different elementary metrics are the best for different subsets. It confirms the observation of the benefits achieved for the combination of the elementary metrics that are not the best themselves but somehow complementary, utilizing various “nature” of image data (different features, data representation, transforms, etc.) and based on different assumptions. To illustrate the idea of this “compensation”, the scatter plots obtained for elementary metrics and their combination for the KADID-10k database are presented in Figure 4. Different outliers may be easily observed for elementary metrics and the NN-based metric provides the smallest number of obvious outliers and practically linear dependence between MOS and the metric.

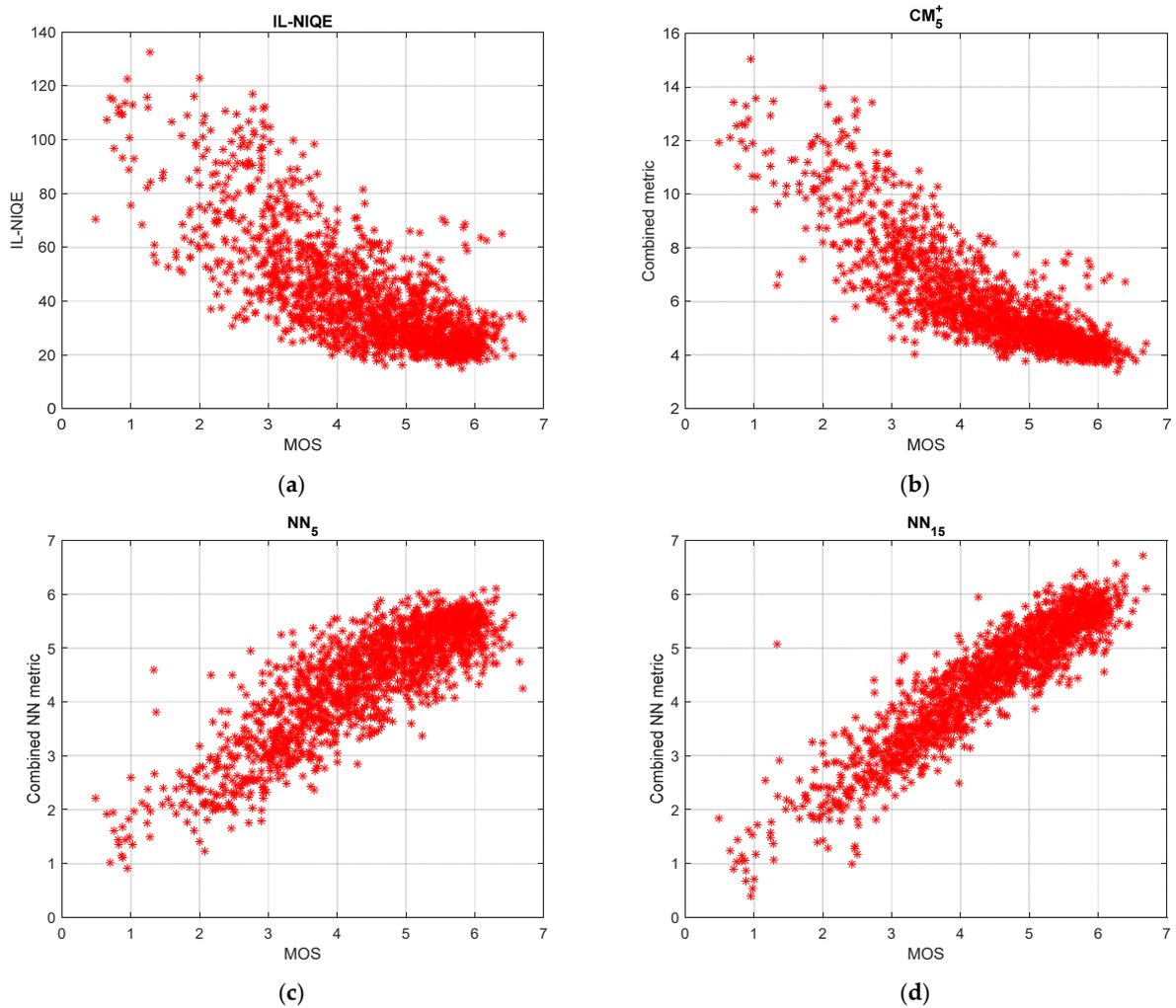


Figure 2. Scatter plots illustrating the correlation between the subjective (MOS values) and objective metrics for the NA subset: (a) the “best” elementary metric (IL-NIQE); (b) the “best” combined weighted sum of 5 metrics (CM₅⁺ metric); (c) the “best” combined NN-based metrics using 5 inputs; (d) the “best” combined NN-based metrics using 15 inputs.

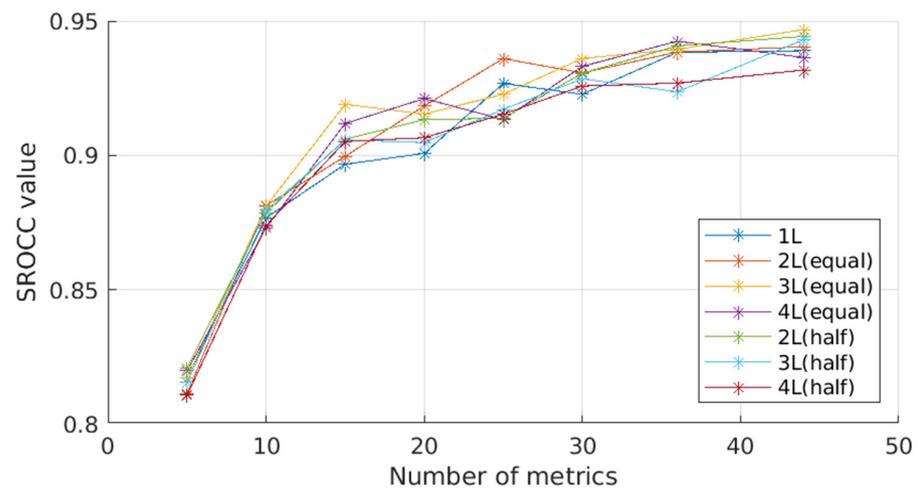
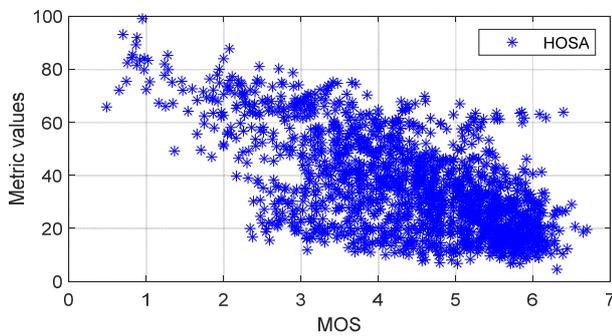


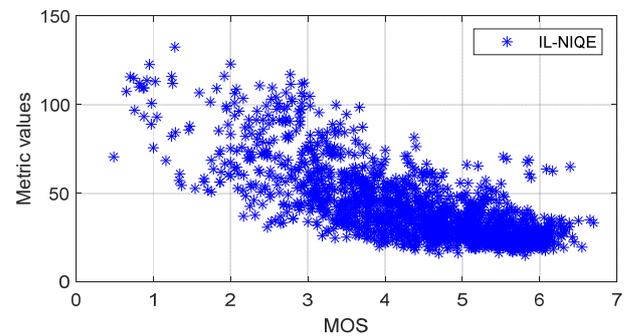
Figure 3. The performance of the NN-based metrics for various numbers of input metrics.

Table 8. Results of the verification of the universality of the proposed NN-based metric: SROCC values obtained for LIVE database and various subsets from the TID2013 and KADID-10k databases; the highest SROCC correlations are marked with bold fonts.

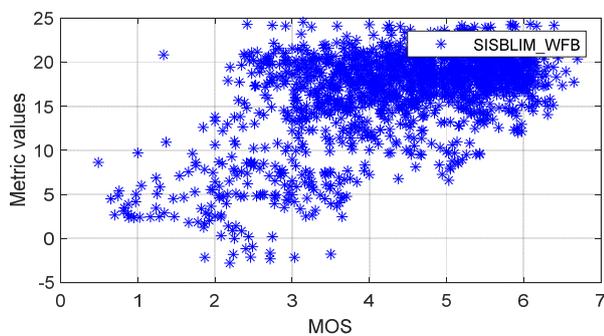
Database	LIVE	TID2013				KADID-10k				
Subsets	Noise+SC	Noises	ImProc	RS	Blur	Compr	Noises	RS	RS2	
Distortions #	All	1–7,19	1,2,4–7,19	8–11,21	1–11,19,21	1–3	9,10	11–15	1–3,9–15	1,9–15
5 metrics										
SISBLIM_WFB	0.8233	0.1272	0.2102	0.6823	0.3789	0.8625	0.7083	0.5265	0.6590	0.6204
IL-NIQE	0.8537	0.6763	0.6092	0.7770	0.7221	0.8766	0.7916	0.6207	0.7794	0.7474
HOSA	0.9430	0.3693	0.5469	0.7987	0.5690	0.8846	0.8545	0.3206	0.6540	0.5886
CORNIA	0.9550	0.0978	0.1253	0.8771	0.4735	0.8908	0.8438	0.3628	0.6623	0.5884
MSGF	0.9166	0.1851	0.3399	0.4845	0.3085	0.8107	0.8214	0.2552	0.4840	0.4020
NN metric	0.8921	0.7880	0.7641	0.8459	0.8208	0.8492	0.7742	0.4502	0.7135	0.6750
10 metrics										
MLV	0.2326	0.1162	0.1588	0.5124	0.1755	0.8489	0.2757	0.2213	0.4017	0.2799
SISBLIM_WFB	0.8233	0.1272	0.2102	0.6823	0.3789	0.8625	0.7083	0.5265	0.6590	0.6204
TCLT	0.9950	0.0395	0.1487	0.7099	0.3151	0.7683	0.7086	0.1131	0.4622	0.3941
IL-NIQE	0.8537	0.6763	0.6092	0.7770	0.7221	0.8766	0.7916	0.6207	0.7794	0.7474
HOSA	0.9430	0.3693	0.5469	0.7987	0.5690	0.8846	0.8545	0.3206	0.6540	0.5886
CORNIA	0.9550	0.0978	0.1253	0.8771	0.4735	0.8908	0.8438	0.3628	0.6623	0.5884
MSGF	0.9166	0.1851	0.3399	0.4845	0.3085	0.8107	0.8214	0.2552	0.4840	0.4020
PSI	0.1638	0.3090	0.4925	0.3596	0.0105	0.8771	0.1072	0.0839	0.2445	0.0468
NMC	0.3607	0.2344	0.2865	0.3606	0.0485	0.6812	0.3087	0.0906	0.2171	0.0520
BLUR	0.3361	0.4312	0.6477	0.3554	0.0393	0.8510	0.2328	0.2521	0.1965	0.0205
NN metric	0.8342	0.8484	0.8362	0.9051	0.8813	0.8954	0.8203	0.6217	0.7964	0.7648



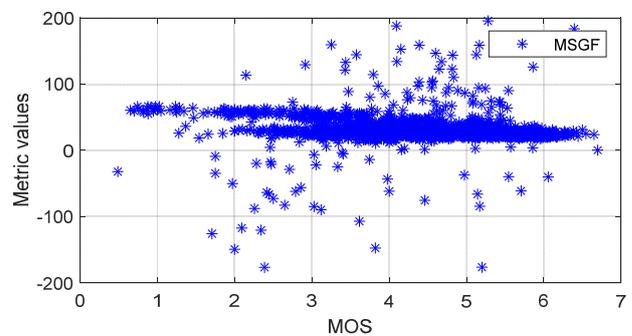
(a)



(b)



(c)



(d)

Figure 4. Cont.

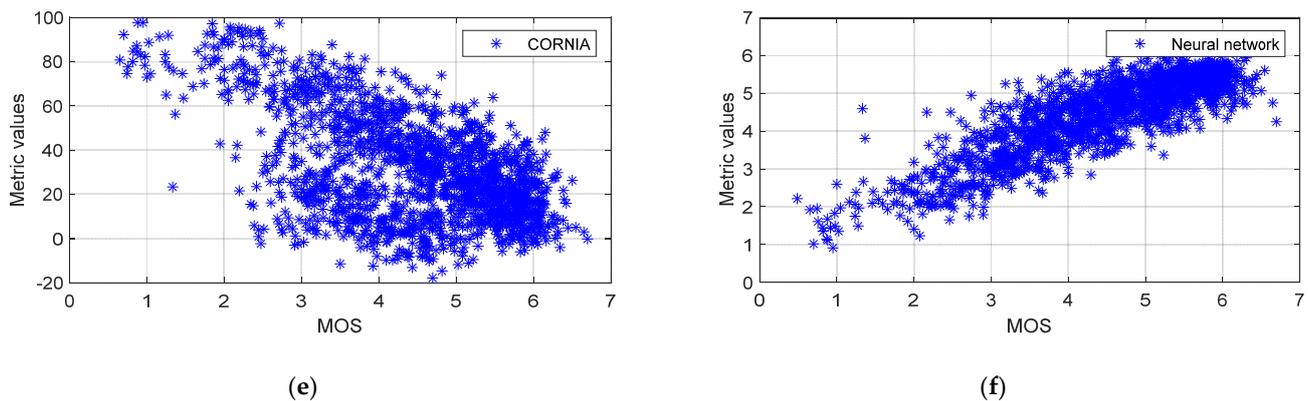


Figure 4. Scatter plots obtained for five elementary metrics and the NN-based combined metric for the KADID-10k database: (a) HOSA; (b) IL-NIQE; (c) SISBLIM_WFB; (d) MSGF; (e) CORNIA; (f) combined NN-based metric using 5 inputs.

4. Discussion

The visual analysis of scatter plots presented in Figure 2c,d shows one important benefit of the NN-based combined metrics. The dependence of the metric on MOS values is practically linear and this produces high SROCC and PLCC, simultaneously. This is caused by the fact that MOS values have been used as the target function in the NN training. For the designed CM_5^+ metric based on the weighted sum (Figure 2b) the dependence is not linear and, due to this, PLCC can be sufficiently smaller than SROCC. Meanwhile, this drawback can be easily removed by an appropriate fitting made after the calculation of CM^+ and optimization of its coefficients.

For some of the CM and CM^+ metrics, particularly based on the combination of a relatively small number of elementary metrics, the optimization of their coefficients according to the maximization of Spearman's correlation, has led to very low values of Pearson's correlation and high nonlinearity of the relation between them and MOS values. One of the possible approaches to avoid this problem is the choice of the PLCC as the goal function, similarly as in some earlier papers [29,30], however, in this case, smaller SROCC values may be achieved.

Although for 5 elementary metrics better SROCC results may be obtained using the weighted sum, even in comparison with the NN-based approach, the advantages of the neural networks may be observed when more elementary metrics are used as inputs, as illustrated in Figure 3.

The choice of elementary metrics made by the Lasso algorithm during the NN-based design differs from the "best" combinations used in the CM and CM^+ metrics. Comparing the case of five elementary metrics, the following metrics have been selected by the Lasso: SISBLIM_WFB, IL-NIQE, HOSA, CORNIA, and MSGF, whereas the highest SROCC value for the CM_5^+ metric has been obtained for the combination of ARISMC, IL-NIQE, CORNIA, WNJE, and QAC as presented in Table 5. Interestingly, only two of the metrics selected in these two cases are the same.

As shown in Table 8, for 10 elementary metrics, the performance of the NN-based metric is the best for 4 out of 5 subsets of the KADID-10k database and close to the best in the remaining case. Thus, in general, although optimized for the database TID2013, the performance of the designed NN-based metrics is appropriately good for other databases.

The results presented in the paper confirm the usefulness of the various combination methods of elementary no-reference image quality metrics for the evaluation of images subject to distortions typical for RS images. Nevertheless, considering the results obtained for the whole TID2013 database, further experiments would be necessary to improve the correlation of such metrics with subjective quality scores as well as for the enhancement of the universality of the combined NR metrics. A design of such metrics sensitive to various types of distortions should be one of the natural directions of further research.

Author Contributions: Conceptualization, A.R., O.I., V.L. and K.O.; methodology, A.R., O.I., V.L. and K.O.; software, A.R., O.I. and K.O.; validation, O.I., J.F. and K.O.; formal analysis, O.I. and V.L.; investigation, A.R., O.I., V.L. and K.O.; resources, A.R., O.I. and K.O.; data curation, A.R., O.I. and K.O.; writing—original draft preparation, O.I., V.L. and K.O.; writing—review and editing, V.L. and K.O.; visualization, O.I., J.F. and K.O.; supervision, V.L. and K.O.; project administration, V.L. and K.O.; funding acquisition, V.L. and K.O. All authors have read and agreed to the published version of the manuscript.

Funding: The research is partially co-financed by the Polish National Agency for Academic Exchange (NAWA) and the Ministry of Education and Science of Ukraine under the project no. PPN/BUA/2019/1/00074 entitled “Methods of intelligent image and video processing based on visual quality metrics for emerging applications”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: LIVE (<https://live.ece.utexas.edu/research/Quality/subjective.htm>) (accessed on 21 December 2021), TID2013 (<http://www.ponomarenko.info/tid2013.htm>), KADID-10k (<http://database.mmsp-kn.de/kadid-10k-database.html>) (accessed on 21 December 2021).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Gonzalez-Roglich, M.; Zvoleff, A.; Noon, M.; Liniger, H.; Fleiner, R.; Harari, N.; Garcia, C. Synergizing Global Tools to Monitor Progress towards Land Degradation Neutrality: Trends.Earth and the World Overview of Conservation Approaches and Technologies Sustainable Land Management Database. *Environ. Sci. Policy* **2019**, *93*, 34–42. [CrossRef]
- Khorram, S.; van der Wiele, C.F.; Koch, F.H.; Nelson, S.A.C.; Potts, M.D. Future Trends in Remote Sensing. In *Principles of Applied Remote Sensing*; Springer: Cham, Switzerland, 2016; pp. 277–285, ISBN 978-3-319-22559-3.
- Kussul, N.; Lemoine, G.; Gallego, F.J.; Skakun, S.V.; Lavreniuk, M.; Shelestov, A. Parcel-Based Crop Classification in Ukraine Using Landsat-8 Data and Sentinel-1A Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2500–2508. [CrossRef]
- Deledalle, C.-A.; Denis, L.; Tabti, S.; Tupin, F. MuLoG, or How to Apply Gaussian Denoisers to Multi-Channel SAR Speckle Reduction? *IEEE Trans. Image Process.* **2017**, *26*, 4389–4403. [CrossRef] [PubMed]
- Zhong, P.; Wang, R. Multiple-Spectral-Band CRFs for Denoising Junk Bands of Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2260–2275. [CrossRef]
- van Zyl Marais, I.; Steyn, W.H.; du Preez, J.A. Onboard Image Quality Assessment for a Small Low Earth Orbit Satellite. In Proceedings of the 7th IAA Symposium on Small Satellites for Earth Observation, Berlin, Germany, 8 May 2009; Paper IAA B7-0602.
- Vasicek, Z.; Bidlo, M.; Sekanina, L.; Torresen, J.; Glette, K.; Furuholmen, M. Evolution of Impulse Bursts Noise Filters. In Proceedings of the 2009 NASA/ESA Conference on Adaptive Hardware and Systems, San Francisco, CA, USA, 29 July–1 August 2009; pp. 27–34.
- Agudelo-Medina, O.A.; Benitez-Restrepo, H.D.; Vivone, G.; Bovik, A. Perceptual Quality Assessment of Pan-Sharpned Images. *Remote Sens.* **2019**, *11*, 877. [CrossRef]
- Yuan, T.; Zheng, X.; Hu, X.; Zhou, W.; Wang, W. A Method for the Evaluation of Image Quality According to the Recognition Effectiveness of Objects in the Optical Remote Sensing Image Using Machine Learning Algorithm. *PLoS ONE* **2014**, *9*, e86528. [CrossRef]
- Jagalingam, P.; Hegde, A.V. A Review of Quality Metrics for Fused Image. *Aquat. Procedia* **2015**, *4*, 133–142. [CrossRef]
- 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]
- Lin, W.; Jay Kuo, C.-C. Perceptual Visual Quality Metrics: A Survey. *J. Vis. Commun. Image Represent.* **2011**, *22*, 297–312. [CrossRef]
- Niu, Y.; Zhong, Y.; Guo, W.; Shi, Y.; Chen, P. 2D and 3D Image Quality Assessment: A Survey of Metrics and Challenges. *IEEE Access* **2019**, *7*, 782–801. [CrossRef]
- Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
- Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale Structural Similarity for Image Quality Assessment. In Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 9–12 November 2003; pp. 1398–1402.

16. Rubel, O.; Rubel, A.; Lukin, V.; Egiazarian, K. Blind DCT-Based Prediction of Image Denoising Efficiency Using Neural Networks. In Proceedings of the 2018 7th European Workshop on Visual Information Processing (EUVIP), Tampere, Finland, 26–28 November 2018; pp. 1–6.
17. Rubel, O.; Rubel, A.; Lukin, V.; Carli, M.; Egiazarian, K. Blind Prediction of Original Image Quality for Sentinel SAR Data. In Proceedings of the 2019 8th European Workshop on Visual Information Processing (EUVIP), Rome, Italy, 28–31 October 2019; pp. 105–110.
18. Yan, J.; Bai, X.; Xiao, Y.; Zhang, Y.; Lv, X. No-Reference Remote Sensing Image Quality Assessment Based on Gradient-Weighted Natural Scene Statistics in Spatial Domain. *J. Electron. Imaging* **2019**, *28*, 1. [[CrossRef](#)]
19. Ferzli, R.; Karam, L.J. A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur (JNB). *IEEE Trans. Image Process.* **2009**, *18*, 717–728. [[CrossRef](#)] [[PubMed](#)]
20. Narvekar, N.D.; Karam, L.J. A No-Reference Perceptual Image Sharpness Metric Based on a Cumulative Probability of Blur Detection. In Proceedings of the 2009 International Workshop on Quality of Multimedia Experience (QoMEX), San Diego, CA, USA, 29–31 July 2009; pp. 87–91.
21. Zhu, T.; Karam, L. A No-Reference Objective Image Quality Metric Based on Perceptually Weighted Local Noise. *EURASIP J. Image Video Process.* **2014**, *2014*, 5. [[CrossRef](#)]
22. Li, S.; Yang, Z.; Li, H. Statistical Evaluation of No-Reference Image Quality Assessment Metrics for Remote Sensing Images. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 133. [[CrossRef](#)]
23. Yang, J.; Zhao, Y.; Yi, C.; Chan, J.C.-W. No-Reference Hyperspectral Image Quality Assessment via Quality-Sensitive Features Learning. *Remote Sens.* **2017**, *9*, 305. [[CrossRef](#)]
24. Ieremeiev, O.; Lukin, V.; Ponomarenko, N.; Egiazarian, K. Combined No-Reference IQA Metric and Its Performance Analysis. *Electron. Imaging* **2019**, *2019*, 260. [[CrossRef](#)]
25. Zhang, L.; Zhang, L.; Bovik, A.C. A Feature-Enriched Completely Blind Image Quality Evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. [[CrossRef](#)]
26. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “Completely Blind” Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [[CrossRef](#)]
27. Ponomarenko, N.; Jin, L.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Astola, J.; Vozel, B.; Chehdi, K.; Carli, M.; Battisti, F.; et al. Image Database TID2013: Peculiarities, Results and Perspectives. *Signal Process. Image Commun.* **2015**, *30*, 57–77. [[CrossRef](#)]
28. 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](#)]
29. Okarma, K. Combined Image Similarity Index. *Opt. Rev.* **2012**, *19*, 349–354. [[CrossRef](#)]
30. 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](#)]
31. Rassool, R. VMAF Reproducibility: Validating a Perceptual Practical Video Quality Metric. In Proceedings of the 2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Cagliari, Italy, 7–9 June 2017; pp. 1–2.
32. Kazakeviciute-Januskeviciene, G.; Janusonis, E.; Bausys, R.; Limba, T.; Kiskis, M. Assessment of the Segmentation of RGB Remote Sensing Images: A Subjective Approach. *Remote Sens.* **2020**, *12*, 4152. [[CrossRef](#)]
33. Christophe, E.; Leger, D.; Maillhes, C. Quality Criteria Benchmark for Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2103–2114. [[CrossRef](#)]
34. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
35. Fazzini, P.; De Felice Proia, G.; Adamo, M.; Blonda, P.; Petracchini, F.; Forte, L.; Tarantino, C. Sentinel-2 Remote Sensed Image Classification with Patchwise Trained ConvNets for Grassland Habitat Discrimination. *Remote Sens.* **2021**, *13*, 2276. [[CrossRef](#)]
36. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [[CrossRef](#)] [[PubMed](#)]
37. Venkatanath, N.; Praneeth, D.; Chandrasekhar, B.M.; Channappayya, S.S.; Medasani, S.S. Blind Image Quality Evaluation Using Perception Based Features. In Proceedings of the 2015 Twenty First National Conference on Communications (NCC), Mumbai, India, 27 February–1 March 2015; pp. 1–6.
38. Moorthy, A.K.; Bovik, A.C. A Two-Step Framework for Constructing Blind Image Quality Indices. *IEEE Signal Process. Lett.* **2010**, *17*, 513–516. [[CrossRef](#)]
39. Saad, M.A.; Bovik, A.C.; Charrier, C. Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain. *IEEE Trans. Image Process.* **2012**, *21*, 3339–3352. [[CrossRef](#)]
40. Moorthy, A.K.; Bovik, A.C. Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality. *IEEE Trans. Image Process.* **2011**, *20*, 3350–3364. [[CrossRef](#)] [[PubMed](#)]
41. Ye, P.; Kumar, J.; Kang, L.; Doermann, D. Unsupervised Feature Learning Framework for No-Reference Image Quality Assessment. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1098–1105.
42. Xue, W.; Zhang, L.; Mou, X. Learning without Human Scores for Blind Image Quality Assessment. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 995–1002.
43. Gu, K.; Zhai, G.; Lin, W.; Yang, X.; Zhang, W. No-Reference Image Sharpness Assessment in Autoregressive Parameter Space. *IEEE Trans. Image Process.* **2015**, *24*, 3218–3231. [[CrossRef](#)]

44. Crété-Roffet, F.; Dolmiere, T.; Ladret, P.; Nicolas, M. The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric. In *Human Vision and Electronic Imaging XII*; International Society for Optics and Photonics: San Jose, CA, USA, 2007; p. 649201.
45. Vu, P.V.; Chandler, D.M. A Fast Wavelet-Based Algorithm for Global and Local Image Sharpness Estimation. *IEEE Signal Process. Lett.* **2012**, *19*, 423–426. [[CrossRef](#)]
46. Feichtenhofer, C.; Fassold, H.; Schallauer, P. A Perceptual Image Sharpness Metric Based on Local Edge Gradient Analysis. *IEEE Signal Process. Lett.* **2013**, *20*, 379–382. [[CrossRef](#)]
47. Wang, Z.; Sheikh, H.R.; Bovik, A.C. No-Reference Perceptual Quality Assessment of JPEG Compressed Images. In Proceedings of the 2002 International Conference on Image Processing (ICIP), Rochester, NY, USA, 22–25 September 2002; Volume 1, pp. I-477–I-480.
48. Li, L.; Lin, W.; Wang, X.; Yang, G.; Bahrami, K.; Kot, A.C. No-Reference Image Blur Assessment Based on Discrete Orthogonal Moments. *IEEE Trans. Cybern.* **2016**, *46*, 39–50. [[CrossRef](#)]
49. Xu, J.; Ye, P.; Li, Q.; Du, H.; Liu, Y.; Doermann, D. Blind Image Quality Assessment Based on High Order Statistics Aggregation. *IEEE Trans. Image Process.* **2016**, *25*, 4444–4457. [[CrossRef](#)]
50. Min, X.; Gu, K.; Zhai, G.; Liu, J.; Yang, X.; Chen, C.W. Blind Quality Assessment Based on Pseudo-Reference Image. *IEEE Trans. Multimed.* **2018**, *20*, 2049–2062. [[CrossRef](#)]
51. Gu, K.; Zhai, G.; Yang, X.; Zhang, W. Hybrid No-Reference Quality Metric for Singly and Multiply Distorted Images. *IEEE Trans. Broadcast.* **2014**, *60*, 555–567. [[CrossRef](#)]
52. Chandler, D.M. Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy. *J. Electron. Imaging* **2010**, *19*, 011006. [[CrossRef](#)]
53. Hassen, R.; Wang, Z.; Salama, M.M.A. Image Sharpness Assessment Based on Local Phase Coherence. *IEEE Trans. Image Process.* **2013**, *22*, 2798–2810. [[CrossRef](#)]
54. Wu, Q.; Wang, Z.; Li, H. A Highly Efficient Method for Blind Image Quality Assessment. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 339–343.
55. Bahrami, K.; Kot, A.C. A Fast Approach for No-Reference Image Sharpness Assessment Based on Maximum Local Variation. *IEEE Signal Process. Lett.* **2014**, *21*, 751–755. [[CrossRef](#)]
56. Wu, Q.; Li, H.; Meng, F.; Ngan, K.N.; Zhu, S. No Reference Image Quality Assessment Metric via Multi-Domain Structural Information and Piecewise Regression. *J. Vis. Commun. Image Represent.* **2015**, *32*, 205–216. [[CrossRef](#)]
57. Gu, K.; Lin, W.; Zhai, G.; Yang, X.; Zhang, W.; Chen, C.W. No-Reference Quality Metric of Contrast-Distorted Images Based on Information Maximization. *IEEE Trans. Cybern.* **2017**, *47*, 4559–4565. [[CrossRef](#)] [[PubMed](#)]
58. Golestaneh, S.A.; Chandler, D.M. No-Reference Quality Assessment of JPEG Images via a Quality Relevance Map. *IEEE Signal Process. Lett.* **2014**, *21*, 155–158. [[CrossRef](#)]
59. Gong, Y.; Sbalzarini, I.F. Image Enhancement by Gradient Distribution Specification. In *Lecture Notes in Computer Science, Proceedings of the Computer Vision—ACCV 2014 Workshops, Singapore, 1–2 November 2014*; Jawahar, C.V., Shan, S., Eds.; Springer: Cham, Switzerland, 2015; Volume 9009, pp. 47–62, ISBN 978-3-319-16630-8.
60. Zhang, Y.; Moorthy, A.K.; Chandler, D.M.; Bovik, A.C. C-DIVINE: No-Reference Image Quality Assessment Based on Local Magnitude and Phase Statistics of Natural Scenes. *Signal Process. Image Commun.* **2014**, *29*, 725–747. [[CrossRef](#)]
61. Liu, L.; Hua, Y.; Zhao, Q.; Huang, H.; Bovik, A.C. Blind Image Quality Assessment by Relative Gradient Statistics and Adaboosting Neural Network. *Signal Process. Image Commun.* **2016**, *40*, 1–15. [[CrossRef](#)]
62. Zhang, Y.; Chandler, D.M. No-Reference Image Quality Assessment Based on Log-Derivative Statistics of Natural Scenes. *J. Electron. Imaging* **2013**, *22*, 043025. [[CrossRef](#)]
63. Rakhshanfar, M.; Amer, M.A. Sparsity-Based No-Reference Image Quality Assessment for Automatic Denoising. *Signal Image Video Process.* **2018**, *12*, 739–747. [[CrossRef](#)]
64. Ma, K.; Liu, W.; Liu, T.; Wang, Z.; Tao, D. DipIQ: Blind Image Quality Assessment by Learning-to-Rank Discriminable Image Pairs. *IEEE Trans. Image Process.* **2017**, *26*, 3951–3964. [[CrossRef](#)]
65. Li, L.; Yan, Y.; Lu, Z.; Wu, J.; Gu, K.; Wang, S. No-Reference Quality Assessment of Deblurred Images Based on Natural Scene Statistics. *IEEE Access* **2017**, *5*, 2163–2171. [[CrossRef](#)]
66. Wu, Q.; Li, H.; Meng, F.; Ngan, K.N.; Luo, B.; Huang, C.; Zeng, B. Blind Image Quality Assessment Based on Multichannel Feature Fusion and Label Transfer. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 425–440. [[CrossRef](#)]
67. Xue, W.; Mou, X.; Zhang, L.; Bovik, A.C.; Feng, X. Blind Image Quality Assessment Using Joint Statistics of Gradient Magnitude and Laplacian Features. *IEEE Trans. Image Process.* **2014**, *23*, 4850–4862. [[CrossRef](#)] [[PubMed](#)]
68. Ponomarenko, N.N.; Lukin, V.V.; Ereemeev, O.I.; Egiazarian, K.O.; Astola, J.T. *Sharpness Metric for No-Reference Image Visual Quality Assessment*; Egiazarian, K.O., Agaian, S.S., Gotchev, A.P., Recker, J., Wang, G., Eds.; International Society for Optical Engineering: Bellingham, WA, USA, 2012; p. 829519.
69. Liu, L.; Liu, B.; Huang, H.; Bovik, A.C. No-Reference Image Quality Assessment Based on Spatial and Spectral Entropies. *Signal Process. Image Commun.* **2014**, *29*, 856–863. [[CrossRef](#)]
70. Lukin, V.V.; Ponomarenko, N.N.; Ereemeev, O.I.; Egiazarian, K.O.; Astola, J. *Combining Full-Reference Image Visual Quality Metrics by Neural Network*; Rogowitz, B.E., Pappas, T.N., de Ridder, H., Eds.; International Society for Optical Engineering: Bellingham, WA, USA, 2015; p. 93940K.

71. Bosse, S.; Maniry, D.; Wiegand, T.; Samek, W. A Deep Neural Network for Image Quality Assessment. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3773–3777.
72. Bosse, S.; Maniry, D.; Muller, K.-R.; Wiegand, T.; Samek, W. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Trans. Image Process.* **2018**, *27*, 206–219. [[CrossRef](#)] [[PubMed](#)]
73. 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 (QoME), Berlin, Germany, 5–7 June 2019; pp. 1–3.