*Article*

# Full-Reference Quality Metric Based on Neural Network to Assess the Visual Quality of Remote Sensing Images

**Oleg Ieremeiev** [1] , **Vladimir Lukin** [1], **Krzysztof Okarma** [2,*] and **Karen Egiazarian** [3]

[1]  Department of Information and Communication Technologies, National Aerospace University, 61070 Kharkov, Ukraine; 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

[3]  Computational Imaging Group, Tampere University, 33720 Tampere, Finland; karen.eguiazarian@tuni.fi

[*]  Correspondence: okarma@zut.edu.pl

check for
updates

**Abstract:** Remote sensing images are subject to different types of degradations. The visual quality of such images is important because their visual inspection and analysis are still widely used in practice. To characterize the visual quality of remote sensing images, the use of specialized visual quality metrics is desired. Although the attempts to create such metrics are limited, there is a great number of visual quality metrics designed for other applications. Our idea is that some of these metrics can be employed in remote sensing under the condition that those metrics have been designed for the same distortion types. Thus, image databases that contain images with types of distortions that are of interest should be looked for. It has been checked what known visual quality metrics perform well for images with such degradations and an opportunity to design neural network-based combined metrics with improved performance has been studied. It is shown that for such combined metrics, their Spearman correlation coefficient with mean opinion score exceeds 0.97 for subsets of images in the Tampere Image Database (TID2013). Since different types of elementary metric pre-processing and neural network design have been considered, it has been demonstrated that it is enough to have two hidden layers and about twenty inputs. Examples of using known and designed visual quality metrics in remote sensing are presented.

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

## 1. Introduction

Currently, there are a great number of applications of remote sensing (RS) [1,2]. There are many reasons behind this [3,4]. Firstly, modern RS sensors are able to provide data (images) from which useful information can be retrieved for large territories with appropriate accuracy (reliability). Secondly, there exist systems capable of carrying out frequent observations (monitoring) of given terrains that, in turn, allow the analysis of changes or development of certain processes [5,6]. Due to the modern tendency to acquire multichannel images (a set of images with different wavelengths and/or polarizations [7–10]) and their pre-processing (that might include co-registration, geometric and radiometric correction, calibration, etc. [1]), RS data can be well prepared for further analysis and processing.

However, this does not mean that the quality of RS images is perfect. There are numerous factors that influence RS image quality (in wide sense) and prevent the solving of various tasks of RS data processing. For example, a part of a sensed terrain can be closed by clouds and this sufficiently decreases the quality (usefulness) of such optical or infrared images [11]. This type of quality degradation

may be very troublesome not only because of necessary cloud cover detection and removal [12,13], but due to its specificity, since in natural RGB images captured by cameras, clouds are typically not considered as a distortion. Hence, there are no such kinds of distortion in any dataset that may be used for the verification of image quality metrics with the use of subjective scores, as it is done in this paper. Unfortunately, the lack of such appropriate datasets significantly limits the studies related to the evaluation of cloud cover in terms of image quality assessment.

Noise of different origin and type is, probably, one of degrading factors met the most often [8–11]. Different types of smearing and blur can be frequently encountered as well [11,14,15]. Lossy compression, often used in remote sensing data transferring and storage [16–19], can also lead to noticeable degradations. Image pre-filtering or interpolation can result in residual and/or newly introduced degradations [8,10,20]. The presence of multiple degradations is possible as well, like blurred noisy images or noisy images compressed in a lossy manner [21,22].

Therefore, in practice, different situations are possible:

- an image seems to be perfect, i.e., no degradations can be visually detected (sharpness is satisfactory, no noise is visible, no other degradations are observed);
- an image is multichannel and there are component images of very high quality and component images of quite low quality [9,23]; for RS data with a large number of components, i.e., hyperspectral images, this can be detected by component-wise visualization and analysis of images;
- an acquired image is originally degraded in some way, e.g., due to the principle of imaging system operation; good examples are synthetic aperture radar (SAR) images, for which a speckle noise is always present [8,24].

This means that the quality of original (acquired) images should be characterized quantitatively using some metrics. An efficiency of RS image pre-processing (e.g., denoising or lossy compression) should be characterized as well. In this sense, there are several groups of metrics (criteria) that can be used for this purpose. Firstly, there are practical situations when full-reference metrics can be applied. This happens, for example, in lossy compression of data when a metric can be calculated using original and compressed images [17–19]. Secondly, no-reference metrics can be used when distortion-free data are not available [11,24]. Usually, some parameters of an image are estimated to calculate a no-reference metric in that case. Note that there are quite successful attempts to predict full-reference metrics without having reference images [25,26]. Finally, there are many metrics that characterize image quality (or efficiency of image processing) from the viewpoint of quality of solving the final tasks [27–29]. These can be, e.g., the area under the curve [30] or the probabilities of correct classification [31].

It is obvious that many metrics are correlated. For example, RS image classification criteria depend on the quality of the original data, although the efficiency of classification is also dependent on a used set of features, an applied classifier and an used training method. In this paper, we concentrate on metrics characterizing the quality of original images or images after pre-processing, such as denoising or lossy compression, focusing on full-reference metrics and, in particular, visual quality metrics.

Conventional metrics, such as mean square error (MSE) or peak signal-to-noise ratio (PSNR), are still widely used in the analysis of RS images or evaluation of the efficiency of their processing [29,32–34]. Meanwhile, there is an obvious tendency to apply visual quality metrics [33,35–40]. There are several papers utilizing the Structural SIMilarity (SSIM) [41], which is probably the oldest—except for the Universal Image Quality Index (UQI) [42], being its direct predecessor—visual quality metric [33,35,36]; some other visual quality metrics have been designed and tested recently [37–40] for particular applications, such as image pan-sharpening, fusion, and object detection. However, the number of papers where visual quality metrics are employed is still limited [33–40,43–45].

Nevertheless, there are several reasons to apply visual quality metrics. It is known that the human vision system (HVS) pays primary attention to image sharpness [46–49], being highly correlated with both HVS-based visual quality metrics and many tasks of image processing, such as denoising,

lossy compression, edge detection, segmentation, and classification. So, it is obviously reasonable to know what known metrics can be applied, which are the best among them, and if the metric performance can be further improved. Similar problems have already been partly solved in multimedia applications where: (a) a lot of HVS-metrics have been proposed [46–49]; (b) methodologies of their testing (verification) have been proposed [46–49] (c) databases for the metrics' verification have been created [46–52]; (d) preliminary conclusions have been drawn [46–50] that allow finding good metrics for particular type(s) of distortions; (e) ways to improve metrics' performance have been put forward, including the design of combined metrics [53–57] or neural network (NN)-based metrics (some examples are given in [58–60]).

However, this potential has not been exploited in RS imaging. In our opinion, there are several reasons behind this. Firstly, there are no special databases of RS images that can be used to evaluate the quality metrics. Secondly, different people (not professionals) participate in metric verification for grayscale or color images since the opinions of any kind of customer are important and they are processed in a robust manner [47–50]. Meanwhile, the situation is different for RS images as the number of channels can be other than one or three and multichannel data may be visualized in a different manner, and this can influence their perception. It is also difficult to find a lot of people trained to analyze and classify RS images offering their opinions concerning image quality (carrying out image ranking). These factors restrict research directed towards visual metric design and verification for RS images.

Nevertheless, we believe that a preliminary analysis of metrics' applicability for characterization of RS image visual quality can be performed. For this purpose, it is possible to use existing databases containing some images degraded by distortions typical for RS applications. Here, two aspects are worth mentioning. Firstly, a database that has images with all or almost all types of distortions that take place in remote sensing has to be chosen, hence the Tampere Image Database (TID2013) [50] is a good option in this sense (see details in the next Section). Secondly, one can argue that the analysis and perception of traditional color images and RS images are different things. We partly agree with this statement, but it should be stressed that there is an obvious tendency towards convergence of these types of images. For example, digital cameras are installed on modern unmanned aerial vehicles (UAVs) with flight altitude of about 50 m. Hence, there is a question as to whether an acquired image is still a photo or RS data. If there is almost no difference, an application of known quality metrics for RS images seems to be worth investigating.

Taking this into account, TID2013 has been considered in more detail and the types of distortions have been selected for further studies (Section 2). Then, Spearman rank order correlation coefficients (SROCCs) are calculated for many known metrics and the best of them are determined (Section 3). The ways to combine several elementary metrics into combined ones are introduced in Section 4, and their design with the use of trained neural networks is discussed in Section 5. The obtained results are analyzed, and preliminary conclusions are given in Section 6, followed by the examination of computational efficiency (Section 7), whereas Section 8 contains additional verification results. Finally, the conclusions are drawn.

## 2. TID2013 and Some of Its Useful Properties

Since the research community that mainly deals with remote sensing is usually less interested in the latest results in multimedia and color imaging, some important tendencies in the latter area should be briefly described. There are two slightly contradictory desires observed in research and design. On the one hand, good (the best, the most appropriate) HVS-metrics are needed for each particular application where the term "good" or "appropriate" includes a complex of properties, such as: (a) high values of a correlation factor (e.g., SROCC) between a metric and mean opinion score (MOS) obtained as the result of specialized perceptual experiments with volunteers, (b) simplicity and/or high speed of the metric's calculation, (c) the metric's monotonous behavior (e.g., a better visual quality corresponds to a larger metric value) and clear correspondence of visual quality to the metric's value (e.g., it is

often desired to establish what metric value corresponds to invisibility of distortions [61]). On the other hand, there is a desire to create a universal metric capable of performing well for numerous and various types of distortions. In this case, researchers try to get maximal SROCC values for universal databases like TID2013.

In general, TID2013 was created to present a wide variety of distortion types, namely: 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), Sparse sampling and reconstruction (#24). As can be seen, there are types of distortions that are atypical for remote sensing applications (e.g., ##15, 18, 22, 23) whilst there are many types of distortions that are more or less typical for multichannel imaging [62] (see brief discussion in Introduction).

There are 25 test color images in TID2013 and there are five levels of each type of distortion. In other words, there are 25 reference images and 3000 distorted images (120 distorted images for each test image). For each test image, 120 distorted images are partly compared between each other in a tristimulus manner (a better-quality image is chosen among two distorted ones having the corresponding reference image simultaneously presented on the screen). This has been done by many observers (volunteers that have participated in experiments) and the results are jointly processed. As the result, each distorted image has a mean opinion score that can potentially vary from 0 to 9, but, in fact, varies from about 0.2 to about 7.2 (the larger the better).

Many visual quality metrics have been studied for TID2013. A traditional approach to analysis or verification is to calculate a metric value for all distorted images and then to calculate the SROCC or Kendall rank order correlation coefficient (KROCC) [63] between metric values and MOS. SROCC values approaching unity (or −1) show that there is a strict (although possibly nonlinear) dependence between a given metric and MOS and that such a metric can be considered a candidate for practical use. A metric can be considered universal if it provides high SROCC for all considered types of distortion. For example, both PSNR and SSIM provide a SROCC of about 0.63 for a full set of distortion types (see data in Table 5 in [50]). These are, certainly, not the best results since some modern elementary metrics produce a SROCC approaching 0.9 [50]. The term "elementary metric" is further used to distinguish most metrics considered in [50] from the combined and NN-based metrics proposed recently.

A slightly different situation takes place if SROCC is determined for a particular type of distortion or the so-called subsets that include images with several types of distortion. There are several subsets considered in [50] including, for example, the subset called "Color". Considering the similarity of distortions with those encountered in RS images, two subsets are of prime interest. The first one, called "Noise", includes images with 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), Lossy compression of noisy images (#21). The second subset, called "Actual", includes images with distortions ##1, 3, 4, 5, 6, 8, 9, 10 (JPEG compression), 11 (JPEG2000 compression), 19, and 21.

In practice, RS images can be degraded by additive Gaussian noise (it is a conventional noise model for optical images [11]), by noise with different intensities in component images [9], and by spatially correlated noise [64]. Due to image interpolation or deblurring, images corrupted by masked or high-frequency noise can also be met [14]. A quantization noise may occur due to image calibration (range changing). There are also numerous reasons why blur can be observed in RS images [11,14]. Image denoising is a typical stage of image pre-processing [9,10] where specific residual noise can be

observed, whereas a multiplicative noise is typical for SAR images [7,8]. Distortions due to compression take place in many practical cases. Certainly, JPEG and JPEG2000 considered in TID2013 are not the only options [17–19] but they can be treated as representative of the Discrete Cosine Transform (DCT) and wavelet-based compression techniques.

Hence, TID2013 images in general, and the subsets "Noise" and "Actual" in particular, provide a good opportunity for preliminary (rough) analysis of the applicability of existing metrics to the visual quality assessment of RS images. Nevertheless, among the known metrics, only some of them are particularly suitable for color images. The other ones can be determined for grayscale images and their mean value is usually calculated if this metric is applied component-wise to color images. Nevertheless, both types are analyzed since they are of interest for our study.

## 3. Analysis of Elementary Metrics' Performance for TID2013 Subsets

Currently, there is a great number of visual quality metrics. Some of them were designed 10 or even 20 years ago, some other ones were proposed recently. Some can be considered as the modifications of PSNR (e.g., PSNR-HVS-M [65]) or SSIM (e.g., Multi-Scale Structural SIMilarity—MS-SSIM [66]). Some other metrics have been designed based on the other principles. Some metrics are expressed in dB and vary in wide limits, whereas the other ones vary in the range of 0 to 1. There are metrics for which smaller values correspond to better quality (e.g., DCTune [67]) as well. However, a detailed analysis of the principle of operation for all of the elementary metrics considered here is outside of the scope of this paper, since what is most important of all is their performance for the subsets under interest. To avoid the presentation of many mathematical formulas in a long appendix, similarly as in the paper [68], a brief description of the metrics is summarized in Table 1.

The SROCC values for fifty elementary metrics are presented in Table 2 for all types of distortions in descending order, as well as for the subsets "Noise", "Actual", and "Noise&Actual" that includes images with all types of distortions present, at least, in one subset. As can be seen, there are a few quite universal metrics, e.g., the Mean Deviation Similarity Index (MDSI) [69], Perceptual SIMilarity (PSIM) [70], and Visual Saliency-Induced Index (VSI) [71], for which SROCC almost reaches 0.9. As expected, the SROCC values for subsets are larger than for all types of distortions. The "champion" for the subset "Noise" is MDSI (SROCC = 0.928), the best results for the subset "Actual" (SROCC = 0.939) are provided by several metrics (MDSI, PSNRHA [72], PSNRHMAm [73]); for both subsets together, the largest SROCC (equal to 0.937) is again provided by MDSI. Nevertheless, the performance of some of the metrics applied for color images is dependent on the method of color to grayscale conversion. However, the results presented in the paper for the NN-based metrics have been obtained assuming the same type of RGB to YCbCr conversion for all individual metrics, using the first component Y as the equivalent of the greyscale image. Such results can be considered appropriate for many practical applications. Nevertheless, it is still interesting if SROCC values can be further improved.

## 4. Design of Combined Image Quality Metrics

The idea of combined (hybrid) metrics for general purpose IQA assumes that different metrics utilize various kinds of image data and therefore different features and properties of images may be used in parallel. Therefore, one may expect good results for a combination of metrics that come from various "families" of metrics, complementing each other. This assumption has been initially motivated by the construction of the SSIM formula [41] where three factors representing luminance, contrast and structural distortions are multiplied. Probably the first approach to the combined full-reference metrics [54] has been proposed as the nonlinear combination of three metrics: MS-SSIM [66], Visual Information Fidelity (VIF) [74] and Singular Value Decomposition-based R-SVD [75], and verified for TID2008 [76], leading to SROCC = 0.8715 for this dataset. Nevertheless, in this research, the optimization goal has been chosen as the maximization of the Pearson's linear correlation coefficient (PCC) without the use of nonlinear fitting. The proposed combined metric is the product of three above mentioned metrics raised to different powers with optimized exponents.

Since the correlation of the R-SVD metric with subjective quality scores (MOS values) is relatively low, similarly to MSVD [77] (see Table 2), better results may be obtained by replacing it with Feature SIMilarity (FSIM) [78], leading to the Combined Image Similarity Metric (CISI) [55] with SROCC = 0.8742 for TID2008, although again optimized towards the highest PCC. Nevertheless, the highest SROCC obtained by optimizing the CISI weights for the whole TID2013 is equal to 0.8596. Another modification [56], leading to SROCC = 0.9098 for TID2008, has been based on four metrics, where FSIMc has been improved by the use of optimized weights for gradient magnitude and phase congruency components with added RFSIM [79].

**Table 1.** Brief descriptions of the elementary metrics used in experiments.

| Year | Full-Reference Metric | Description |
|------|----------------------|-------------|
| - | MSE | Classical pixel-based mean square error |
| - | PSNR | Peak signal-to-noise ratio based on MSE expressed in dB |
| 1993 | WSNR [80] | Classical SNR with weighting appropriate to the HVS |
| 2000 | NQM [81] | Noise Quality Measure based on degradation model |
| 2002 | UQI [42] | Universal Image Quality Index measuring the local differences in mean, standard deviation and correlation |
| 2004 | SSIM [41] | Structural Similarity being the improved version of UQI |
| 2003 | MS-SSIM [66] | Multi-scale version of SSIM |
| 2005 | IFC [82] | Information Fidelity Criterion utilizing the Natural Scene Statistics (NSS) originating from statistical models in an information-theoretic setting |
| 2006 | VIF [74] | Visual Information Fidelity metric based on the measurement of the mutual information between the subbands of the wavelet transform for reference and distorted images |
| 2006 | VIFP [74] | Pixel-domain version of VIF with lower computational cost |
| 2006 | QILV [83] | Quality Index based on Local Variance (SSIM-inspired) |
| 2006 | MSVD [77] | Metric based on the Singular Value Decomposition of Image Blocks |
| 2006 | PSNRHVS [84] | PSNR that takes into account less sensitivity of HVS to distortions in high spatial frequencies |
| 2007 | PSNRHVSM [65] | PSNR that takes into account less sensitivity of HVS to distortions in high spatial frequencies and masking effect |
| 2007 | VSNR [85] | Visual SNR-based metric calculated in the wavelet domain |
| 2009 | CWSSIM [86] | SSIM-based metric utilizing similarity of components in complex wavelet domain (Complex Wavelet SSIM) |
| 2010 | RFSIM [79] | Riesz-transform based Feature SIMilarity (extension of SSIM) with feature mask defined as the edge regions of the image considered as key locations |
| 2011 | FSIM [78] | Feature Similarity metric based on phase congruency and gradient magnitude measured similarly as in the SSIM |
| 2011 | FSIMc [78] | Color version of FSIM with additional use of YIQ chrominance components |
| 2011 | IWSSIM [87] | SSIM-based metric utilizing the information content weighting |

**Table 1.** *Cont.*

| Year | Full-Reference Metric | Description |
|------|----------------------|-------------|
| 2011 | PSNRHAy [72] | Modification of PSNRHVS that takes into account contrast change and mean shift of gray scale image |
| 2011 | PSNRHMAy [72] | Modification of PSNRHVSM that takes into account contrast change and mean shift of gray scale image |
| 2011 | PSNRHA [72] | Modification of PSNRHVS that takes into account contrast change and mean shift of color image |
| 2011 | PSNRHMA [72] | Modification of PSNRHVSM that takes into account contrast change and mean shift of color image |
| 2011 | ADM [88] | Metric incorporating detail losses and additive impairments with HVS processing |
| 2012 | GSM [89] | SSIM-based metric utilizing the similarity of gradient images |
| 2012 | SR-SIM [90] | Spectral Residual SIMilarity metric using a specific visual saliency map closely related to perceived quality |
| 2013 | ESSIM [91] | Metric based of the similarity of edge strength |
| 2013 | SFF [92] | Sparse Feature Fidelity metric utilizing Independent Component Analysis (ICA) for training the features detector, using visual attention and visual threshold |
| 2013 | WASH [93] | Wavelet-based SHarp features metric utilizing the high sensitivity of HVS to sharpness and zero-crossings |
| 2013 | IGM [94] | Metric utilizing the Internal Generative Mechanism decomposing the image into predicted visual content and the disorderly portion with the residual content |
| 2014 | VSI [71] | Visual Saliency-Induced Index using the weighting function to reflect the importance of a local region |
| 2014 | GMSD [95] | Gradient Magnitude Similarity Deviation based on global variation of gradient-based local quality map |
| 2014 | IQM2 [96] | SSIM-inspired metric utilizing pyramid wavelet transform |
| 2015 | DSS [97] | DCT Subband Similarity metric measuring and weighting changes in structural information in subbands |
| 2016 | MCSD [98] | Multiscale Contrast Similarity Deviation with a resorting of the contrast features to multiscale representation |
| 2016 | ADD_SSIM [99] | Analysis of Distortion Distribution affected by image content and distortion used for pooling with SSIM |
| 2016 | ADD_GSIM [99] | Analysis of Distortion Distribution affected by image content and distortion used for pooling with GSIM |
| 2016 | UNIQUE [100] | Unsupervised Image Quality Estimation utilizing color space transformation, mean subtraction, whitening and linear decoder to obtain sparse representations |
| 2017 | MS-UNIQUE [101] | Multi-model and Sharpness-weighted Unsupervised Image Quality Estimation (extension of UNIQUE metric) |
| 2017 | PSNRHMAm [73] | Modification of PSNRHMA that includes updated HVS-model and masking effect |
| 2017 | PSIM [70] | Perceptual SIMilarity based on fusion of micro- and macro-structures and color information similarity |
| 2018 | HaarPSI [102] | Haar Wavelet-based Perceptual Similarity Index |
| 2018 | RVSIM [103] | Riesz transform and Visual contrast sensitivity-based feature SIMilarity index utilizing Log-Gabor filter and monogenic signal similarity matrix |

**Table 1.** *Cont.*

| Year | Full-Reference Metric | Description |
|---|---|---|
| 2018 | CVSSI [104] | Contrast and Visual Saliency Similarity-Induced Index using local contrast quality map and the global vs. quality map with their weighted standard deviation |
| 2018 | SSIM4 [105] | Modification of SSIM that incorporates HVS ability to reconstruct image fragments from non-local similarity |
| 2018 | CSSIM [105] | Modification of SSIM for color images |
| 2018 | CSSIM4 [105] | Modification of SSIM for color images that incorporates HVS ability to reconstruct image fragments from non-local similarity |
| 2018 | DSI [106] | DisSimilarity Index for low contrast and noise-like textured images considering the masking effect of non-predictable energy of image regions |

**Table 2.** Spearman rank order correlation coefficient (SROCC) values for all distortion types and three subsets of the Tampere Image Database (TID2013).

| | Full-Reference Metric | All Distortions | Noise | Actual | Noise & Actual |
|---|---|---|---|---|---|
| 1 | VSI [71] | 0.8967 | 0.9101 | 0.9258 | 0.9218 |
| 2 | PSIM [70] | 0.8926 | 0.9189 | 0.9309 | 0.9303 |
| 3 | MDSI [69] | 0.8897 | 0.9275 | 0.9387 | 0.9374 |
| 4 | HaarPSI [102] | 0.8730 | 0.9063 | 0.9168 | 0.9190 |
| 5 | MS-UNIQUE [101] | 0.8708 | 0.8812 | 0.8977 | 0.8981 |
| 6 | UNIQUE [100] | 0.8600 | 0.8607 | 0.8813 | 0.8799 |
| 7 | PSNRHMAm [73] | 0.8541 | 0.9221 | 0.9387 | 0.9315 |
| 8 | SFF [92] | 0.8518 | 0.8787 | 0.9059 | 0.8946 |
| 9 | FSIMc [78] | 0.8510 | 0.9022 | 0.9150 | 0.9164 |
| 10 | CSSIM [105] | 0.8417 | 0.8728 | 0.8878 | 0.8892 |
| 11 | ADD_GSIM [99] | 0.8310 | 0.9023 | 0.9151 | 0.9159 |
| 12 | PSNRHA [72] | 0.8198 | 0.9230 | 0.9388 | 0.9322 |
| 13 | PSNRHMA [72] | 0.8137 | 0.9151 | 0.9343 | 0.9250 |
| 14 | ESSIM [91] | 0.8121 | 0.8411 | 0.8866 | 0.8619 |
| 15 | CVSSI [104] | 0.8090 | 0.9248 | 0.9350 | 0.9341 |
| 16 | SR-SIM [90] | 0.8076 | 0.9070 | 0.9211 | 0.9206 |
| 17 | MCSD [98] | 0.8045 | 0.9224 | 0.9326 | 0.9323 |
| 18 | GSM [89] | 0.8028 | 0.8408 | 0.8841 | 0.8583 |
| 19 | IGM [94] | 0.8023 | 0.9099 | 0.9220 | 0.9227 |
| 20 | ADD_SSIM [99] | 0.8023 | 0.9008 | 0.9119 | 0.9120 |
| 21 | FSIM [78] | 0.8011 | 0.8969 | 0.9108 | 0.9117 |
| 22 | GMSD [95] | 0.8004 | 0.9211 | 0.9314 | 0.9318 |
| 23 | IQM2 [96] | 0.7955 | 0.8995 | 0.9103 | 0.9122 |
| 24 | DSS [97] | 0.7915 | 0.8766 | 0.8904 | 0.8945 |
| 25 | ADM [88] | 0.7861 | 0.9113 | 0.9201 | 0.9189 |
| 26 | PSNRHAy [72] | 0.7794 | 0.9184 | 0.9272 | 0.9275 |
| 27 | MS-SSIM [66] | 0.7872 | 0.8733 | 0.8872 | 0.8881 |
| 28 | IWSSIM [87] | 0.7775 | 0.8783 | 0.8934 | 0.8937 |
| 29 | RFSIM [79] | 0.7721 | 0.8731 | 0.8793 | 0.8847 |
| 30 | SSIM4 [105] | 0.7657 | 0.8743 | 0.8887 | 0.8902 |
| 31 | PSNRHMAy [72] | 0.7570 | 0.9107 | 0.9209 | 0.9226 |
| 32 | CSSIM4 [105] | 0.7394 | 0.8432 | 0.8668 | 0.8604 |
| 33 | DSI [106] | 0.7114 | 0.7493 | 0.7801 | 0.7838 |
| 34 | VIF [74] | 0.6816 | 0.8422 | 0.8585 | 0.8532 |
| 35 | VSNR [85] | 0.6809 | 0.8691 | 0.8817 | 0.8815 |
| 36 | RVSIM [103] | 0.6748 | 0.8192 | 0.8449 | 0.8423 |

**Table 2.** *Cont.*

| | Full-Reference Metric | All Distortions | Noise | Actual | Noise & Actual |
|---|---|---|---|---|---|
| 37 | PSNRHVS [84] | 0.6536 | 0.9172 | 0.9257 | 0.9263 |
| 38 | PSNR | 0.6396 | 0.8217 | 0.8246 | 0.8335 |
| 39 | MSE | 0.6396 | 0.8217 | 0.8246 | 0.8335 |
| 40 | SSIM [41] | 0.6371 | 0.7574 | 0.7877 | 0.7812 |
| 41 | NQM [81] | 0.6349 | 0.8362 | 0.8572 | 0.8527 |
| 42 | PSNRHVSM [65] | 0.6246 | 0.9061 | 0.9175 | 0.9188 |
| 43 | VIFP [74] | 0.6084 | 0.7835 | 0.8151 | 0.8056 |
| 44 | QILV [83] | 0.5975 | 0.6604 | 0.6948 | 0.7049 |
| 45 | WSNR [80] | 0.5796 | 0.8804 | 0.8966 | 0.8952 |
| 46 | CWSSIM [86] | 0.5551 | 0.7943 | 0.8160 | 0.8051 |
| 47 | UQI [42] | 0.5444 | 0.6482 | 0.6904 | 0.6824 |
| 48 | IFC [82] | 0.5229 | 0.7201 | 0.7598 | 0.7468 |
| 49 | WASH [93] | 0.2903 | 0.3290 | 0.4173 | 0.3139 |
| 50 | MSVD [77] | 0.1261 | 0.1123 | 0.1309 | 0.1424 |

Another idea [57], utilizing the support vector regression approach to the optimization of PCC for five databases with additional context classification for distortion types, has achieved SROCC = 0.9495 for seven combined metrics with SROCC = 0.9403 using four of them. Nevertheless, al these results have been obtained for the less demanding TID2008 [76], being the earlier version of TID2013 [50], containing only 1700 images (in comparison to 3000) with a smaller number of distortion types and levels.

Recently, another approach to combined metrics, based on the use of the median and alpha-trimmed mean of up to five initially linearized metrics, has been proposed [53]. The best results obtained for TID2013 are SROCC = 0.8871 for the alpha-trimmed mean of five metrics: Information Fidelity Criterion (IFC) [82], DCTune [67], FSIMc [78], Sparse Feature Fidelity (SFF) [92] and PSNRHMAm [73]. Slightly worse results (SROCC = 0.8847) have been obtained using the median of nearly the same metrics (only replacing IFC [82] with a pixel-based version of Visual Information Fidelity—VIFP [74]).

## 5. Neural Network Design and Training for the Considered Subsets

In recent years, neural networks (NN) have demonstrated a very high potential in solving many tasks related to image processing. Their use is often treated as a remedy to get benefits in design and performance improvement. Hence, the peculiarities and possibilities of the NN use for our application are briefly considered, i.e., in the design of new, more powerful full-reference metrics for images with the aforementioned types of distortions. For this purpose, the requirements for such metrics are recalled below.

A good NN-based metric should provide a reasonable advantage in performance compared to elementary metrics. Since we deal with SROCC as one quantitative criterion of metric performance, it should be considerably improved compared to the already reached values of 0.93 . . . 0.94. Since the maximal value of SROCC is unity, its improvement by 0.02 . . . 0.03 can be considered as sufficient. The other relevant aspects are input parameters and NN structure. Since a typical requirement for a full-reference metric is to perform rather fast, input parameters should be calculated easily and quickly. Certainly, their calculation can be done in parallel or accelerated somehow, but anyway none of the input parameters should be too complex. The structure of a used NN should possibly be simple as well. A smaller number of hidden layers and fewer neurons in them without a loss of performance are desired. A smaller number of input parameters can also be advantageous.

A brief analysis of existing solutions shows the following:

1.  Neural networks have already been used in the design of full-reference quality metrics (see, e.g., [58,107–110]); the metric [107] employs feature extraction from reference and distorted images and uses deep learning in convolution metrics design, providing SROCC = 0.94 for all types of

distortions in TID2013; E. Prashnani et al. [108] have slightly improved the results of [107] due to exploiting a new pairwise-learning framework; Seo et al. [109] reached SROCC = 0.961 using deep learning;

2.  There can be different structures of NNs (despite the popularity of convolutional networks, standard multilayer ones can still be effective enough) and different sets of input parameters (both certain features and elementary metrics can be used).

Keeping this in mind, our idea is to use a set of elementary quality metrics as inputs and apply the NNs with a quite simple structure for solving our task—to get a combined metric (or several combined metrics) with performance sufficiently better than for the best elementary metric. Then, a set of particular tasks to be solved arises, namely:

*   how many elementary metrics should be used?
*   what elementary metrics should be employed?
*   what structure of the NN should be chosen and how to optimize its parameters?
*   are some pre-processing operations of input data needed?
*   what should be the NN output and its properties?

Regarding the last question, since we plan to exploit TID2013 in our design, it should be recalled that the quality of images in this database is characterized by mean opinion score (MOS). The main properties of MOS in TID2013 are determined by the methodology of experiments carried out by observers. Potentially, it was possible that MOS could be from 0 to 9, but, as the result of experiments, MOS varies in the limits from 0.24 to 7.21 [53]. Moreover, the analysis of MOS and image quality [53] has shown that four gradations of image quality are possible with respect to MOS:

1.  Excellent quality (MOS > 6.05);
2.  Good quality (5.25 < MOS ≤ 6.05);
3.  Middle quality (3.94 < MOS ≤ 5.25);
4.  Bad quality (MOS ≤ 3.94).

This "classification" is a little bit subjective, hence some explanations are needed. Images are considered to have excellent quality if distortions in them cannot be visually noticed. For images with good quality, MOS values have the ranks from 201 to 1000 and distortions can be noticed by careful visual inspection. If MOS values have the ranks from 1001 to 2000, the image quality is classified as middle (the distortions are visible, but they are not annoying). The quality of the other images is conditionally classified as bad—the distortions are mostly annoying.

Figure 1 illustrates the examples of distorted images for the same reference image (#16 in TID2013) that has neutral content and is of a medium complexity, being similar to RS images. As it has been stated earlier, there is an obvious tendency to convergence of these types of images. The values of MOS and three elementary metrics are presented for them as well. The image in Figure 1a corresponds to the first group (excellent quality) and it is really difficult to detect distortions. The image in Figure 1b belongs to the second group and distortions are visible, especially in homogeneous image regions. The image in Figure 1c is a good representative of the third group of images for which distortions are obvious but they are not annoying yet. Finally, the image in Figure 1d is an example of a bad quality image. The presented values of metrics show how they correspond to quality degradation and can characterize the distortion level. As may be noticed, not all presented metrics reflect the subjective quality perfectly, e.g., there is a higher PSNR value for the bad quality image (Figure 1d) than for the middle quality image. Similarly, PSNRHA and MDSI metrics do not correspond to their MOS values for almost unnoticeable distortions, i.e., excellent and good quality.
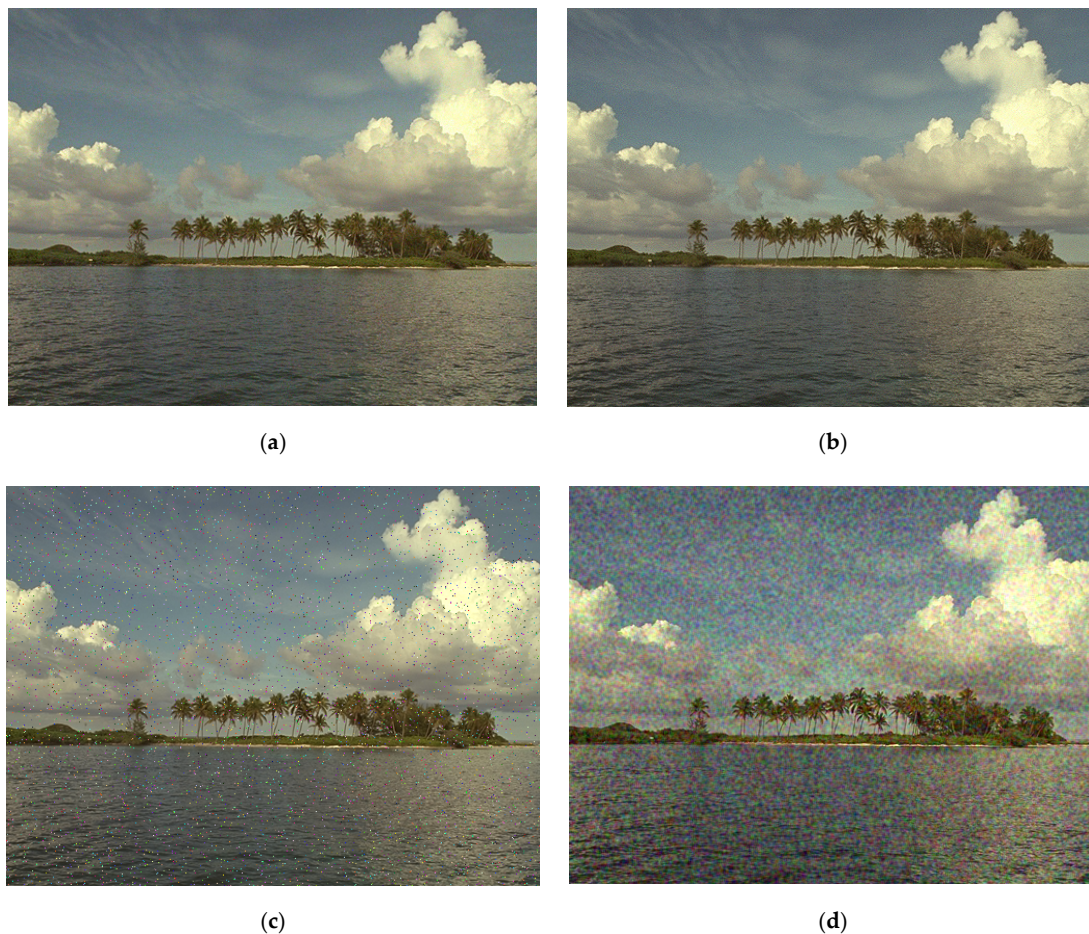
**Figure 1.** Examples of the test image with different quality: (**a**) excellent (distortion type #2—additive noise in color components)—mean opinion score (MOS) = 6.15, peak signal-to-noise ratio (PSNR) = 40.773 dB, PSNRHA = 36.935 dB, Mean Deviation Similarity Index (MDSI) = 0.19; (**b**) good (distortion type #5—high frequency noise)—MOS = 5.6154, PSNR = 34.799 dB, PSNRHA = 39.5 dB, MDSI = 0.1892; (**c**) middle (distortion type #6—impulse noise)—MOS = 3.9737, PSNR = 27.887 dB, PSNRHA = 29.208 dB, MDSI = 0.3413; (**d**) bad (distortion type #3—spatially correlated noise)—MOS = 3.2368, PSNR = 28.842 dB, PSNRHA = 26.1 dB, MDSI = 0.3887.

Taking this into account, it has been decided that the NN output should be in the same limits as the MOS. This means that error minimization with respect to MOS can be used as the target function in the NN training. If needed, MOS (NN output) can be easily recalculated to another scale like, e.g., from 0 to 1.

The penultimate question concerns input pre-processing. It is known that it is often recommended in the NN theory to carry out some preliminary normalization of input data (features) if they have different ranges of variation [61]. This is true for elementary metrics, as for example, PSNR and PSNR-HVS-M, both expressed in dB, can vary in wide limits (even from 10 dB to 60 dB) but, for distorted images in TID2013, they vary in narrower limits. PSNR used for setting five levels of distortions, has five values approximately reached (21, 24, 27, 30, and 33 dB) although, in fact, the values of this metric for images in TID2013 vary from about 13 dB to ≈41 dB. Consecutively, PSNR-HVS-M varies from 14 dB to 59 dB, which mainly corresponds to "operation limits" starting from very annoying distortions to practically perfect quality (invisible distortions). The MDSI varies from 0.1 to 0.55 for images in TID2013 where the larger values correspond to lower visual quality. Some other metrics like SSIM, MS-SSIM, and FSIM vary in the limits from 0 to 1, where the latter limit corresponds to perfect quality. In fact, most values of these metrics are concentrated in the upper third part of

this interval (see the scatter plot in Figure 2 for the color version of the metric FSIM, referred to as FSIMc). An obvious general tendency of the increase in FSIMc when MOS becomes larger is observed. Meanwhile, two important phenomena may also be noticed. Firstly, there is a certain diversity of the metric's values for the same MOS. Secondly, the dependence of FSIMc on MOS (or MOS on FSIMc) is nonlinear.
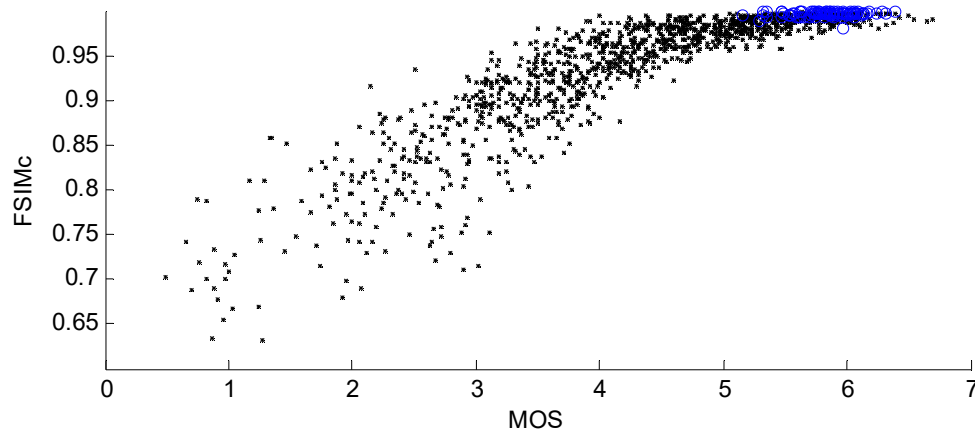


**Figure 2.** The scatter plot of color version of Feature SIMilarity (FSIMc) vs. MOS (blue circles correspond to images for which distortions have not been detected).

In this sense, linearization (fitting) is often used to get not only the high values of SROCC but the conventional (Pearson) correlation factor (coefficient) as well [69]. Then, two hypotheses are possible. The first one is that the NN, being nonlinear and able to adapt to peculiarities of input data, will "manage" this nonlinearity of input–output dependence "by itself" (denote this hypothesis as H1). The second hypothesis (H2) is that elementary metric pre-processing in the form of fitting can be beneficial for further improvement of combined metric performance (optimization).

Concerning fitting needed for realization of H2, there are several commonly accepted options. One of them is to apply the Power Fitting Function (PFF) $y(x) = a \cdot x^b + c$, where a, b, and c are the adjustable parameters. The fitting results can be characterized by the root mean square error (RMSE) of the scatter plot points after fitting with respect to the fitted curve (the smaller RMSE, the better). The results obtained for the elementary metrics considered above (in Table 2) are given in the left part of Table 3. As one can see, the best fitting result (the smallest RMSE) is obtained for the metric MDSI (it equals to 0.3945). The results for the metrics that are among the best according to SROCC, e.g., Contrast and Visual Saliency Similarity Induced index (CVSSI), Multiscale Contrast Similarity Deviation (MCSD), PSNRHA, Gradient Magnitude Similarity Deviation (GMSD), see data in Table 2, are almost equally good, too.

Another fitting model that can be applied is, e.g., Poly2 $y(x) = p_1 \cdot x^2 + p_2 \cdot x + p_3$, where $p_1$, $p_2$, and $p_3$ are parameters to be adjusted to produce the best fit. The results are very similar. For example, for MDSI, the minimal RMSE equal to 0.3951 is observed. For PSNRHA, RMSE = 0.4013 is achieved, i.e., slightly better than using the PFF. The detailed data obtained for Poly2 fit are presented in the right part of Table 3. For clarity, the better of the two RMSE results are marked by the bold format in Table 3. Then, three options are possible in combined NN design: 1) to use the PFF for all elementary metrics; 2) to apply Poly2 for all elementary metrics; 3) to choose the best fitting for each elementary metric and to apply it.

Earlier in [53], different possible monotonous functions including linear fitting have been considered for combining elementary metrics. As practice has shown, they produce worse results with higher RMSE in comparison to PFF and Poly2.

The next considered issue is what structures and parameters of NNs can be chosen and optimized. As it has been noted above, we concentrate on conventional structures. We prefer to apply the

multilayer NN instead of a deep learning approach because in this way many design aspects are clear. It prevents the overtraining problems and allows for a good generalization. In particular, the number of neurons in the input layer can be equal to the number of elementary metrics used. In addition, this solution can be easily adopted by using TID2013, avoiding the need for a large training set of data.

Concerning this number, different options are possible, including the use of all 50 considered elementary metrics. It is also possible to restrict to the best metrics from Table 2 and employ elementary metrics that have certain properties, for example, metrics that have SROCC > 0.9 for the considered types of distortions (there are 22 such metrics), SROCC > 0.92 (14 metrics), or SROCC > 0.93 (7 metrics).

Nevertheless, the theory of NNs states that it is reasonable to apply such input parameters that can "add" or "complement" information to each other, i.e., are not highly correlated. One possible approach is the evaluation of cross-correlation function to determine the similarity between all pairs of metrics and after that to exclude the worst in highly correlated pairs.

Instead of that, in this paper, a very useful approach called Lasso regularization [111] (Lasso is an abbreviation for Least Absolute Shrinkage and Selection Operator) has been applied for the selection of the most unique metrics for NNs. In machine learning, Lasso and Ridge regularizations are used to introduce additional limitations to the model and decrease the problem of overfitting.

The key feature of Lasso is that this method may introduce zero weights for the "noisy" and the least important data. For the task of metrics combination, it means that Lasso regularization can determine the elementary metrics that are the least useful for combining and leave the other ("most informative") ones. This study adopts the implementation of Lasso available in MATLAB® so that, for different thresholds, zero coefficients for metrics that can be excluded can be estimated. According to that, the number of such non-zero values (NNZ) for each metric has been determined, assuming the following conditions: (a) NNZ > 20 (out of 100 values); (b) NNZ > 30; (c) NNZ > 40; (d) NNZ > 50; (e) NNZ > 60.

**Table 3.** Metric fitting results and parameters for the Power Fitting Function (PFF) and second order polynomial (Poly2); better RMSE values are shown in bold format.

| Metric | PFF – $y(x) = a \cdot x^b + c$ | | | | Poly2 – $y(x) = p_1 \cdot x^2 + p_2 \cdot x + p_3$ | | | |
|---|---|---|---|---|---|---|---|---|
| | RMSE | a | b | c | RMSE | p1 | p2 | p3 |
| PSNR | 0.6830 | −126.636 | −0.0554 | 109.1416 | **0.6777** | −0.00447 | 0.47070 | −5.81476 |
| MSE | **0.6811** | −2.4997 | 0.1690 | 9.3328 | 0.6822 | 0.00002 | −0.01776 | 5.59889 |
| WSNR [80] | 0.5216 | 103.726 | 0.0372 | −113.987 | **0.5155** | −0.00221 | 0.28682 | −2.91023 |
| NQM [81] | 0.6037 | 0.8643 | 0.5904 | −1.3442 | **0.6004** | −0.00189 | 0.23355 | −0.18685 |
| UQI [42] | **0.8393** | 6.5341 | 0.4358 | −0.8949 | 0.8411 | −2.21482 | 6.68761 | 1.12020 |
| SSIM [41] | 0.7645 | 3.2558 | 4.5929 | 2.8783 | **0.7611** | 11.54027 | −11.10435 | 5.57541 |
| MS–SSIM [66] | **0.5536** | 3.5705 | 12.6686 | 2.2151 | 0.5828 | 58.28322 | −87.43399 | 34.73077 |
| IFC [82] | **0.7404** | 30.6531 | 0.0374 | −27.7813 | 0.7679 | −0.03235 | 0.66730 | 2.26364 |
| VIF [74] | **0.5600** | 4.9142 | 0.6852 | 1.1082 | 0.5623 | −1.19541 | 5.56502 | 1.65296 |
| VIFP [74] | 0.6514 | 11.1695 | 0.2507 | −4.6922 | **0.6450** | −6.69260 | 11.72200 | 0.58570 |
| MSVD [77] | 1.2091 | 463.362 | −1.7245 | 4.0580 | 1.2091 | 0.00010 | −0.02582 | 5.67944 |
| QILV [83] | **0.8924** | 2.1258 | 12.8721 | 3.2304 | 0.9399 | 3.37891 | −0.84093 | 2.48204 |
| VSNR [85] | 0.5505 | −40.9961 | −0.1030 | 33.6452 | **0.5445** | −0.00271 | 0.27019 | −0.85345 |
| PSNRHVS [84] | 0.4176 | −54.0574 | −0.4961 | 14.4993 | **0.4116** | −0.00463 | 0.45391 | −4.93928 |
| PSNRHVSM [65] | 0.4472 | −78.8179 | −0.7655 | 9.8824 | **0.4356** | −0.00347 | 0.37023 | −3.98164 |
| CWSSIM [86] | **0.7294** | 2.4648 | 1858.3573 | 3.0546 | 0.9874 | 4040.905 | −7844.04 | 3807.931 |
| RFSIM [79] | 0.5086 | 4.9998 | 0.8521 | 1.0324 | **0.5085** | −0.46671 | 5.20948 | 1.29949 |
| PSNRHAy [72] | 0.4114 | −79.7808 | −0.7054 | 11.7318 | **0.4085** | −0.00524 | 0.49796 | −5.72493 |
| PSNRHMAy [72] | 0.4332 | −120.791 | −0.9339 | 9.0571 | **0.4247** | −0.00380 | 0.39867 | −4.58245 |
| PSNRHA [72] | 0.4057 | −271.292 | −1.1468 | 9.8101 | **0.4013** | −0.00726 | 0.66096 | −8.98189 |
| PSNRHMA [72] | 0.4296 | −437.562 | −1.3364 | 8.5281 | **0.4186** | −0.00590 | 0.57515 | −8.11850 |
| FSIM [78] | **0.4813** | 3.6227 | 13.3796 | 2.1510 | 0.5021 | 86.92251 | −139.2986 | 58.00009 |
| FSIMc [78] | **0.4699** | 3.7012 | 11.1900 | 2.1265 | 0.4821 | 71.85416 | −112.4054 | 46.25250 |
| IWSSIM [87] | **0.5336** | 3.4602 | 11.7455 | 2.2851 | 0.5653 | 48.88158 | −71.94606 | 28.60147 |
| ADM [88] | 0.5133 | 5.2079 | 1.0584 | 1.1913 | **0.5116** | −0.17437 | 5.53250 | 0.98740 |

**Table 3.** *Cont.*

| Metric | PFF – $y(x) = a \cdot x^b + c$ | | | | Poly2 – $y(x) = p_1 \cdot x^2 + p_2 \cdot x + p_3$ | | | |
|---|---|---|---|---|---|---|---|---|
| | RMSE | a | b | c | RMSE | p1 | p2 | p3 |
| GSM [89] | **0.6085** | 3.8941 | 62.8385 | 2.0375 | 0.6108 | 2510.109 | −4831.844 | 2327.548 |
| IGM [94] | **0.4154** | 4.8469 | 4.6108 | 0.3577 | 0.4183 | 27.89277 | −34.32709 | 11.64504 |
| SR−SIM [90] | **0.4485** | 3.6224 | 26.1269 | 2.1720 | 0.4774 | 303.1233 | −541.8017 | 244.300 |
| SFF [92] | **0.5037** | 4.3154 | 17.0200 | 1.4609 | 0.5255 | 189.9484 | −324.4132 | 140.106 |
| GMSD [95] | **0.4201** | −10.7552 | 0.5948 | 6.2403 | 0.4267 | 61.89223 | −30.21982 | 5.84554 |
| ESSIM [91] | **0.5862** | 4.1907 | 794.155 | 1.7213 | 0.5924 | 469,230.2 | −935,844.8 | 466,620.3 |
| WASH [93] | 1.1077 | −3.9034 | −0.8187 | 9.1013 | **1.0504** | −29.57805 | 52.04451 | −18.07274 |
| VSI [71] | **0.4577** | 3.5501 | 37.8918 | 2.3297 | 0.4783 | 583.2767 | −1076.422 | 498.8442 |
| IQM2 [96] | 0.4716 | 3.6621 | 1.2572 | 2.1072 | **0.4685** | 1.16719 | 2.58221 | 2.04645 |
| DSS [97] | **0.5109** | 3.3472 | 3.2571 | 2.2928 | 0.5270 | 5.36353 | −2.05047 | 2.23715 |
| ADD_GSIM [99] | **0.4544** | 3.7670 | 155.109 | 1.9912 | 0.4883 | 10,718.78 | −21,043.36 | 10,330.15 |
| ADD_SSIM [99] | **0.4705** | 3.9271 | 142.763 | 1.8464 | 0.4977 | 10,885.49 | −21,376.10 | 10,496.23 |
| MCSD [98] | **0.4123** | −10.6097 | 0.5096 | 6.4149 | 0.4353 | 90.00587 | −36.64613 | 5.78630 |
| MDSI [69] | **0.3945** | −17.9432 | 1.6790 | 6.7208 | 0.3951 | −15.49132 | −3.91786 | 6.91601 |
| UNIQUE [100] | **0.6600** | 4.3775 | 0.4905 | 1.3221 | 0.6677 | −1.93108 | 5.38180 | 2.21278 |
| MS−UNIQUE [101] | **0.6478** | 4.3483 | 0.6346 | 1.4211 | 0.6499 | −1.21208 | 4.97049 | 2.01005 |
| PSIM [70] | **0.4144** | 3.9219 | 583.789 | 1.9809 | 0.4268 | 215,288.1 | −428,850.3 | 213,567.9 |
| CVSSI [104] | **0.4137** | −10.9029 | 0.5469 | 6.3852 | 0.4363 | 63.31016 | −32.15855 | 5.81023 |
| PSNRHMAm [73] | 0.4086 | −258.9957 | −1.1396 | 9.7006 | **0.4042** | −0.00702 | 0.64070 | −8.57968 |
| DSI [106] | **0.7916** | −2.4387 | 0.1568 | 7.6098 | 0.8908 | 0.00027 | −0.06153 | 5.10611 |
| CSSIM [105] | **0.5238** | 3.3778 | 19.8756 | 2.2658 | 0.5439 | 152.6187 | −260.5828 | 113.4502 |
| CSSIM4 [105] | **0.6121** | 2.8845 | 72.9625 | 2.6397 | 0.7174 | 351.6808 | −635.0208 | 288.5030 |
| SSIM4 [105] | **0.5315** | 3.2548 | 9.2413 | 2.3957 | 0.5624 | 27.46582 | −35.54652 | 13.56676 |
| HaarPSI [102] | **0.4347** | 4.5283 | 2.1854 | 1.4475 | 0.4357 | 5.15546 | −0.66241 | 1.46847 |
| RVSIM [103] | **0.5678** | 5.5441 | 1.1196 | −0.0078 | 0.5679 | 0.48043 | 5.25852 | −0.20338 |

Since the network output is the combined metric, one neuron is present in the output layer, but the number of hidden layers and the number of neurons in each hidden layer can be different. Two variants have been analyzed—two and four hidden layers. Furthermore, two variants of the numbers of neurons in the hidden layers have been considered, namely an equal number of neurons in each hidden layer and a twice smaller number of neurons in each successive hidden layer.

As the activation function in hidden layers, the hyperbolic tangent sigmoid transfer function is used, which also provides the normalization in the range (−1,1), whereas a linear function is used for the output layer.

An example of the NN structure is presented in Figure 3. In this case, there are nine inputs and the number of neurons in hidden layers decreases gradually.
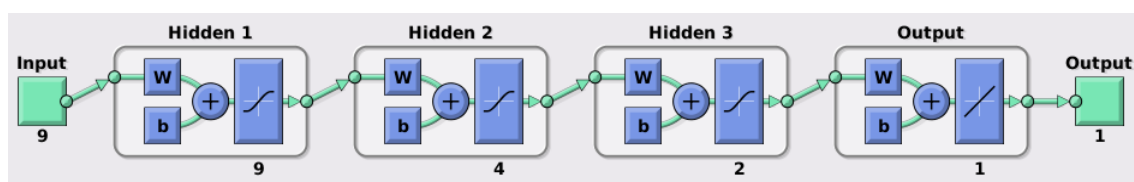


**Figure 3.** An exemplary structure of the neural network with 9 inputs used in experiments.

In addition to the NN structure, some other factors might influence its performance, such as the methodology of NN training and testing, stability of training results, the number of epochs, etc. Due to the limited set of 1625 images for the NN training and testing, this constraint complicates our task. According to a traditional methodology, an available set should be divided into training and verification ones in some proportion. In the conducted experiments, 70% of images have been used for training and the remaining images for verification. Since the division of images into sets is random, training and verification results can be random as well. To partly get around this uncertainty, the best

data are presented below for each version of the trained NN (producing the largest SROCC at training stage) of a given structure.

Concerning the training stage—each NN has been trained to provide as high SROCC as possible. Obviously, some other training strategies are also possible. In particular, it is possible to use the Pearson correlation coefficient (PCC) to be maximized. Nevertheless, the results of the additional use of the PCC for the training are not considered in this paper.

To give more understanding of which NN structures have been analyzed and what parameters have been used, the main characteristics are presented in Table 4, where the number of input elementary metrics is provided in each case. As can be seen, there are quite a lot of possible NN structures.

**Table 4.** The numbers of elementary metrics used as the inputs for different configurations of neural networks (NN)s under conditions of different fitting and restrictions imposed on an elementary metric set.

| Pre-Processing Setup | Elementary Metrics Used | Without Fitting | Power2 | Best of PFF and Poly2 |
|---|---|---|---|---|
| ABSENT (NO LASSO) | All 50 | 50 | 50 | 50 |
| | SROCC > 0.8 | 43 | 43 | 43 |
| | SROCC > 0.9 | 22 | 22 | 22 |
| | SROCC > 0.92 | 14 | 14 | 14 |
| | SROCC > 0.93 | 7 | 7 | 7 |
| LASSO | NNZ > 20 | 43 | 39 | 41 |
| | NNZ > 30 | 38 | 25 | 29 |
| | NNZ > 40 | 24 | 20 | 21 |
| | NNZ > 50 | 16 | 12 | 11 |
| | NNZ > 60 | 11 | 9 | 6 |
| | NNZ > 70 | 7 | 6 | 6 |

## 6. Neural Network Training and Verification Results

The main criteria of the NN training and verification are the SROCC values. Four SROCCs have been analyzed: SROCCtrain(Max), SROCCtrain(Lasso), SROCCtest(Max), and SROCCtest(Lasso) that correspond to the training and test (verification) cases using maximal and Lasso-determined numbers of inputs. Starting from the NN with two hidden layers, where preliminary fitting is not used, the obtained results are presented in Figure 4. Digits near each point in the presented plots show the numbers of neurons in hidden layers. Data in Figure 4a relate to the case when the numbers of neurons in hidden layers are the same, whilst the plots in Figure 4b correspond to situation when the numbers of neurons steadily diminish.
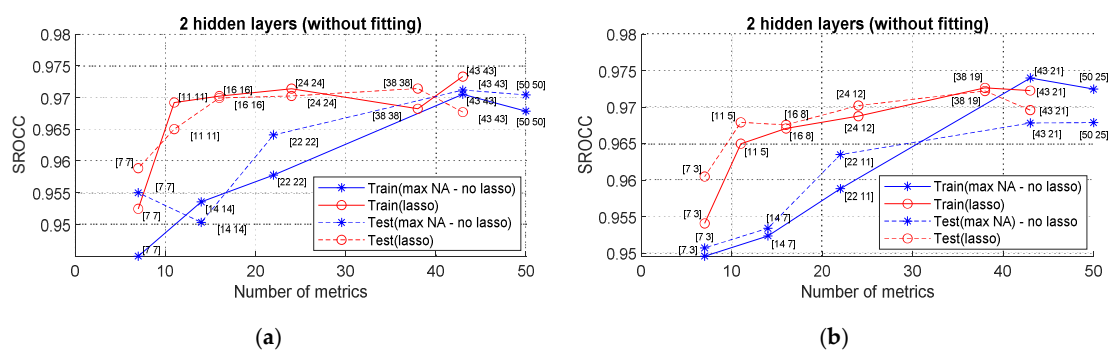


**Figure 4.** SROCC values for NNs with two hidden layers with different numbers of inputs: (**a**) equal number of neurons in hidden layers; (**b**) twice smaller number of neurons in hidden layers.

The analysis of data shows the following:

- If Lasso is not used, the increase in the number of metrics leads to a general tendency of increasing both SROCCtrain and SROCCtest; meanwhile, for more than 40 inputs, the improvement is not observed;
- If Lasso regularization is applied, the results are not so good if the number of inputs ($N_{inp}$) is smaller than 20; but if it exceeds 20, the performance practically does not depend on $N_{inp}$; this means that the Lasso method allows the simplification of the NN structure, minimizing the number of inputs and neurons in the other layers;
- The best (largest) SROCC values exceed 0.97, demonstrating that a sufficient improvement (compared to the best elementary metric) is attained due to the use of the NN and parameters' optimization;
- SROCCtrain and SROCCtest are practically the same for each configuration of the analyzed NN, so training results can be considered as stable;
- No essential difference in results has been found for equal or non-equal numbers of neurons in hidden layers.

Concerning another number of hidden layers, namely four, the obtained plots are given in Figure 5. The analysis of the obtained data makes it possible to draw two main conclusions. Firstly, there are no obvious advantages in comparison to the case of using NNs with two hidden layers. Secondly, for all other conclusions given above, concerning SROCCtrain and SROCCtest, the influence of Ninp and Lasso regularization are the same, i.e., it is reasonable to apply a limited number, e.g., 24 elementary metrics determined by Lasso.

The next question to be answered is: "does preliminary fitting help?" The answer to it is presented in Figure 6 with two sets of plots. Both sets are obtained for the NNs with two hidden layers and with an equal number of neurons in them. The plots in Figure 6a are obtained for the PFF, and in Figure 6b—for the best fit. A comparison of the plots corresponding to each other in Figure 6 shows that there is no sufficient difference in the choice of fitting. Moreover, comparison to the corresponding plots in Figure 4a indicates that preliminary fitting does not produce sufficient performance improvement compared to the cases when it is not used. This means that the trained NNs provide this pre-processing by themselves.

The other conclusions that stem from the analysis of the plots in Figure 6 are practically the same as earlier. The lasso method ensures that the performance close to optimal can be provided for $N_{inp}$ slightly larger than 20. The maximal attained values of SROCC are above 0.97 and smaller than 0.975. The results for four hidden layers and non-equal numbers of neurons in the hidden layers have been analyzed as well, and the best results are practically at the same level.
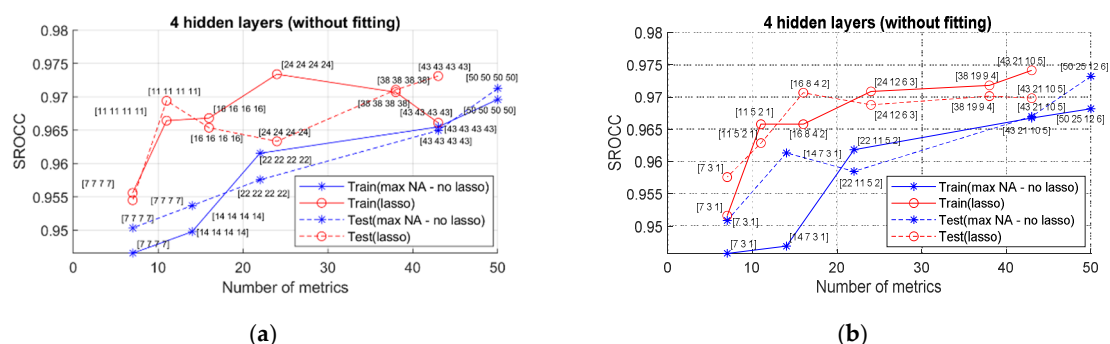


(a)  (b)

**Figure 5.** SROCC values for NNs with four hidden layers with different numbers of inputs: (**a**) equal number of neurons in hidden layers; (**b**) twice smaller number of neurons in hidden layers.
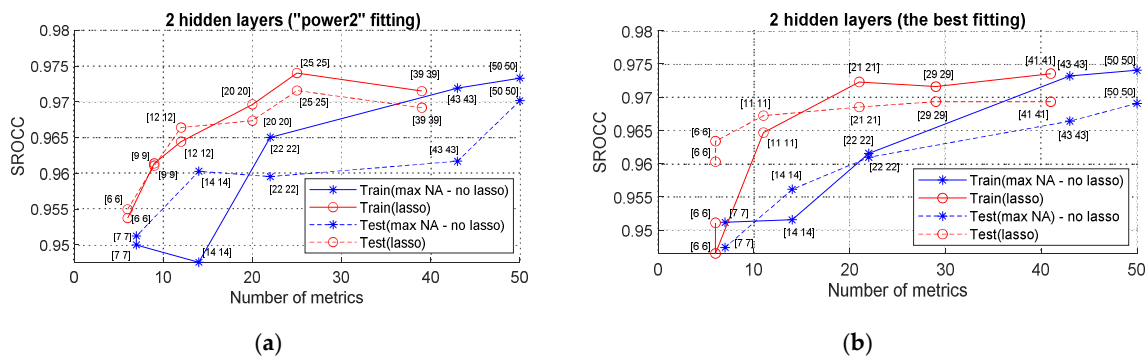
**Figure 6.** SROCC values for NNs with two hidden layers with different numbers of inputs and an equal number of neurons in hidden layers with two ways of fitting: (**a**) PFF; (**b**) the best fitting.

To prove that the number of hidden layers does not have any essential influence on the performance of the combined NN-based metrics, the metric characteristics for three and five hidden layers of NNs that used PFF for elementary metric pre-processing and equal numbers of neurons in all layers are presented in Figure 7. The analysis shows that the maximal attained values of the SROCC are even smaller than for NNs with two layers.
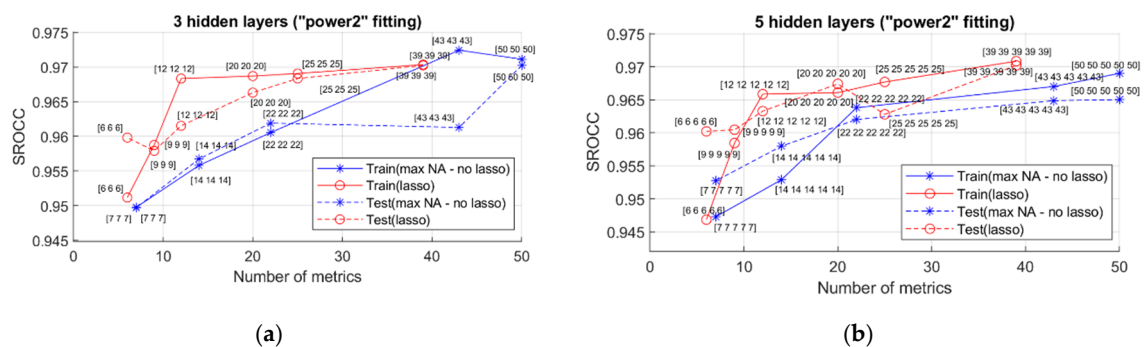


**Figure 7.** SROCC values for NNs with equal numbers of neurons in hidden layers for PFF with: (**a**) three layers; (**b**) five layers.

Completing the analysis based on SROCC, it is possible to state the following:

- there are many configurations of NNs that provide approximately the same SROCC;
- keeping in mind the desired simplicity, the use of NNs with two hidden layers without fitting, and with a number of inputs of about 20, is recommended.

Nevertheless, two other aspects are interesting as well—what are the elementary metrics "recommended" by Lasso in this case, and are the conclusions drawn from SROCC analysis in agreement with conclusions that stem from the analysis for other criteria? Answering the first question, two good NN configurations can be found in Figure 4a, namely NNs with 16 and 24 inputs. The elementary metrics used by the NN with 16 inputs, as well as with 24 inputs, are listed in Table 5. As can be seen, all 16 metrics from the first set are also present in the second set.

The first observation is that the metrics MDSI, CVSSI, PSNRHA, GMSD, IGM, HaarPSI, ADM, IQM2, which are among the top-20 in Table 2, are present among the chosen ones. Some moderately good metrics, such as DSS, are also chosen. There are elementary metrics that are efficient according to data in Table 2 but they are not chosen, for example, MCSD, PSNRHMAm, PSIM. Although for PSNRHMAm, the reason for excluding this metric by Lasso may be its high correlation with PSNRHVS and PSNRHA, the situation is not as clear for MCSD and PSIM. Meanwhile, the sets contain such metrics as WASH and MSVD that, according to data in Table 2, do not perform well. In addition,

both sets contain PSNR, and the second set also contains MSE, strictly connected to PSNR. This means that, although Lasso allows making the sets of recommended elementary metrics narrower, the result of its operation is not optimal.

**Table 5.** Elementary metrics used by the neural network for 16 and 24 inputs with SROCC values.

| Full-Reference Metric | SROCC (All Distortions) | Lasso (16 Inputs) | Lasso (24 Inputs) |
|---|---|---|---|
| MDSI [69] | 0.8897 | X | X |
| HaarPSI [102] | 0.8730 | X | X |
| MS-UNIQUE [101] | 0.8708 | | X |
| PSNRHA [72] | 0.8198 | X | X |
| CVSSI [104] | 0.8090 | X | X |
| IGM [94] | 0.8023 | X | X |
| GMSD [95] | 0.8004 | X | X |
| IQM2 [96] | 0.7955 | X | X |
| DSS [97] | 0.7915 | | X |
| ADM [88] | 0.7861 | X | X |
| MS-SSIM [66] | 0.7872 | | X |
| RFSIM [79] | 0.7721 | X | X |
| CSSIM4 [105] | 0.7394 | | X |
| DSI [106] | 0.7114 | X | X |
| VIF [74] | 0.6816 | | X |
| PSNRHVS [84] | 0.6536 | X | X |
| PSNR | 0.6396 | X | X |
| MSE | 0.6396 | | X |
| NQM [81] | 0.6349 | | X |
| QILV [83] | 0.5975 | | X |
| CWSSIM [86] | 0.5551 | X | X |
| IFC [82] | 0.5229 | X | X |
| WASH [93] | 0.2903 | X | X |
| MSVD [77] | 0.1261 | X | X |

Nevertheless, there are several positive outcomes of the design using Lasso. They become obvious from the analysis of data presented in Table 6. It may be observed that there are several good configurations of NNs that provide a SROCC of about 0.97 for the number of inputs of about 20 (this is also shown in plots). Moreover, these NNs ensure RMSE values that are considerably smaller than for the best elementary metric after linearization (see data in Table 3 where the best values are larger than 0.39). In addition, the values of the Pearson correlation coefficient (PCC) are also large and exceed 0.97, indicating very good linearity properties of the designed combined metrics.

Having calculated the SROCC, RMSE and PCC values, it is possible to carry out a more thorough analysis. The first observation is that SROCC, RMSE and PCC are highly correlated in our case. Larger SROCC and PCC correspond to smaller RMSE. The best results, according to all three criteria, are produced by the NN with configuration #3, although, considering the NN complexity, the configuration #2 is good as well. The number of inputs smaller than 16 (e.g., 11 or 12 in configurations #1, #4, #5) leads to worse values of the considered criteria. The use of the NN configurations with preliminary fitting, a decreasing number of neurons in hidden layers, and a larger number of hidden layers (configurations ##4–9), does not produce improvements in comparison to the corresponding configurations #1 and #2. Thus, the application of the NN configuration #2 will be further analyzed.

**Table 6.** Performance characteristics of designed NN-based combined quality metrics.

| | NN Configuration | The Best Network | | | Average of Top 5 Results (with Standard Deviation) | | |
|---|---|---|---|---|---|---|---|
| | | SROCC | RMSE | PCC | SROCC | RMSE | PCC |
| 1 | (NNZ60) 11 inputs, 2 hidden layers, equal number of neurons, no fitting (11/11) | 0.9683 | 0.2873 | 0.9706 | 0.9658 (0.0018) | 0.2919 (0.0071) | 0.9697 (0.0015) |
| 2 | (NNZ50) 16 inputs, 2 hidden layers, equal number of neurons, no fitting (16/16) | 0.9702 | 0.2699 | 0.9741 | 0.9666 (0.0022) | 0.2863 (0.0105) | 0.9708 (0.0021) |
| 3 | (NNZ40) 24 inputs, 2 hidden layers, equal number of neurons, no fitting (24/24) | 0.9712 | 0.2603 | 0.9760 | 0.9694 (0.0011) | 0.2698 (0.0070) | 0.9742 (0.0013) |
| 4 | 12 inputs, 2 hidden layers, equal number of neurons, power2 fitting (12/12) | 0.9652 | 0.2933 | 0.9694 | 0.9649 (0.0003) | 0.2961 (0.0026) | 0.9688 (0.0006) |
| 5 | 11 inputs, 2 hidden layers, equal number of neurons, "best" fitting (11/11) | 0.9656 | 0.2927 | 0.9695 | 0.9648 (0.0009) | 0.2992 (0.0055) | 0.9681 (0.0012) |
| 6 | 16 inputs, 2 hidden layers, twice smaller number of neurons, no fitting (16/8) | 0.9672 | 0.2840 | 0.9713 | 0.9664 (0.0006) | 0.2890 (0.0046) | 0.9703 (0.0010) |
| 7 | 16 inputs, 3 hidden layers, equal number of neurons, no fitting (16/16/16) | 0.9665 | 0.2912 | 0.9698 | 0.9654 (0.0007) | 0.2950 (0.0034) | 0.9691 (0.0007) |
| 8 | 16 inputs, 4 hidden layers, equal number of neurons, no fitting (16/16/16/16) | 0.9664 | 0.2923 | 0.9696 | 0.9654 (0.0010) | 0.2982 (0.0072) | 0.9684 (0.0015) |
| 9 | 16 inputs, 4 hidden layers, twice smaller number of neurons, no fitting (16/8/4/2) | 0.9674 | 0.2775 | 0.9726 | 0.9673 (0.0008) | 0.2841 (0.0054) | 0.9713 (0.0011) |

Table 6 contains three columns marked by the heading "The best network" and three columns marked by "The top 5 results". It has been mentioned earlier that the results of the NN learning depend on the random division of distorted images into training and testing sets. Because of this, to analyze the stability of training, we have calculated the mean SROCC, RMSE and PCC values for the top five results of NN training for each configuration as well as their standard deviations provided in brackets. A comparison of SROCC, RMSE and PCC for the top five results to the corresponding values for the best network shows that the difference is small. Moreover, the conclusions that can be drawn as the result of the analysis of these "average" results concerning NN performance fully coincide with conclusions drawn from the analysis for the best network.

## 7. Analysis of Computational Efficiency

In addition, the computational efficiency should be briefly discussed. For the NN-based metric with 16 inputs, the calculation of PSNR, MDSI, PSNRHVS, ADM, GMSD, WASH, IQM2, CVSSI and HaarPSI is very fast or fast, the calculation of IFC and RFSIM requires several times longer, whereas the calculation of MSVD, CWSSIM, PSNRHA, IGM, and DSI, takes even more time (about one order of magnitude). Thus, even with parallel calculations of elementary metrics, a calculation of the NN-based metric needs sufficiently more time than for such good elementary metrics as MDSI, GMSD or HaarPSI.

Hence, further research directions should be directed to the combination of possibly fast elementary metrics, providing a possibly good balance between the MOS prediction monotonicity (as well as accuracy) and computational efficiency. The average calculation time of the considered elementary metrics determined for 512 × 384 pixels images from the TID2013 dataset is provided in Table 7. Time data was evaluated using a notebook with Intel i5 4th generation CPU and 8 GB RAM controlled by the Linux Ubuntu 18.04 operating system, using MATLAB® 2019b software.

**Table 7.** Average calculation time for the images from the TID2013 database.

| | Metric | Calculation Time [s] | Speed | Use in the Proposed Metric |
|---|---|---|---|---|
| 1 | MSE | 0.0622 | very fast | yes (only for 24 inputs) |
| 2 | GMSD [95] | 0.0760 | very fast | yes |
| 3 | MCSD [98] | 0.0794 | very fast | no |
| 4 | PSNRHVSM [65] | 0.0819 | very fast | no |
| 5 | QILV [83] | 0.0913 | very fast | yes (only for 24 inputs) |
| 6 | MDSI [69] | 0.0918 | very fast | yes |
| 7 | CVSSI [104] | 0.0920 | very fast | yes |
| 8 | SSIM [41] | 0.1053 | very fast | no |
| 9 | SR-SIM [90] | 0.1065 | very fast | no |
| 10 | GSM [89] | 0.1094 | very fast | no |
| 11 | MS-SSIM [66] | 0.1139 | very fast | yes (only for 24 inputs) |
| 12 | ESSIM [91] | 0.1160 | very fast | no |
| 13 | UQI [42] | 0.1180 | very fast | no |
| 14 | VIFP [74] | 0.1277 | very fast | no |
| 15 | HaarPSI [102] | 0.1280 | very fast | yes |
| 16 | PSIM [70] | 0.1282 | very fast | no |
| 17 | SFF [92] | 0.1400 | very fast | no |
| 18 | CSSIM [105] | 0.1491 | very fast | no |
| 19 | PSNR | 0.1538 | very fast | yes |
| 20 | ADD_SSIM [99] | 0.1813 | fast | no |
| 21 | ADD_GSIM [99] | 0.1837 | fast | no |
| 22 | WASH [93] | 0.1854 | fast | yes |
| 23 | WSNR [80] | 0.1874 | fast | no |
| 24 | PSNRHVS [84] | 0.1932 | fast | yes |
| 25 | DSS [97] | 0.2284 | fast | yes (only for 24 inputs) |
| 26 | IQM2 [96] | 0.2331 | fast | yes |
| 27 | ADM [88] | 0.2568 | fast | yes |
| 28 | VSNR [85] | 0.2602 | fast | no |
| 29 | VSI [71] | 0.2653 | fast | no |
| 30 | FSIMc [78] | 0.2888 | fast | no |
| 31 | UNIQUE [100] | 0.2938 | fast | no |
| 32 | FSIM [78] | 0.3000 | fast | no |
| 33 | NQM [81] | 0.3018 | fast | yes (only for 24 inputs) |
| 34 | PSNRHMAm [73] | 0.4291 | medium | no |
| 35 | IWSSIM [87] | 0.4500 | medium | no |
| 36 | PSNRHMAy [72] | 0.5259 | medium | no |
| 37 | RFSIM [79] | 0.6073 | medium | yes |
| 38 | PSNRHAy [72] | 0.6239 | medium | no |
| 39 | MS-UNIQUE [99] | 0.6917 | medium | yes (only for 24 inputs) |
| 40 | VIF [74] | 0.8343 | medium | yes (only for 24 inputs) |
| 41 | IFC [82] | 0.8574 | medium | yes |
| 42 | SSIM4 [105] | 1.0142 | slow | no |
| 43 | RVSIM [103] | 1.0736 | slow | no |
| 44 | PSNRHMA [72] | 1.3862 | slow | no |
| 45 | PSNRHA [72] | 1.4027 | slow | yes |
| 46 | CSSIM4 [105] | 1.4421 | slow | yes (only for 24 inputs) |
| 47 | CWSSIM [86] | 1.5497 | slow | yes |
| 48 | MSVD [77] | 2.7140 | slow | yes |
| 49 | DSI [106] | 2.8284 | slow | yes |
| 50 | IGM [94] | 9.9101 | very slow | yes |

## 8. Verification for Three-Channel Remote Sensing Images

The analysis of metrics' performance and their verification for multichannel RS images is a complex problem. Obviously, the best solution could be to have a database of reference and distorted RS images and MOS for each distorted image. This way of a future research is, in general, possible and

expedient but it requires considerable time and effort. Firstly, a large number of observers have to be attracted to experiments. Secondly, these observers should have some skills in the analysis of RS data; this is the main problem. Thirdly, the set of images to be viewed and assessed has to be somehow agreed in the RS community.

Because of this, we are currently able to perform only some preliminary tests. The aim of the very first test is to show that the designed metric (in fact, MOS predicted by the trained NN) produces reasonable results for particular images. Regarding the images of Figure 1, for each of them the true MOS value is available. Furthermore, processing the value of each metric (PSNR, PSNRHA, MDSI) by using the linearization with the parameters previously processed (power fitting function) is also possible to obtain the corresponding estimation of MOS. Comparing these values with the MOS value predicted by NN (Table 8), it is clear that NN provides good results. Generalizing from this specific case to the TID2013 Noise&Actual subset, an interesting scatter plot can be obtained (Figure 8). A remarkable aspect is that the MOS values predicted from the selected elementary metrics can even be negative, whereas this drawback is absent when the MOS values are predicted by using NN. It is also seen that the NN-based metric provides high linearity of the relation between true and predicted MOS. Problems of MDSI for small MOS values are observed as well.

**Table 8.** True and Predicted MOS Values.

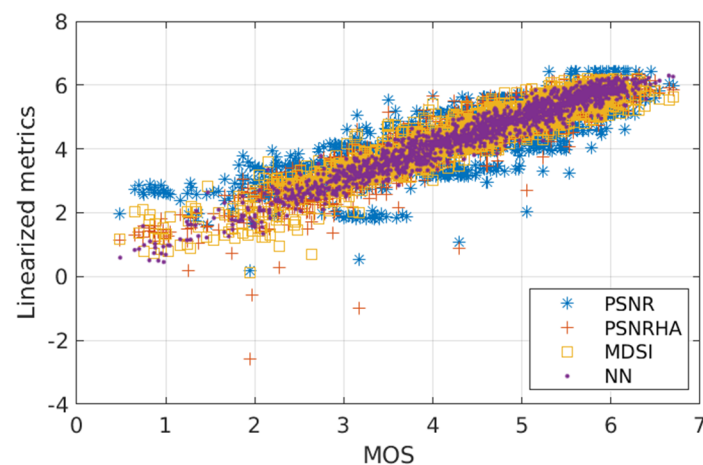| Image | True MOS | Estimated MOS Using the Single Metric | | | MOS Predicted by the Designed Neural Network |
| | | PSNR | PSNRHA | MDSI | |
| --- | --- | --- | --- | --- | --- |
| Figure 1a | 6.08 | 6.4176 | 5.8410 | 5.9314 | 5.8329 |
| Figure 1b | 5.97 | 5.5941 | 6.1279 | 5.9685 | 5.8697 |
| Figure 1c | 4.67 | 5.6024 | 4.9626 | 4.7508 | 4.7202 |
| Figure 1d | 3.87 | 4.5033 | 4.5790 | 4.4500 | 4.2479 |



**Figure 8.** Scatter plot illustrating the correlation of the four methods of MOS prediction presented in Table 8 with true MOS values (range 0–7) for the TID2013 Noise&Actual subset.

The further studies relate to four test images presented in Figure 9. These are three-channel pseudo-color images called Frisco, Diego2, Diego3, and Diego4, respectively, all of size $512 \times 512$ pixels, 24 bits per pixel, from the visible range of the Landsat sensor. The reasons for choosing them are twofold, as these images are of different complexity and they have been already used in some experiments [112]. The images Frisco and Diego4 are quite simple since they contain large homogeneous regions, whereas the image Diego2 has a very complex structure (a lot of small-size details and textures), and the image Diego3 is of a medium complexity.

**Figure 9.** Three-channel remote sensing (RS) test images: (**a**) Frisco; (**b**) Diego2; (**c**) Diego3; (**d**) Diego4.

A standard requirement of visual quality metrics is monotonicity, i.e., monotonous increasing or decreasing if the "intensity" of a given type of distortion increases. This property can be easily checked for many different types of distortions. These images have been compressed using the lossy AGU method [113], providing different quality and compression ratios (CR). This has been performed by changing the quantization step (QS) where a larger QS relates to larger introduced distortions and worse visual quality, respectively. Nevertheless, considering the possible extension of the proposed approach for multispectral RS images, as well as the highly demanded development of the RS image quality assessment database in the future, some more typical approaches to RS data compression should eventually be applied for this purpose, such as the CCSDS 123.0-B-2 standard [114].

The quality (excellent, good and so on) is determined according to the results in [61]. The collected data are presented in Table 9. An obvious tendency is that all metrics including the designed one become worse if QS (and CR, respectively) increases. The values of the NN-based metric are larger than MOS values predicted from elementary metrics for the test image Frisco, but smaller for the test image Diego2. Probably, this property is partly connected with image complexity. However, there are

evidences that the designed metric "behaves" correctly. In Figure 10, the compressed images using QS = 40 are presented, for which distortions are always visible. A visual inspection (comparison) of these images to the reference images in Figure 9a,b shows that distortions are more visible for the test image Diego2. This is clearly confirmed by the values of PSNRHA (36.32 dB and 32.63 dB, respectively—see Table 9). PSNR shows the same tendency, although MDSI does not indicate this. Hence, some cases may be encountered when conclusions drawn from the analysis of different quality metrics can be different.

**Table 9.** Predicted MOS Values for Two RS Test Images Using Linearization.

| RS Image | QS | Quality | PSNR | | PSNRHA | | MDSI | | Designed Metric |
|---|---|---|---|---|---|---|---|---|---|
| | | | Value | Predicted MOS | Value | Predicted MOS | Value | Predicted MOS | |
| Frisco | 10 | excellent | 43.0815 | 6.3461 | 44.8013 | 6.3442 | 0.1442 | 6.0263 | 6.5005 |
| | 20 | excellent | 38.7741 | 5.7441 | 40.2754 | 5.8940 | 0.1920 | 5.5972 | 6.3730 |
| | 30 | good | 36.5231 | 5.4008 | 37.9569 | 5.6184 | 0.2183 | 5.3271 | 6.2117 |
| | 40 | good | 34.9522 | 5.1477 | 36.3188 | 5.4009 | 0.2371 | 5.1194 | 6.0510 |
| Diego2 | 10 | excellent | 42.3026 | 6.2421 | 43.8623 | 6.2590 | 0.1250 | 6.1743 | 5.6921 |
| | 20 | good | 36.4450 | 5.3885 | 38.0930 | 5.6356 | 0.1734 | 5.7735 | 5.4895 |
| | 30 | good | 32.8438 | 4.7884 | 34.8640 | 5.1893 | 0.2082 | 5.4338 | 5.2050 |
| | 40 | middle | 30.3845 | 4.3373 | 32.6321 | 4.8251 | 0.2347 | 5.1473 | 4.8780 |



(**a**)                                                                (**b**)
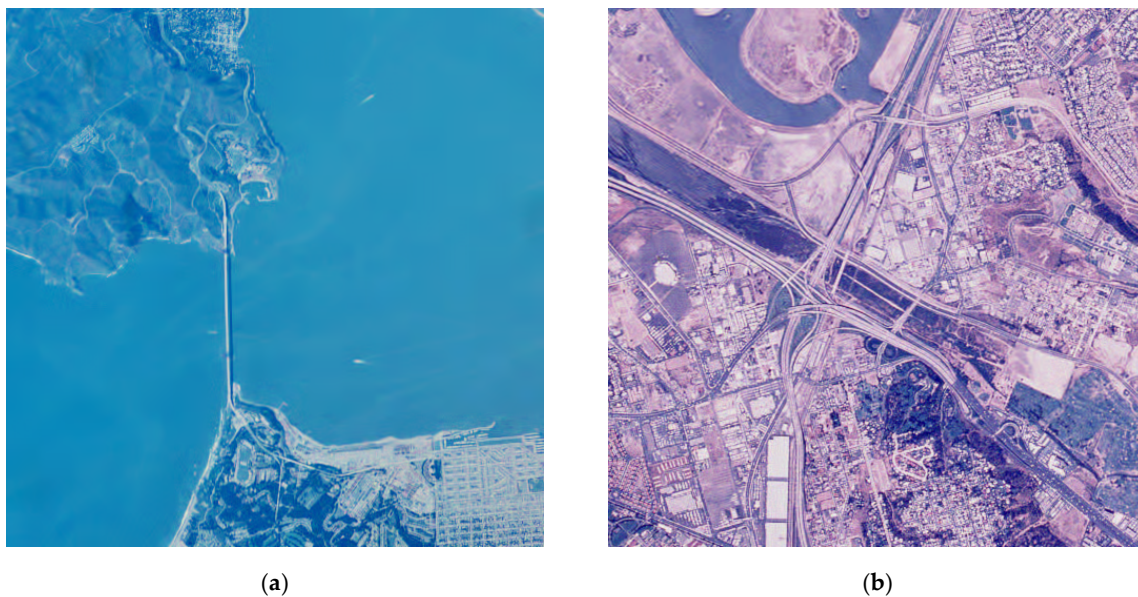
**Figure 10.** Compressed images with quantization step (QS) = 40: (**a**) Frisco; (**b**) Diego2.

The analysis for the case of additive white Gaussian noise that has been added to the considered four three-channel images has also been carried out. Four values of noise variance have been used to correspond to four upper levels of distortions exploited in TID2013. The mean MOS values (averaged for 25 test images in TID2013) have also been obtained that correspond to these values of noise variance. For each image, PSNR, PSNRHA, and MDSI have been calculated and the corresponding predicted values of MOS have been determined. The NN-based metric has been calculated as well. The obtained results are presented in Table 10.

**Table 10.** Predicted MOS values for four RS test images distorted by an additive white Gaussian noise.

| RS Image | Noise Variance | Quality | PSNR | | PSNRHA | | MDSI | | Designed Metric | Mean MOS |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Value | Predicted MOS | Value | Predicted MOS | Value | Predicted MOS | | |
| Frisco | 32.5 | good | 32.99 | 5.40 | 34.33 | 5.62 | 0.21 | 5.38 | 5.86 | 5.67 |
| | 65 | middle | 30.01 | 4.90 | 31.45 | 5.20 | 0.25 | 4.96 | 5.28 | 5.23 |
| | 130 | middle | 27.03 | 4.36 | 28.46 | 4.40 | 0.29 | 4.48 | 4.86 | 4.85 |
| | 260 | bad | 24.06 | 3.77 | 25.56 | 4.11 | 0.33 | 3.96 | 4.24 | 3.77 |
| Diego2 | 32.5 | good | 32.95 | 5.39 | 34.42 | 5.61 | 0.18 | 5.75 | 5.45 | 5.67 |
| | 65 | middle | 29.95 | 4.89 | 31.40 | 5.20 | 0.21 | 5.44 | 5.18 | 5.23 |
| | 130 | middle | 26.96 | 4.35 | 28.41 | 4.70 | 0.24 | 5.04 | 4.81 | 4.85 |
| | 260 | bad | 23.98 | 3.75 | 25.63 | 4.13 | 0.29 | 4.54 | 4.32 | 3.77 |
| Diego3 | 32.5 | good | 32.94 | 5.39 | 34.49 | 5.61 | 0.19 | 5.67 | 5.71 | 5.67 |
| | 65 | middle | 29.96 | 4.90 | 31.41 | 5.20 | 0.22 | 5.33 | 5.36 | 5.23 |
| | 130 | middle | 26.97 | 4.35 | 28.42 | 4.70 | 0.25 | 4.92 | 4.93 | 4.85 |
| | 260 | bad | 23.99 | 3.75 | 25.50 | 4.10 | 0.30 | 4.41 | 4.26 | 3.77 |
| Diego4 | 32.5 | good | 32.94 | 5.39 | 34.40 | 5.62 | 0.16 | 5.91 | 6.18 | 5.67 |
| | 65 | middle | 29.97 | 4.90 | 31.40 | 5.25 | 0.19 | 5.62 | 5.84 | 5.23 |
| | 130 | middle | 26.98 | 4.35 | 28.42 | 4.70 | 0.23 | 5.25 | 5.52 | 4.85 |
| | 260 | bad | 23.98 | 3.75 | 25.77 | 4.16 | 0.26 | 4.81 | 5.01 | 3.77 |

The analysis shows that all metrics become worse (PSNR, PSNRHA and the designed metric decrease, and MDSI increases) if noise variance increases, i.e., the monotonicity property is preserved. The predicted MOS values are quite close to the mean MOS, whereas for good and middle quality PSNRHA, MDSI and the NN-based metric provide better MOS prediction than PSNR. However, for bad quality images, the situation is the opposite.

## 9. Conclusions

The task of assessing the visual quality of remote sensing images subject to different types of degradations has been considered. It has been shown that there are no commonly accepted metrics and, therefore, their design is desired. The problems that can be encountered have been mentioned and their solution has been proposed. The already existing database of distorted color images, for which it is possible to choose images with the types of distortions often observed in remote sensing, has been employed for this purpose. Its use allows the determination of existing visual quality metrics that perform well for the types of degradations that are of interest. The best of such metrics provides a SROCC with a MOS of about 0.93, which is considered as a very good result. Moreover, the database TID2013 allows the design of visual quality metrics based on the use of elementary visual quality metrics as inputs. Several configurations of NNs and methods of input data pre-processing have been studied. It has been shown that even simple NNs without input pre-processing (linearization) having two hidden layers are able to provide SROCC values of about 0.97. PCC values are of the same order, meaning that the relation between NN output and MOS values is practically linear.

Some elementary metrics and the designed one have then been verified for three-channel remote sensing images with two types of degradations, demonstrating the monotonicity of the proposed metric's behavior. Besides, it has been shown that the designed metric produces an accurate evaluation of MOS that allows the classification of remote sensing images according to their quality. Since, as shown in the paper, from the visual quality point of view, color images and RS images (after applying the necessary operation to visualize them) have similar characteristics, this manuscript provides the first evidence that a full-reference image quality metric, developed using a dataset (TID2013) that is actually not a remote sensing dataset, works well when applied to RGB images obtained by remote sensing data.

In the future, the ways of accelerating NN-based metrics by means of restricting the set of possible inputs are planned to be analyzed, considering the computational efficiency of the input metrics as well. We also hope that some perceptual experiments, with the help of some specialists in RS image analysis

for image quality assessment, will be carried out. Although for different types of RS data, different types of degradations can be important, we have considered only those ones that are quite general and are present in TID2013. For example, the characteristics of clouds can be additional feature(s) used as elementary metrics in the quality characterization of particular types of RS data in future research.

One of the limitations of the proposed method is the necessity of calculation of several metrics, which are not always fast. To overcome this issue, some of them may be calculated in parallel, although hardware acceleration possibilities should be provided in such cases.

Another shortcoming is the fact that many metrics are developed for grayscale images and hence they may be applied for single channel only (or independently for three channels leading to three independent results). Hence, a relevant direction of our further research will be related to the optimization of color to greyscale conversion methods used for the individual elementary metrics. In some cases, an appropriate application of elementary metrics for multichannel images may also require changes of data types and dynamic ranges. Although the color to greyscale conversion based on the International Telecommunication Union (ITU) Recommendation BT.601-5 (with the use of RGB to YCbCr conversion in fact limiting the range of the Y component to the range [16,235]), has been assumed in this paper, the results of individual elementary metrics obtained using various color spaces and conversion methods may lead to further increases in the combined metrics' performance.

## References

1. Schowengerdt, R.A. *Remote Sensing, Models, and Methods for Image Processing*, 3rd ed.; Academic Press: Burlington, MA, USA, 2007; ISBN 978-0-12-369407-2.

2. 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 International Publishing: Cham, Switzerland, 2016; pp. 277–285; ISBN 978-3-319-22559-3.

3. 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]

4. Marcuccio, S.; Ullo, S.; Carminati, M.; Kanoun, O. Smaller Satellites, Larger Constellations: Trends and Design Issues for Earth Observation Systems. *IEEE Aerosp. Electron. Syst. Mag.* **2019**, *34*, 50–59. [CrossRef]

5. 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]

6. European Space Agency First Applications from Sentinel-2A. Available online: http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2/First_applications_from_Sentinel-2A (accessed on 8 May 2020).

7. Joshi, N.; Baumann, M.; Ehammer, A.; Fensholt, R.; Grogan, K.; Hostert, P.; Jepsen, M.; Kuemmerle, T.; Meyfroidt, P.; Mitchard, E.; et al. A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote Sens.* **2016**, *8*, 70. [CrossRef]

8.　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]

9.　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]

10.　Ponomaryov, V.; Rosales, A.; Gallegos, F.; Loboda, I. Adaptive Vector Directional Filters to Process Multichannel Images. *IEICE Trans. Commun.* **2007**, *E90-B*, 429–430. [CrossRef]

11.　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, 4–8 May 2009. Paper IAA B7-0602.

12.　Hagolle, O.; Huc, M.; Pascual, D.V.; Dedieu, G. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images. *Remote Sens. Environ.* **2010**, *114*, 1747–1755. [CrossRef]

13.　Li, X.; Wang, L.; Cheng, Q.; Wu, P.; Gan, W.; Fang, L. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. *ISPRS J. Photogramm. Remote Sens.* **2019**, *148*, 103–113. [CrossRef]

14.　Li, Y.; Clarke, K.C. Image deblurring for satellite imagery using small-support-regularized deconvolution. *ISPRS J. Photogramm. Remote Sens.* **2013**, *85*, 148–155. [CrossRef]

15.　Xia, H.; Liu, C. Remote Sensing Image Deblurring Algorithm Based on WGAN. In *Service-Oriented Computing—ICSOC 2018 Workshops*; Liu, X., Mrissa, M., Zhang, L., Benslimane, D., Ghose, A., Wang, Z., Bucchiarone, A., Zhang, W., Zou, Y., Yu, Q., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2019; Volume 11434, pp. 113–125; ISBN 978-3-030-17641-9.

16.　Christophe, E. Hyperspectral Data Compression Tradeoff. In *Optical Remote Sensing*; Prasad, S., Bruce, L.M., Chanussot, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 9–29; ISBN 978-3-642-14211-6.

17.　Blanes, I.; Magli, E.; Serra-Sagrista, J. A Tutorial on Image Compression for Optical Space Imaging Systems. *IEEE Geosci. Remote Sens. Mag.* **2014**, *2*, 8–26. [CrossRef]

18.　Zemliachenko, A.N.; Kozhemiakin, R.A.; Uss, M.L.; Abramov, S.K.; Ponomarenko, N.N.; Lukin, V.V.; Vozel, B.; Chehdi, K. Lossy compression of hyperspectral images based on noise parameters estimation and variance stabilizing transform. *J. Appl. Remote Sens.* **2014**, *8*, 083571. [CrossRef]

19.　Penna, B.; Tillo, T.; Magli, E.; Olmo, G. Transform Coding Techniques for Lossy Hyperspectral Data Compression. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1408–1421. [CrossRef]

20.　Xu, T.; Fang, Y. Remote sensing image interpolation via the nonsubsampled contourlet transform. In Proceedings of the 2010 International Conference on Image Analysis and Signal Processing, Zhejiang, China, 9–11 April 2010; pp. 695–698.

21.　Al-Shaykh, O.K.; Mersereau, R.M. Lossy compression of noisy images. *IEEE Trans. Image Process.* **1998**, *7*, 1641–1652. [CrossRef] [PubMed]

22.　Shi, X.; Wang, L.; Shao, X.; Wang, H.; Tao, Z. Accurate estimation of motion blur parameters in noisy remote sensing image. In *Proceedings of the SPIE Sensing Technology + Applications*; Huang, B., Chang, C.-I., Lee, C., Li, Y., Du, Q., Eds.; SPIE: Baltimore, MD, USA, 21 May 2015; p. 95010X.

23.　Abramov, S.; Uss, M.; Lukin, V.; Vozel, B.; Chehdi, K.; Egiazarian, K. Enhancement of Component Images of Multispectral Data by Denoising with Reference. *Remote Sens.* **2019**, *11*, 611. [CrossRef]

24.　Dellepiane, S.G.; Angiati, E. Quality Assessment of Despeckled SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 691–707. [CrossRef]

25.　Rubel, O.; Lukin, V.; Rubel, A.; Egiazarian, K. NN-Based Prediction of Sentinel-1 SAR Image Filtering Efficiency. *Geosciences* **2019**, *9*, 290. [CrossRef]

26.　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.

27.　Christophe, E.; Leger, D.; Mailhes, C. Quality criteria benchmark for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2103–2114. [CrossRef]

28.　García-Sobrino, J.; Blanes, I.; Laparra, V.; Camps-Valls, G.; Serra-Sagristà, J. Impact of Near-Lossless Compression of IASI L1C data on Statistical Retrieval of Atmospheric Profiles. In Proceedings of the 4th On-Board Payload Data Compression Workshop (OBPDC), Venice, Italy, 23–24 October 2014.

29. Lukin, V.; Abramov, S.; Krivenko, S.; Kurekin, A.; Pogrebnyak, O. Analysis of classification accuracy for pre-filtered multichannel remote sensing data. *Expert Syst. Appl.* **2013**, *40*, 6400–6411. [CrossRef]

30. Hanley, J.A.; McNeil, B.J. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **1983**, *148*, 839–843. [CrossRef]

31. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; Mapping science series; Lewis Publications: Boca Raton, FL, USA, 1999; ISBN 978-0-87371-986-5.

32. Zhou, H.; Gao, H.; Tian, X. A Super-resolution Reconstruction Method of Remotely Sensed Image Based on Sparse Representation. *Sens. Transducers J.* **2013**, *160*, 229–236.

33. Aswathy, C.; Sowmya, V.; Soman, K.P. Hyperspectral Image Denoising Using Low Pass Sparse Banded Filter Matrix for Improved Sparsity Based Classification. *Procedia Comput. Sci.* **2015**, *58*, 26–33. [CrossRef]

34. Kumar, T.G.; Murugan, D.; Rajalakshmi, K.; Manish, T.I. Image enhancement and performance evaluation using various filters for IRS-P6 Satellite Liss IV remotely sensed data. *Geofizika* **2015**, *32*, 179–189. [CrossRef]

35. Yang, K.; Jiang, H. Optimized-SSIM Based Quantization in Optical Remote Sensing Image Compression. In Proceedings of the 2011 Sixth International Conference on Image and Graphics, Hefei, China, 12–15 August 2011; pp. 117–122.

36. Khosravi, M.R.; Sharif-Yazd, M.; Moghimi, M.K.; Keshavarz, A.; Rostami, H.; Mansouri, S. MRF-Based Multispectral Image Fusion Using an Adaptive Approach Based on Edge-Guided Interpolation. *J. Geogr. Inf. Syst.* **2017**, *9*, 114–125. [CrossRef]

37. Agudelo-Medina, O.A.; Benitez-Restrepo, H.D.; Vivone, G.; Bovik, A. Perceptual Quality Assessment of Pan-Sharpened Images. *Remote Sens.* **2019**, *11*, 877. [CrossRef]

38. Moreno-Villamarin, D.E.; Benitez-Restrepo, H.D.; Bovik, A.C. Predicting the Quality of Fused Long Wave Infrared and Visible Light Images. *IEEE Trans. Image Process.* **2017**, *26*, 3479–3491. [CrossRef] [PubMed]

39. Jagalingam, P.; Hegde, A.V. A Review of Quality Metrics for Fused Image. *Aquat. Procedia* **2015**, *4*, 133–142. [CrossRef]

40. 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]

41. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

42. Wang, Z.; Bovik, A. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]

43. Guo, J.; Yang, F.; Tan, H.; Wang, J.; Liu, Z. Image matching using structural similarity and geometric constraint approaches on remote sensing images. *J. Appl. Remote Sens.* **2016**, *10*, 045007. [CrossRef]

44. Liu, D.; Li, Y.; Chen, S. No-reference remote sensing image quality assessment based on the region of interest and structural similarity. In Proceedings of the 2nd International Conference on Advances in Image Processing—ICAIP '18, Chengdu, China, 16–18 June 2018; ACM Press: New York, NY, USA, 2018; pp. 64–67.

45. Krivenko, S.S.; Abramov, S.K.; Lukin, V.V.; Vozel, B.; Chehdi, K. Lossy DCT-based compression of remote sensing images with providing a desired visual quality. In *Proceedings of the Image and Signal Processing for Remote Sensing XXV*; Bruzzone, L., Bovolo, F., Benediktsson, J.A., Eds.; SPIE: Strasbourg, France, 7 October 2019; p. 36.

46. Lin, W.; Jay Kuo, C.-C. Perceptual visual quality metrics: A survey. *J. Vis. Commun. Image Represent.* **2011**, *22*, 297–312. [CrossRef]

47. Chandler, D.M. Seven Challenges in Image Quality Assessment: Past, Present, and Future Research. *ISRN Signal Process.* **2013**, *2013*, 1–53. [CrossRef]

48. 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]

49. Zhai, G.; Min, X. Perceptual image quality assessment: A survey. *Sci. China Inf. Sci.* **2020**, *63*, 211301. [CrossRef]

50. 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]

51. 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.

52. Sun, W.; Zhou, F.; Liao, Q. MDID: A multiply distorted image database for image quality assessment. *Pattern Recognit.* **2017**, *61*, 153–168. [CrossRef]

53. Ieremeiev, O.; Lukin, V.; Ponomarenko, N.; Egiazarian, K. Robust linearized combined metrics of image visual quality. *Electron. Imaging* **2018**, *2018*, 260-1–260-6. [CrossRef]

54. Okarma, K. Combined Full-Reference Image Quality Metric Linearly Correlated with Subjective Assessment. In *Artificial Intelligence and Soft Computing*; Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6113, pp. 539–546; ISBN 978-3-642-13207-0.

55. Okarma, K. Combined image similarity index. *Opt. Rev.* **2012**, *19*, 349–354. [CrossRef]

56. Okarma, K. Extended Hybrid Image Similarity—Combined Full-Reference Image Quality Metric Linearly Correlated with Subjective Scores. *Electron. Electr. Eng.* **2013**, *19*, 129–132. [CrossRef]

57. Liu, T.-J.; Lin, W.; Kuo, C.-C.J. Image Quality Assessment Using Multi-Method Fusion. *IEEE Trans. Image Process.* **2013**, *22*, 1793–1807. [CrossRef] [PubMed]

58. Lukin, V.V.; Ponomarenko, N.N.; Ieremeiev, 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 Optics and Photonics: San Francisco, CA, USA, 2015; p. 93940K.

59. 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]

60. 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.

61. Ponomarenko, N.; Lukin, V.; Astola, J.; Egiazarian, K. Analysis of HVS-Metrics' Properties Using Color Image Database TID. In *Advanced Concepts for Intelligent Vision Systems*; Battiato, S., Blanc-Talon, J., Gallo, G., Philips, W., Popescu, D., Scheunders, P., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9386, pp. 613–624; ISBN 978-3-319-25902-4.

62. Lukin, V.; Zemliachenko, A.; Krivenko, S.; Vozel, B.; Chehdi, K. Lossy Compression of Remote Sensing Images with Controllable Distortions. In *Satellite Information Classification and Interpretation*; B. Rustamov, R., Ed.; IntechOpen: London, UK, 2019; ISBN 978-1-83880-566-1.

63. Yule, G.U.; Kendall, M.G. *An Introduction to the Theory of Statistics*, 14th ed.; Charles Griffin & Co.: London, UK, 1950.

64. Goossens, B.; Pizurica, A.; Philips, W. Removal of Correlated Noise by Modeling the Signal of Interest in the Wavelet Domain. *IEEE Trans. Image Process.* **2009**, *18*, 1153–1165. [CrossRef] [PubMed]

65. Ponomarenko, N.; Silvestri, F.; Egiazarian, K.; Carli, M.; Astola, J.; Lukin, V. On between-coefficient contrast masking of DCT basis functions. In Proceedings of the Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM 2007, Scottsdale, AZ, USA, 25–26 January 2007; Volume 4.

66. 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.

67. Solomon, J.A.; Watson, A.B.; Ahumada, A. Visibility of DCT basis functions: Effects of contrast masking. In Proceedings of the IEEE Data Compression Conference (DCC'94), Snowbird, UT, USA, 29–31 March 1994; IEEE Computer Society Press: Los Alamitos, CA, USA, 1994; pp. 361–370.

68. Avcibaş, I.; Sankur, B.; Sayood, K. Statistical evaluation of image quality measures. *J. Electron. Imaging* **2002**, *11*, 206–223. [CrossRef]

69. Ziaei Nafchi, H.; 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]

70. Gu, K.; Li, L.; Lu, H.; Min, X.; Lin, W. A Fast Reliable Image Quality Predictor by Fusing Micro- and Macro-Structures. *IEEE Trans. Ind. Electron.* **2017**, *64*, 3903–3912. [CrossRef]

71. Zhang, L.; Shen, Y.; Li, H. VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment. *IEEE Trans. Image Process.* **2014**, *23*, 4270–4281. [CrossRef]

72. Ponomarenko, N.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Carli, M. Modified image visual quality metrics for contrast change and mean shift accounting. In Proceedings of the 2011 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), Polyana-Svalyava, Ukraine, 23–35 February 2011; pp. 305–311.

73. Ieremeiev, O.; Lukin, V.; Ponomarenko, N.; Egiazarian, K. Full-reference metrics multidistortional analysis. *Electron. Imaging* **2017**, *2017*, 27–35. [CrossRef]

74. Sheikh, H.R.; Bovik, A.C. Image information and visual quality. *IEEE Trans. Image Process.* **2006**, *15*, 430–444. [CrossRef]

75. Mansouri, A.; Aznaveh, A.M.; Torkamani-Azar, F.; Jahanshahi, J.A. Image quality assessment using the singular value decomposition theorem. *Opt. Rev.* **2009**, *16*, 49–53. [CrossRef]

76. 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.

77. Shnayderman, A.; Gusev, A.; Eskicioglu, A.M. An SVD-based grayscale image quality measure for local and global assessment. *IEEE Trans. Image Process.* **2006**, *15*, 422–429. [CrossRef] [PubMed]

78. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [CrossRef] [PubMed]

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

80. Mitsa, T.; Varkur, K.L. Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms. In Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, Minneapolis, MN, USA, 27–30 April 1993; Volume 5, pp. 301–304.

81. Damera-Venkata, N.; Kite, T.D.; Geisler, W.S.; Evans, B.L.; Bovik, A.C. Image quality assessment based on a degradation model. *IEEE Trans. Image Process.* **2000**, *9*, 636–650. [CrossRef] [PubMed]

82. Sheikh, H.R.; Bovik, A.C.; de Veciana, G. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* **2005**, *14*, 2117–2128. [CrossRef] [PubMed]

83. Aja-Fernandez, S.; Estepar, R.S.J.; Alberola-Lopez, C.; Westin, C.-F. Image Quality Assessment based on Local Variance. In Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA, 30 August–3 September 2006; pp. 4815–4818.

84. Egiazarian, K.; Astola, J.; Ponomarenko, N.; Lukin, V.; Battisti, F.; Carli, M. Two new full-reference quality metrics based on HVS. In Proceedings of the Second International Workshop on Video Processing and Quality Metrics, Scottsdale, AZ, USA, 22–24 January 2006; Volume 4.

85. Chandler, D.M.; Hemami, S.S. VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images. *IEEE Trans. Image Process.* **2007**, *16*, 2284–2298. [CrossRef]

86. Sampat, M.P.; Wang, Z.; Gupta, S.; Bovik, A.C.; Markey, M.K. Complex Wavelet Structural Similarity: A New Image Similarity Index. *IEEE Trans. Image Process.* **2009**, *18*, 2385–2401. [CrossRef]

87. Wang, Z.; Li, Q. Information Content Weighting for Perceptual Image Quality Assessment. *IEEE Trans. Image Process.* **2011**, *20*, 1185–1198. [CrossRef]

88. Li, S.; Zhang, F.; Ma, L.; Ngan, K.N. Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments. *IEEE Trans. Multimed.* **2011**, *13*, 935–949. [CrossRef]

89. Liu, A.; Lin, W.; Narwaria, M. Image Quality Assessment Based on Gradient Similarity. *IEEE Trans. Image Process.* **2012**, *21*, 1500–1512. [CrossRef]

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

91. Zhang, X.; Feng, X.; Wang, W.; Xue, W. Edge Strength Similarity for Image Quality Assessment. *IEEE Signal Process. Lett.* **2013**, *20*, 319–322. [CrossRef]

92. Chang, H.-W.; Yang, H.; Gan, Y.; Wang, M.-H. Sparse Feature Fidelity for Perceptual Image Quality Assessment. *IEEE Trans. Image Process.* **2013**, *22*, 4007–4018. [CrossRef] [PubMed]

93. Reenu, M.; David, D.; Raj, S.S.A.; Nair, M.S. Wavelet Based Sharp Features (WASH): An Image Quality Assessment Metric Based on HVS. In Proceedings of the 2013 2nd International Conference on Advanced Computing, Networking and Security, Mangalore, India, 15–17 December 2013; pp. 79–83.

94. Wu, J.; Lin, W.; Shi, G.; Liu, A. Perceptual Quality Metric with Internal Generative Mechanism. *IEEE Trans. Image Process.* **2013**, *22*, 43–54. [CrossRef] [PubMed]

95. Xue, W.; Zhang, L.; Mou, X.; Bovik, A.C. Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index. *IEEE Trans. Image Process.* **2014**, *23*, 684–695. [CrossRef] [PubMed]

96. Dumic, E.; Grgic, S.; Grgic, M. IQM2: New image quality measure based on steerable pyramid wavelet transform and structural similarity index. *Signal Image Video Process.* **2014**, *8*, 1159–1168. [CrossRef]

97. Balanov, A.; Schwartz, A.; Moshe, Y.; Peleg, N. Image quality assessment based on DCT subband similarity. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 2105–2109.

98. Wang, T.; Zhang, L.; Jia, H.; Li, B.; Shu, H. Multiscale contrast similarity deviation: An effective and efficient index for perceptual image quality assessment. *Signal Process. Image Commun.* **2016**, *45*, 1–9. [CrossRef]

99. Gu, K.; Wang, S.; Zhai, G.; Lin, W.; Yang, X.; Zhang, W. Analysis of Distortion Distribution for Pooling in Image Quality Prediction. *IEEE Trans. Broadcast.* **2016**, *62*, 446–456. [CrossRef]

100. Temel, D.; Prabhushankar, M.; AlRegib, G. UNIQUE: Unsupervised Image Quality Estimation. *IEEE Signal Process. Lett.* **2016**, *23*, 1414–1418. [CrossRef]

101. Prabhushankar, M.; Temel, D.; AlRegib, G. MS-UNIQUE: Multi-model and Sharpness-weighted Unsupervised Image Quality Estimation. *Electron. Imaging* **2017**, *2017*, 30–35. [CrossRef]

102. 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]

103. Yang, G.; Li, D.; Lu, F.; Liao, Y.; Yang, W. RVSIM: A feature similarity method for full-reference image quality assessment. *EURASIP J. Image Video Process.* **2018**, *2018*, 6. [CrossRef]

104. Jia, H.; Zhang, L.; Wang, T. Contrast and Visual Saliency Similarity-Induced Index for Assessing Image Quality. *IEEE Access* **2018**, *6*, 65885–65893. [CrossRef]

105. Ponomarenko, M.; Egiazarian, K.; Lukin, V.; Abramova, V. Structural Similarity Index with Predictability of Image Blocks. In Proceedings of the 2018 IEEE 17th International Conference on Mathematical Methods in Electromagnetic Theory (MMET), Kiev, Ukraine, 2–5 July 2018; pp. 115–118.

106. Egiazarian, K.; Ponomarenko, M.; Lukin, V.; Ieremeiev, O. Statistical Evaluation of Visual Quality Metrics for Image Denoising. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 6752–6756.

107. Bosse, S.; Maniry, D.; Muller, K.-R.; Wiegand, T.; Samek, W. Neural network-based full-reference image quality assessment. In Proceedings of the 2016 Picture Coding Symposium (PCS), Nuremberg, Germany, 4–7 December 2016; pp. 1–5.

108. Prashnani, E.; Cai, H.; Mostofi, Y.; Sen, P. PieAPP: Perceptual Image-Error Assessment Through Pairwise Preference. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1808–1817.

109. Seo, S.; Ki, S.; Kim, M. JND-SalCAR: A Novel JND-based Saliency-Channel Attention Residual Network for Image Quality Prediction. *arXiv* **2019**, arXiv:1902.05316 Cs.

110. Athar, S.; Wang, Z. A Comprehensive Performance Evaluation of Image Quality Assessment Algorithms. *IEEE Access* **2019**, *7*, 140030–140070. [CrossRef]

111. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288. [CrossRef]

112. Kozhemiakin, R.A.; Rubel, O.; Abramov, S.K.; Lukin, V.V.; Vozel, B.; Chehdi, K. Efficiency Analysis for 3D Filtering of Multichannel Images. In *Image and Signal Processing for Remote Sensing XXII*; Bruzzone, L., Bovolo, F., Eds.; SPIE: Edinburgh, UK, 18 October 2016; p. 1000406.

113. Ponomarenko, N.; Lukin, V.; Egiazarian, K.; Astola, J. DCT Based High Quality Image Compression. In *Image Analysis*; Kalviainen, H., Parkkinen, J., Kaarna, A., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2005; Volume 3540, pp. 1177–1185. ISBN 978-3-540-26320-3.

114. Consultative Committee for Space Data Systems (CCSDS). *Low-Complexity Lossless and Near-Lossless Multispectral and Hyperspectral Image Compression*; CCSDS 123.0-B-2 Recommended Standard, Blue Book, no. 1; CCDSS NASA: Washington, DC, USA, 2019.