remote sensing

MDPI

Review

Image Quality Assessment Methods for Multispectral Pan-Sharpening Images: A Comprehensive Review

Igor Stępień and Mariusz Oszust *

Department of Computer and Control Engineering, Rzeszow University of Technology, Wincentego Pola 2, 35-029 Rzeszow, Poland; istepien@kia.prz.edu.pl
* Correspondence: marosz@kia.prz.edu.pl

Highlights

What are the main findings?

- Current evaluation metrics often rely on unrealistic scale invariance assumptions and arbitrarily chosen parameters, making them unreliable in operational scenarios where the panchromatic band does not fully cover the spectral range of the multispectral image.
- While deep learning models achieve high accuracy on their training data, they exhibit severe overfitting when applied to different sensors, meaning their reliability drops sharply for images from new or unseen satellite sensors.

What are the implications of the main findings?

- Quality evaluation must shift away from abstract mathematical scores toward task-driven assessments that verify whether fused images actually improve the performance of automated remote sensing applications.
- Future research should prioritize domain adaptation and spatial quality mapping to develop models that remain robust across different sensors and diverse, label-scarce environments.

Abstract

Multispectral pan-sharpening aims to fuse high-resolution panchromatic and low-resolution multispectral imagery. However, this process introduces spatial artifacts and spectral distortions. Assessing the quality of fused images remains a fundamental challenge due to the absence of full-resolution ground-truth data. This paper provides a comprehensive review of Image Quality Assessment (IQA) frameworks tailored for pan-sharpened imagery. After overviewing major fusion approaches, including Component Substitution (CS), Multi-Resolution Analysis (MRA), Variational Optimization (VO), and Deep Learning (DL), the review analyzes the evaluation techniques used to benchmark them. It then systematically examines the evolution of evaluation protocols, from classical reference-based metrics relying on Wald's protocol to full-resolution consistency models and recent no-reference (NR) algorithms. The analysis highlights critical methodological bottlenecks within the field, including unrealistic scale-invariance assumptions in consistency-based metrics, dependence on arbitrary parameters, and severe cross-sensor overfitting in deep learning approaches. Furthermore, the review addresses the mismatch between mathematical fidelity, human visual perception, and practical applicability. Finally, it outlines future research directions, focusing on spatial quality mapping and task-driven assessment protocols that validate fusion efficacy based on its impact on automated remote sensing applications.

Check for updates

Academic Editor: Zhenwei Shi

Received: 6 July 2026
Revised: 1 September 2026
Accepted: 2 September 2026
Published: 4 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: remote sensing; image fusion; pan-sharpening; no-reference metrics; benchmarking protocols; perceptual gap

1. Introduction

The last decade has seen a growing demand for high-quality remote sensing data, driven by the expansion of Earth observation applications ranging from coastal monitoring and dune mapping to urban classification [1,2]. Optical remote sensing systems typically acquire high-resolution panchromatic (PAN) images, which capture structural details but lack spectral depth, alongside low-resolution multispectral (MS) images, which provide rich chromatic information at the expense of spatial clarity [3,4]. To overcome the physical and technological constraints of optical sensors, pan-sharpening (PS) has emerged as a fundamental computational procedure that fuses these two modalities to generate a single high-resolution multispectral image [5]. PS images serve as an essential preprocessing step for numerous downstream remote sensing applications, including change detection [6], object recognition [7], visual image analysis, scene interpretation [8], coastal monitoring, dune mapping, urban classification, wildlife tracking, soil cover mapping, or wetland analysis [2]. However, the fusion process inevitably introduces spatial artifacts and spectral distortions related to radiometric inconsistencies, sensor characteristics, and temporal differences between source image acquisitions [9,10]. These degradations reduce the fidelity of the fused image, making Image Quality Assessment (IQA) a continuous challenge [11,12]. Since subjective evaluation aligns with human perception, it is time-consuming and costly, as it requires extensive manual assessment by human evaluators [13]. Consequently, objective IQA methods have been developed to approximate subjective assessments and automate the quality evaluation process [14].

Evaluating PS image quality is limited by the absence of a full-resolution MS ground truth. Historically, evaluation relies on Wald's protocol at a reduced scale, permitting the use of classical full-reference (FR) metrics. However, this protocol assumes strict scale invariance. Algorithms optimized at a reduced scale frequently introduce spatial degradations or spectral shifts when applied at full resolution. To address this, full-resolution assessment utilizes consistency-based analyzers that evaluate fidelity without reference data [15]. Nevertheless, these metrics introduce distinct algorithmic limitations, relying on arbitrary weighting parameters that can alter method rankings. Also, their underlying assumption of invariant inter-band relationships fails because the broad PAN band rarely matches the exact spectral range of the MS bands. Furthermore, while deep learning (DL) no-reference (NR) models achieve high prediction accuracy on specific datasets, they exhibit severe cross-sensor overfitting and lack interpretability, reducing their reliability in heterogeneous environments.

Several review papers address pan-sharpening techniques and their quantitative evaluation [2,12]. However, they focus on algorithm architecture, leaving their limitations unaddressed. The literature lacks a critical examination dedicated specifically to IQA frameworks for pan-sharpened imagery, particularly regarding deep learning and perceptual quality [16,17]. This paper addresses this gap by providing a systematic review focused on IQA, differentiating itself from existing literature by analyzing the transition from classical reference metrics to consistency analyzers and deep learning no-reference models. Through the examination of methodological bottlenecks, such as scale-invariance assumptions in consistency metrics or cross-sensor overfitting in deep learning, this review outlines future directions that prioritize perceptual alignment and task-driven assessment protocols.

The remainder of this paper is organized as follows. Section 2 presents pan-sharpening fusion techniques. Section 3 describes common distortions and assessment protocols. Section 4 introduces the fundamentals of image quality assessment. Section 5 reviews remote sensing sensors and benchmark datasets. Section 6 surveys existing IQA methods for pan-sharpened imagery, including full-reference, full-resolution, classical no-reference, and deep learning-based approaches. Section 7 provides a critical review of these quality metrics. Section 8 discusses future research directions, and Section 9 concludes the paper.

2. Pan-Sharpening Fusion Techniques

Pan-sharpening methods are classified into three primary categories, as outlined by Song et al. [18]: Component Substitution (CS) [19,20], Multi-Resolution Analysis (MRA) [21], and Variational Optimization (VO) [22]. In recent years, Deep Learning (DL) has emerged as a distinct fourth paradigm, significantly advancing the field [23]. It should be noted that a universally standardized taxonomy remains elusive, as researchers often employ diverse classification criteria. Consequently, specific or hybrid methods are frequently assigned to different classes depending on the focus of a given study [24,25].

2.1. CS-Based Pan-Sharpening Methods

Component substitution methods transform the upsampled source MS image into a specific mathematical space to separate spatial and spectral information. The component representing spatial details, known as intensity, is subsequently radiometrically matched and substituted with the high-resolution PAN image [26,27]. Representative classical approaches include Principal Component Analysis (PCA) [28], Intensity-Hue-Saturation (IHS) [29], and Gram–Schmidt (GS) spectral sharpening [30]. Further advancements introduced adaptive statistical coefficients, exemplified by the Band-Dependent Spatial Detail (BDSD) method [31].

While CS-based methods are computationally efficient and robust in preserving high-frequency spatial details, they exhibit high susceptibility to spectral distortion. By substituting a derived component with the PAN image without accounting for exact spectral response overlaps between sensors, these methods introduce radiometric inconsistencies and global color shifts. Such spectral anomalies degrade the reliability of automated downstream spectral analysis, establishing a primary challenge for objective quality evaluation.

2.2. MRA-Based Pan-Sharpening Methods

Multi-resolution analysis techniques enhance MS resolution by injecting high-frequency details extracted from the PAN image via multiscale decomposition [32]. Common implementations employ discrete wavelet transforms, such as the Additive à Trous Wavelet Transform (ATWT) utilizing a standard unitary injection model [33]. To improve radiometric preservation, subsequent extensions include Model 2 (ATWT-M2), which incorporates sensor-specific spectral response adjustments into the injection gain, and Model 3 (ATWT-M3), which applies adaptive injection gains based on local structural variance [27]. Furthermore, Generalized Laplacian Pyramid (GLP) architectures are frequently utilized, particularly those employing spatial filters designed to match the Modulation Transfer Function (MTF) of the optical sensor, such as MTF-GLP [34].

2.3. VO-Based Pan-Sharpening Methods

Variational optimization conceptualizes pan-sharpening as an ill-posed inverse problem, reconstructing the high-resolution MS image by minimizing an objective energy functional. These models balance data fidelity terms against regularization priors to ensure stability [35,36]. Popular implementations include Bayesian frameworks, Total Variation (TV) regularization [37], and Sparse Representation (SR) strategies utilizing advanced dictio-

nary learning [38]. While VO methods effectively model complex structural dependencies, the numerical optimization process often introduces distinct artifacts. Depending on chosen regularization parameters, VO outputs frequently suffer from unnatural over-smoothing of fine textures or isolated geometric warping. This hinders the automated assessment of spatial quality without pristine reference information.

2.4. DL-Based Pan-Sharpening Methods

Deep learning has emerged as a dominant paradigm, utilizing neural networks to learn non-linear mappings directly from source imagery rather than relying on handcrafted mathematical priors [39]. Early architectures heavily relied on Convolutional Neural Networks (CNNs), such as the Multiscale and Multidepth CNN (MSDCNN) [40] and the Deep Residual Pan-sharpening Neural Network (DRPNN) [41]. Recently, the state-of-the-art has rapidly shifted toward more complex architectures, including Generative Adversarial Networks (GANs) [42], Vision Transformers capable of capturing long-range dependencies like PanFormer [43], or Diffusion Models [44]. While these models achieve remarkable perceptual quality, their data-driven nature introduces unique and severe challenges for IQA. DL models, particularly GANs and diffusion models, are prone to generative hallucinations due to the synthesis of high-frequency textures or artificial structural patterns that appear visually pleasing but possess no physical basis in the original sensor data. Furthermore, these highly parameterized networks frequently overfit to specific training distributions. Consequently, unpredictable spatial and spectral distortions can emerge when a method is applied to unseen satellite sensors. Traditional IQA metrics are insufficient for evaluating these modern architectures, as they frequently assign high mathematical scores to physically inaccurate, hallucinated representations.

2.5. Summary

Each category of pan-sharpening methods presents a distinct degradation profile. While CS methods typically compromise global spectral fidelity to achieve high spatial sharpness, MRA frameworks exhibit the opposite trade-off, often producing spatial blurring or ringing near structural edges. Conversely, DL architectures address these classical mathematical limitations but introduce synthetic artifacts and generalization issues that traditional physics-based models do not generate. The diverse nature of these distortions highlights the inadequacy of relying on a single evaluation approach. Comprehensive assessment requires structured protocols to systematically capture these method-specific degradations. Recognizing this methodological variability establishes the direct necessity for robust objective IQA metrics, which are categorized and analyzed in the subsequent sections of this review.

3. Distortions and Assessment Protocols in Pan-Sharpening

3.1. Common Image Distortions in Pan-Sharpening

IQA quantifies the fidelity, spatial detail, and perceptual characteristics of digital images, while enhancement techniques aim to improve these attributes without requiring prior knowledge of specific degradation sources [45,46]. In remote sensing, pan-sharpening addresses the spatial-spectral trade-off in satellite optical sensors by fusing the high spatial resolution of PAN imagery with the rich chromatic data of low-resolution MS imagery [47]. However, the process introduces artifacts that degrade the final PS product. Quality assessment methods are designed to detect and quantify specific degradations. While sensor-induced errors, such as calibration noise or atmospheric interference, influence overall image quality, they are considered upstream degradation sources [48]. Consequently, the evaluation of pan-sharpening performance focuses on the direct outcomes of the fusion

process, which are broadly classified into two primary categories: spectral and spatial distortions (as illustrated in Figure 1).



(**a**) Reference (**b**) PCA



(**c**) EXP

Figure 1. Exemplary images presenting differences between (**a**) the original reference image, and the fused results exhibiting (**b**) severe spectral distortions produced by PCA, and (**c**) spatial distortions (blur and lack of structural detail) represented by the upsampled MS baseline (EXP) [8,27].

3.1.1. Spectral Distortions

Spectral distortion refers to the alteration of original multispectral reflectances and the violation of inter-band relationships during the fusion process. A source of this distortion is the non-linear relationship between the broad spectral response of the PAN sensor and the narrower responses of the individual MS bands [49]. This mismatch leads to radiometric inconsistencies and visible color deviations in the PS images. While most PS techniques utilize histogram matching or mean-variance scaling as a preprocessing step to mitigate global brightness mismatches, local variations in gray values continue to introduce spectral anomalies [50,51]. Methodologically, CS techniques exhibit high susceptibility to severe spectral distortion because they substitute a derived mathematical component with the PAN image, frequently without accounting for the exact spectral overlap of the sensors [36]. Although specific algorithms, such as Gram–Schmidt, minimize these effects better than traditional IHS methods, color degradation persists. The extent of spectral distortion varies depending on the wavelength, scene structure, and the employed algorithm. Additionally, spectral biases are often land-cover dependent, altering reflectances for specific classes (e.g., water bodies versus vegetation) even when global metrics indicate acceptable overall performance. Consequently, quantitative downstream analysis (e.g., vegetation indexing or classification) on PS data requires caution, as the physical validity of the original MS reflectances is compromised [52].

3.1.2. Spatial Distortions

While MRA methods generally excel at preserving spectral fidelity, they are prone to introducing spatial distortions [53]. These degradations arise from inaccuracies in extracting high-frequency details from the PAN image and improperly injecting them into the upsampled MS bands [54]. The structural inconsistencies between the differing resolutions of the source images challenge the fusion model's ability to maintain coherent geometry. Spatial distortions in PS images manifest in several observable forms that severely affect structural interpretability. For instance, blurring occurs when the fusion method fails to adequately transfer high-frequency details from the PAN image, resulting in a loss of sharpness and reduced spatial separability of ground objects. Closely related is over-smoothing, which arises from excessive filtering of high frequencies, leading to an unnatural, plastic appearance and the loss of natural texture. Aliasing appears as jagged edges or repetitive moiré patterns, typically caused by insufficient sampling or improper filtering of high-frequency spatial information during the downsampling and upsampling phases. Furthermore, excessive injection of high-frequency components often produces over-sharpening and ringing artifacts (the Gibbs phenomenon), which introduce unnatural textures and halo structures around high-contrast edges. Additionally, edge artifacts, common in both CS and MRA approaches, occur when interpolation errors or inaccurate detail injection introduce structural discontinuities along object boundaries. Finally, geometric misregistration between PAN and MS images can introduce double edges or local warping along structural boundaries, which are often overlooked by traditional IQA metrics [27,55]. These artifacts are strongly influenced by the design of the injection model, the spatial resolution ratio between the PAN and MS sensors, and the accuracy of the initial co-registration [52]. Because CS methods prioritize spatial sharpness while MRA methods prioritize spectral consistency, evaluating PS quality requires metrics capable of capturing both dimensions simultaneously. The need for addressing this duality led to the development of joint quality indicators, such as the Quality with No Reference (QNR) index [56], which attempts to quantify both spatial and spectral degradations within a single, no-reference framework. Finally, robust pan-sharpening requires a balance between structural enhancement and radiometric preservation to prevent the introduction of misleading artificial features.

3.2. PS Quality Assessment Protocols

The objective of pan-sharpening is to create images that combine high spatial resolution with the original spectral information. However, since a physical full-resolution ground-truth image does not exist, evaluating the fidelity of the fused image requires specialized frameworks. These assessment protocols dictate how the source and fused data should be prepared and compared. An essential prerequisite for all objective protocols is highly accurate, sub-pixel co-registration of the input PAN and MS imagery. Any geometric misalignment propagates as severe quality degradation during the assessment. Several widely used pan-sharpening protocols have been developed, as summarized in Table 1.

Table 1. Comparison of Image quality assessment protocols.

Protocol	Characteristics & Methodology
Wald's Protocol	**Focus:** Consistency and Synthesis **Metrics:** Any full-reference metric (e.g., ERGAS, SAM, Q-index) **Method:** Generates reduced-scale dataset via simultaneous MTF decimation; degrades fused image to match original MS resolution **Techniques:** MTF decimation filter for consistency

Table 1. *Cont.*

Protocol	Characteristics & Methodology
Zhou's Protocol	**Focus:** Separate Spectral and Spatial Quality **Metrics:** Absolute difference (spectral), SRCC (spatial) **Method:** Compares fused image and resampled MS; uses Laplacian filter **Techniques:** Spatial detail extraction via Laplacian
Khan's Protocol	**Focus:** Spectral and Spatial Distortion **Metrics:** Q4 index across scales **Method:** MTF filters, high-pass complements, Q4 calculations **Techniques:** Separate D_λ and D_S indices
Ghassemian's Protocol	**Focus:** Separate Spectral and Spatial Quality **Metrics:** Any distance measure for spatial; SAM for spectral **Method:** Weighted PAN comparison, downsampling of fused image **Techniques:** Spatial quality with weighted synthetic PAN
VQAP	**Focus:** Subjective Visual Quality **Metrics:** 23 Criteria (spatial, spectral, object-level) **Method:** Expert weighting, automated answer sheet for subjective MOS **Techniques:** Human-perception benchmark

3.2.1. Wald's Protocol

The most widely used method for the objective assessment of MS pan-sharpening outputs is Wald's protocol [52,57]. It relies on two key principles: consistency and synthesis. Following standard multi-resolution evaluation frameworks [58], the protocol establishes a reduced-resolution testing environment to evaluate the synthesis property. This setup is achieved by spatially downsampling both the original PAN and MS images using sensor-matched MTF filters. The principle of synthesis encompasses two main attributes. First, the PS image should closely match the ideal image that would be observed by the corresponding MS sensor at the highest possible resolution. Second, the set of PS images should resemble the set of MS images observed by the corresponding sensor at its maximum resolution. Because high-resolution MS references do not physically exist, evaluating synthesis requires this reduced-scale framework, where the original MS image serves as the ground truth against the fused output of the downsampled inputs [59,60]. Consequently, Wald's protocol does not mandate a specific mathematical metric, but allows the use of any full-reference analyzer (e.g., ERGAS, SAM) on the reduced-scale data. The complementary principle of consistency operates at the full scale. It asserts that a pan-sharpened image, when spatially degraded back to the resolution of the original MS image, must closely resemble that original MS input. Accurate spatial degradation requires the use of decimation filters tuned to the MTF of the capturing sensor, ensuring that the downsampled fused image realistically simulates the sensor's acquisition characteristics [61].

3.2.2. Zhou's Protocol

To bypass the limitations of resolution degradation, Zhou's protocol assesses spectral and spatial qualities separately at full scale (FS) [62]. For spectral quality, the protocol evaluates each band by calculating the average absolute difference between the fused image and the upsampled original MS image. In an ideal fusion scenario, this difference should approach zero. For spatial quality assessment, the protocol extracts spatial details from both the PAN and MS images using a Laplacian filter. It then calculates the Spearman's Rank Correlation Coefficient (SRCC) between the high-pass information extracted

from the PAN image and each band of the PS image. The correlation values approaching 1 indicate that the high-frequency spatial details of the PAN image have been accurately transferred to the pan-sharpened product. A persistent limitation of this FS spatial assessment, however, is the assumption that PAN structural details map uniformly across all MS bands. This assumption frequently violates the actual spectral response characteristics of the optical sensor.

3.2.3. Khan's Protocol

Khan's protocol introduces two separate indices to address spatial and spectral distortions at the full-resolution scale, separating these degradations without relying on the QNR formulation [63]. The distortion is quantified using the $Q4$ metric or its generalized form $Q2^n$ for more than four bands. Let $\hat{\mathbf{E}}$ denote the full-resolution pan-sharpened image, $\mathbf{M}$ the original low-resolution multispectral image, and $\mathbf{P}$ the panchromatic image. To quantify spectral distortion (D_λ), an MTF-based downsampling filter is applied to $\hat{\mathbf{E}}$, producing a degraded image $\hat{\mathbf{E}}_d$. The spectral distortion index is then calculated by comparing this degraded result with the original MS image as:

$$D_\lambda^{\text{Khan}} = 1 - Q4(\hat{\mathbf{E}}_d, \mathbf{M}). \tag{1}$$

For spatial distortion (D_S), high-pass complements of MTF filters are employed to extract high-frequency information. The process involves applying MTF filters to each band of the fused MS image, then subtracting the filtered image from the fused MS to isolate structural details. The same procedure is applied to the original $\mathbf{P}$ image. The $Q4$ index is then calculated between high-frequency details. The procedure is repeated at a reduced scale using the original MS and downsampled PAN images, and the spatial distortion is determined by the absolute difference in $Q4$ detail similarities across the two scales [64].

3.2.4. Ghassemian's Protocol

Ghassemian's protocol evaluates spectral and spatial qualities independently [65]. For spatial quality, a weighted average is computed from the bands of the fused image to generate a synthetic PAN image. The synthesized PAN image is then compared to the original high-resolution PAN image using a suitable distance measure, such as Euclidean distance. A critical factor in this protocol is defining the weights used to generate the synthetic PAN, which should correspond to the Spectral Response Function (SRF) of the specific sensor. To assess spectral quality, the fused image is downsampled to create a low-resolution version, which is then compared directly to the original MS image. The dissimilarity measure can be any standard method that quantifies spectral angular or radiometric differences, such as SAM.

3.2.5. The Visual Quality Assessment Protocol (VQAP)

While mathematical protocols provide objective quantification, the Visual Quality Assessment Protocol (VQAP) was introduced to evaluate image quality subjectively through standardized human perception. Unlike the preceding protocols, VQAP serves as a perceptual benchmark rather than an automated quality assessment framework. VQAP evaluates essential criteria categorized into spatial enhancement (e.g., sharpness), color preservation (spectral fidelity), and object recognition across global, regional, and local scales [66]. Weighting factors, derived from expert responses, are applied to these 23 distinct criteria to reflect their relative importance. Although VQAP ensures a more structured and reproducible form of Mean Opinion Score (MOS) assessment through automated answer sheets, its reliance on subjective human input makes it inherently time-consuming and difficult to integrate into large-scale, automated evaluation pipelines.

4. Fundamentals of Image Quality Assessment

It is essential to distinguish between evaluation protocols and specific IQA metrics. While protocols establish experimental methodology by defining how data should be scaled and filtered, IQA measures are mathematical formulations calculating quantitative error or similarity. Objective quality metrics serve multiple purposes, including monitoring degradation and benchmarking processing algorithms [67,68]. In general image processing, quality assessment heavily relies on the Human Visual System [69]. In pan-sharpening, however, mathematical and radiometric fidelity is as important as perceptual alignment, particularly when fused images serve automated downstream tasks like land cover classification. Comparing objective metrics with subjective human evaluations is therefore necessary to quantify the perceptual gap and justify the performance limits of mathematical models.

To evaluate objective IQA performance against human ratings, standardized statistical correlation metrics are utilized [70].

The SRCC measures prediction monotonicity. It is calculated as:

$$\mathrm{SRCC} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \tag{2}$$

where $d_i = \mathrm{rank}(P_i) - \mathrm{rank}(S_i)$ represents the difference between predicted ($\mathrm{rank}(P_i)$) and subjective ($\mathrm{rank}(S_i)$) ranks, and n is the total number of evaluated images.

The Kendall Rank-Order Correlation Coefficient (KRCC) evaluates ordinal association and is robust to outliers. In its standard τ_a form, it is expressed as:

$$\mathrm{KRCC} = \frac{n_c - n_d}{0.5n(n-1)}, \tag{3}$$

where n_c and n_d denote concordant and discordant pairs. Adjusted variants are used when tied ranks are present.

The Pearson Linear Correlation Coefficient (PLCC) measures linear correlation to evaluate prediction accuracy:

$$\mathrm{PLCC} = \frac{\overline{\mathbf{P}}^T \overline{\mathbf{S}}}{\sqrt{\left(\overline{\mathbf{P}}^T \overline{\mathbf{P}}\right)\left(\overline{\mathbf{S}}^T \overline{\mathbf{S}}\right)}}, \tag{4}$$

where $\overline{\mathbf{P}}$ and $\overline{\mathbf{S}}$ are mean-removed vectors of predicted and subjective scores.

The Root Mean Square Error (RMSE*) quantifies absolute prediction error relative to subjective scores:

$$\mathrm{RMSE}^* = \sqrt{\frac{(\mathbf{P}-\mathbf{S})^T(\mathbf{P}-\mathbf{S})}{n}}, \tag{5}$$

where $\mathbf{P}$ and $\mathbf{S}$ are vectors after non-linear mapping. This metric differs from conventional pixel-wise remote sensing RMSE.

Computing PLCC and RMSE* requires mapping objective scores to the subjective domain due to non-linear scaling. A standard 5-parameter logistic mapping function recommended by the Video Quality Experts Group (VQEG) is employed:

$$\hat{\mathbf{P}} = y_1\left(\frac{1}{2} - \frac{1}{1+\exp(y_2(\mathbf{P}-y_3))}\right) + y_4\mathbf{P} + y_5, \tag{6}$$

where $\hat{\mathbf{P}}$ denotes mapped scores, $\mathbf{P}$ represents raw predictions, and $y_1, \ldots, y_5$ are optimized regression parameters.

5. Remote Sensing Sensors and Datasets

Remote sensing datasets utilized in pan-sharpening are acquired via diverse Earth observation sensors. Their spatial and spectral parameters dictate the scale ratio between PAN and MS channels, which influences the performance of fusion algorithms and the reliability of IQA metrics. The selection of a dataset represents a trade-off between spatial detail, spectral fidelity, and revisit frequency [45]. Understanding sensor-specific configurations is essential for accurate pan-sharpening evaluation.

5.1. Sensor Overview and Characteristics

Table 2 provides a comparative synthesis of the spatial resolutions, scale ratios, and spectral wavelength ranges for major platforms.

Table 2. Comparative summary of spatial resolutions, scale ratios, and spectral wavelength ranges for major Earth observation sensors. For SPOT-5, the scale ratio depends on the PAN acquisition mode.

Satellite	PAN Res. (m)	MS Res. (m)	Scale Ratio	Spectral Range (nm) [PAN / MS]
Landsat-8	15	30	1:2	500–680/450–880 (MS)
SPOT-5	2.5–5	10	1:2 or 1:4	480–710/500–890 (MS)
Gaofen-2	1	4	1:4	450–900/450–890 (MS)
IKONOS	1	4	1:4	450–900/450–850 (MS)
QuickBird	0.6	2.4	1:4	450–900/450–900 (MS)
WorldView-2	0.5	2.0	1:4	450–800/400–1040 (MS)
GeoEye-1	0.5	2.0	1:4	450–800/450–900 (MS)
Pleiades-1	0.5	2.0	1:4	470–830/430–940 (MS)
WorldView-3	0.31	1.24	1:4	450–800/400–1040 (MS + SWIR)

The majority of commercial satellites (e.g., WorldView series, GeoEye-1, Pleiades, QuickBird, IKONOS) operate with a 1:4 scale ratio, imposing strict requirements on fusion algorithms to preserve structural details. The WorldView-3 sensor is distinguished by its extended spectral coverage, which includes Short-Wave Infrared (SWIR) bands [71]. Conversely, sensors such as Landsat-8 and SPOT-5 operate with distinct scale ratios (1:2 and 1:4, respectively) and specific MTF profiles, providing testbeds for evaluating the scale-invariance of IQA metrics [72]. Gaofen-2 (GF-2) is frequently utilized in deep learning benchmarks due to its high global revisit frequency and spatial resolution [73]. While QuickBird and IKONOS remain foundational datasets for evaluating classical algorithms [74,75], their broad PAN bands frequently induce non-linear spectral distortions.

5.2. Benchmark Datasets and Emerging Trends

Contemporary research increasingly relies on standardized datasets, such as the PAirMax benchmark [27], which provides pre-processed, co-registered image pairs. These benchmarks are vital to mitigate variability introduced by different preprocessing pipelines. Beyond multispectral fusion, recent research is extending pan-sharpening to hyperspectral imagery and dynamic CubeSat constellations like PlanetScope. Although CubeSats often operate near a 1:1 scale ratio, they are gaining attention for multi-temporal fusion and super-resolution applications, representing an emerging frontier in operational remote sensing.

5.3. Implications for Quality Assessment

The discrepancies in spectral and spatial configurations have direct implications for IQA. Specifically, when the spectral response of the PAN band does not fully encompass or align with the individual ranges of the MS channels, the scale-invariance assumptions inherent in consistency-based metrics such as the QNR index are invalidated. Furthermore, the MTF profiles governing spatial frequency attenuation during acquisition must be

explicitly considered by metrics like EFM. Failure to account for sensor-specific MTF characteristics or spectral mismatch leads to biased evaluation scores, misrepresenting the actual spatial fidelity.

6. Pan-Sharpening IQA Methods

To evaluate PS images, numerous methods have been developed, employing diverse methodologies. This is because PS images are well-suited for human interpretation and are highly effective for a range of applications. The characteristics of representative PS IQA methods are presented below for better categorization. Although this review primarily focuses on no-reference image quality assessment methods, several widely used full-reference metrics are also included due to their historical importance and continued use in benchmarking and comparative evaluation of pan-sharpening algorithms. Many recent NR-IQA approaches are additionally validated against conventional reference-based measures, making them relevant for a comprehensive analysis of PS image quality assessment methods. Representative pan-sharpening IQA methods can be grouped according to the availability of reference data and the adopted evaluation strategy. Full-reference (FR-IQA) methods require a reference image and remain widely used in remote sensing due to their interpretability and long-established role in benchmarking pan-sharpening performance. This group includes metrics such as ERGAS, Q4, Q-index, Q2n, SAM, SSIM, EFM, QLR, QHR, JQM, object-based quality evaluation procedures, and FS quality evaluation methods. These approaches typically evaluate spectral fidelity, spatial preservation, or structural similarity between the fused and reference images.

A second group consists of reduced-reference and hybrid approaches, which estimate image quality using partial information derived from PAN and MS inputs without relying on a complete ground truth image. Representative examples include QNR or GQNR. These methods are particularly relevant in practical remote sensing scenarios where reference high-resolution multispectral images are unavailable and are commonly used in pan-sharpening evaluation when full reference data are not available.

The third group includes classical NR-IQA methods that do not require any reference image and rely instead on natural scene statistics or perceptual features. Representative examples are NIQE and opinion-unaware learning approaches based on NIQE and IL-NIQE.

The fourth group comprises deep learning-based NR-IQA methods, which learn hierarchical feature representations to predict perceptual quality in a data-driven manner. This category includes DIQA, MDFS, the NR full-resolution evaluation method with end-to-end learning, opinion-aware and opinion-unaware pan-sharpened image quality analyzers, MLPSIE, TBN-PSI, and deep learning-based blind full-resolution assessment methods.

6.1. Full-Reference Methods

Full-reference methods evaluate fused images by comparing them directly against reference datasets. Because many classical reference-based metrics (such as SSIM or UIQI) were originally designed for single-channel grayscale images, their extension to multispectral and hyperspectral remote sensing data relies on two adaptation strategies. The first is band-wise aggregation, where a metric is computed independently for each spectral channel and subsequently averaged to yield a global index (e.g., QLR). The second is a multivariate or hypercomplex framework, which treats multi-band pixel vectors as unified mathematical entities, such as quaternions in Q_4, to jointly evaluate spectral and spatial fidelity across all channels simultaneously.

Wald et al. [52] introduced a metric for evaluating the overall quality of integrating PAN and MS data, termed ERGAS (Erreur Relative Globale Adimensionnelle de Syn-

thèse/Relative Dimensionless Global Error in Synthesis) [52]. For an image comprising N spectral channels, it is defined as

$$\mathrm{ERGAS} = \frac{100}{C}\sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\mathrm{RMSE}_j}{\mu_j^{GT}}\right)^2}, \tag{7}$$

where C represents the spatial resolution ratio between the multispectral and panchromatic images, RMSE_j denotes the root mean square error between the reference and fused images for the j-th spectral channel (i.e., the pixel-wise radiometric error commonly used in remote sensing), and μ_j^{GT} is the mean intensity of the j-th channel of the reference image. ERGAS functions as a distortion metric, where lower values correspond to improved fusion outcomes. An ERGAS value of zero indicates perfect fusion. ERGAS is not upper-bounded, meaning there is no maximum threshold beyond which the index loses its significance [76].

The Universal Image Quality Index (UIQI or Q) was originally developed as an evaluative metric for monochrome images. Its logic was later extended to multi-band images through Q_4, which leverages hypercomplex numbers (quaternions) to represent spectral pixel vectors [77]. This advancement enables a holistic, joint evaluation of a four-band MS image, rather than processing each spectral band independently. The Q_4 index is formulated as follows [78]:

$$Q_4 = \frac{4|\sigma_{uv}| \cdot |\overline{u}| \cdot |\overline{v}|}{(\sigma_u^2 + \sigma_v^2)(|\overline{u}|^2 + |\overline{v}|^2)} = \frac{|\sigma_{uv}|}{\sigma_u \cdot \sigma_v} \cdot \frac{2|\overline{u}| \cdot |\overline{v}|}{|\overline{u}|^2 + |\overline{v}|^2} \cdot \frac{2\sigma_u\sigma_v}{\sigma_u^2 + \sigma_v^2}, \tag{8}$$

where u and v represent the quaternions associated with a four-band reference image and the fused image, respectively, and $\overline{u}$ and $\overline{v}$ denote their corresponding local mean quaternion values computed over sliding blocks. They are defined as:

$$u = u_1 + \mathrm{i}u_2 + \mathrm{j}u_3 + \mathrm{k}u_4, \tag{9}$$

$$v = v_1 + \mathrm{i}v_2 + \mathrm{j}v_3 + \mathrm{k}v_4. \tag{10}$$

As illustrated in Equation (8), the Q_4 metric is composed of three distinct factors. The first term corresponds to the modulus of the hypercomplex correlation coefficient between u and v, reflecting the structural and spectral consistency. The second term assesses the mean bias (luminance distortion), while the third term evaluates contrast distortion across all bands simultaneously. In practice, local statistics (expectations, variances, and covariances) are computed over $M \times M$ sliding blocks, typically set to $M = 16$ or $M = 32$. To obtain a global quality measure for the entire image, the local Q_4 values are averaged across all blocks. The resulting Q_4 index ranges from $[0, 1]$, where the theoretical optimal value of 1 corresponds to a perfect alignment between the fused image and the reference, indicating an ideal fusion outcome [79].

In the Universal Image Quality Index (UIQI), the term "universal" signifies that this assessment method is independent of the specific images being evaluated, the conditions under which they are viewed, or the particular observers conducting the assessment [80]. UIQI is designed to be applicable across diverse image processing applications, enabling comparisons among distinct types of image distortions. To test the effectiveness of this approach, the authors subjected images to various distortions, including salt-and-pepper noise, additive Gaussian noise, multiplicative speckle noise, mean shift, contrast stretching, blurring, and JPEG compression. Their findings demonstrate the metric's ability to gauge the loss of information occurring during image degradation, in contrast to the Mean

Squared Error (MSE) metric, which primarily measures error energy rather than true information loss. The formulation of the proposed quality index is expressed as follows:

$$Q' = \frac{4\sigma_{xy}\,\overline{x}\,\overline{y}}{\left(\sigma_x^2 + \sigma_y^2\right)\left(\overline{x}^2 + \overline{y}^2\right)}, \tag{11}$$

where

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \overline{y} = \frac{1}{N}\sum_{i=1}^{N} y_i, \tag{12}$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2, \quad \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y})^2, \tag{13}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y}). \tag{14}$$

In the modified quality metric formula Q', each variable has a specific role. The metric Q' itself represents the quality measure being calculated, which assesses the similarity between two datasets, x and y. The symbols x_i and y_i refer to individual values within datasets x and y, respectively. The terms $\overline{x}$ and $\overline{y}$ denote the mean values of datasets x and y, calculated as the average of all values in each dataset. The variables σ_x^2 and σ_y^2 are the variances of datasets x and y, quantifying the spread of values around their respective means. Lastly, σ_{xy} is the covariance between datasets x and y, which captures the extent to which the two datasets vary together. Together, these terms allow Q' to effectively measure the correlation between x and y by integrating information on their means, variances, and covariance [81].

The $Q2^n$ quality index serves as a generalization of the UIQI as defined in [80], applying to an original image signal denoted as σ_w and a test image signal labeled σ_x. This relationship can be articulated as follows:

$$Q_{K\times K} = \frac{\sigma_{wx}}{\sigma_w\sigma_x} \cdot \frac{2\overline{w}\,\overline{x}}{\overline{w}^2 + \overline{x}^2} \cdot \frac{2\sigma_w\sigma_x}{\sigma_w^2 + \sigma_x^2}. \tag{15}$$

Here, σ_w signifies the standard deviation of the image feature w, while σ_x represents the standard deviation of the image feature x. These computations are performed within a given $K \times K$ block. The first term corresponds to the Cross-Correlation (CrC), the second term captures changes in mean bias, and the third term gauges alterations in contrast. An extension of the Q_4 index, the new index $Q_{K\times K}$ assumes real values within the range of $[0, 1]$, where 1 signifies the optimal value [82].

Javan et al. [83] introduced an Edge-based Image Quality Metric (EFM) designed to assess the spatial quality of fused images by focusing on the preservation of structural details. The metric operates on the principle that spatial distortions introduced during pan-sharpening manifest as uneven or dispersed edge responses. The evaluation of these spatial responses relies on the MTF, which characterizes the imaging system's resolution. EFM extracts step-edges using the Canny operator, derives the Line Spread Function (LSF) from edge profiles, and compares resulting MTF curves against reference edges [84,85].

The Spectral Angle Mapper (SAM) is a widely adopted metric for quantifying spectral distortion by treating the spectra of pixels as vectors in an N-dimensional space [86]. A key advantage of SAM is its invariance to uniform changes in pixel illumination or overall intensity. For a given pixel, the spectral angle is computed using the following equation:

$$\mathrm{SAM}(\mathbf{u}, \hat{\mathbf{u}}) = \arccos\left(\frac{\langle \mathbf{u}, \hat{\mathbf{u}} \rangle}{\|\mathbf{u}\|_2 \cdot \|\hat{\mathbf{u}}\|_2}\right), \tag{16}$$

where **u** and **û** are the spectral vectors of the reference and pan-sharpened pixels, respectively. Lower SAM values indicate superior spectral fidelity. The Structural Similarity Index Measure (SSIM), introduced by Wang et al. [87], provides an alternative method for evaluating image quality by isolating luminance, contrast, and structure [88,89]. Applying SSIM locally using a sliding window approach generates a spatial SSIM map, offering detailed topographical insight into local degradations [90]. The formulation between a reference block u and a fused block $\hat{u}$ is expressed as:

$$\mathrm{SSIM}(u, \hat{u}) = \frac{(2\mu_u\mu_{\hat{u}} + C_1)(2\sigma_{u\hat{u}} + C_2)}{(\mu_u^2 + \mu_{\hat{u}}^2 + C_1)(\sigma_u^2 + \sigma_{\hat{u}}^2 + C_2)}, \tag{17}$$

where C_1 and C_2 are small positive constants included to stabilize the division when the denominators approach zero. Additional reference measures include the Quality for Low Resolution (QLR) index, which averages SSIM across downsampled spectral bands [91,92]:

$$\mathrm{QLR} = \frac{1}{m}\sum_{i=1}^{m}\mathrm{SSIM}(dsi, (fdi * lpf)\downarrow), \tag{18}$$

and the Quality for High Resolution (QHR) index, which compares PAN images with averaged fused bands [93]:

$$\mathrm{QHR} = \mathrm{SSIM}\left(pan, \frac{1}{m}\sum j = 1^{m} fdj\right), \tag{19}$$

where ds_i denotes the i-th band of the original multispectral image, fd_i and fd_j represent the corresponding bands of the fused product, and lpf is a low-pass filter applied prior to spatial decimation to match the resolution of the original multispectral data. The Joint Quality Measure (JQM) combines QLR and QHR into a single normalized index [94].

6.2. Consistency-Based No-Reference Metrics

Consistency-based methods evaluate fusion quality at full scale without reference data. The QNR index evaluates full-resolution performance by measuring spectral (D_λ) and spatial (D_s) distortions based on inter-band and inter-sensor relationships [95]:

$$\mathrm{QNR} = (1 - D_\lambda)^\alpha (1 - D_s)^\beta, \tag{20}$$

where α and β control component weighting. Generalized variants like GQNR replace UIQI with SSIM to handle complex sensors and SWIR bands [96,97]. Opinion-aware models (Q_{OA}) map quality-conscious features to subjective Mean Opinion Scores using Support Vector Regressors [98]. Recent deep learning approaches utilize tripartite encoder-decoder architectures to regress analytical QNR scores directly from blind inputs [99].

6.3. Statistical No-Reference Metrics

Statistical methods analyze Natural Scene Statistics (NSS) to quantify degradation without reference images.

The Natural Image Quality Evaluator (NIQE) extracts Mean Subtracted Contrast Normalized coefficients, computes feature vectors from local patches, and measures Mahalanobis distance against a pristine Multivariate Gaussian model [100–103]:

$$D = \sqrt{(\nu_1 - \nu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\nu_1 - \nu_2)}, \tag{21}$$

where ν and Σ represent mean vectors and covariance matrices. Kwan et al. adapted NIQE for pan-sharpened satellite data [96]. Other statistical frameworks construct empirical Multivariate Gaussian (MVG) models using multi-band NSS features, chroma maps, and spatial bandpass statistics to capture blur and color inconsistencies [8,104–106].

6.4. Deep Learning-Based Methods

Deep learning models extract hierarchical representations to predict perceptual quality.

The Deep Pansharpened IQA (DIQA) model computes similarities between deep features derived from source images and fused outputs [107]. The Multi-scale Deep Feature Statistic (MDFS) integrates deep features with MVG modeling [108]. Other frameworks employ multi-stream architectures to regress quality scores against Differential Mean Opinion Scores (DMOS) values [109]. The Multi-Level Pan-Sharpening Images Evaluator (MLPSIE) extracts features using pre-trained networks like ResNet-18 [110] and VGG-19 [111] across True-Color and Pseudo-Color representations [112]. Features are reduced using Kernel Principal Component Analysis and regressed via SVR [113]. The Three-Branch Neural Network (TBN-PSI) utilizes a truncated Inception-v3 backbone to process dual-channel inputs with cross-branch feature exchange, optimizing directly for structural similarity loss [114].

6.5. Object-Based and Task-Driven Assessment

Object-based evaluation segments fused images into geographical objects using watershed transforms to compute localized spatial and spectral quality maps [115–117]. Task-driven assessment validates fusion methods based on their impact on automated downstream applications, such as land-cover classification or object detection. This approach addresses the limitation where high mathematical scores obscure critical semantic information. However, task-driven outcomes depend heavily on label quality, classifier architecture, and scene type, making cross-study comparisons challenging.

7. Critical Review of Quality Metrics

The evaluation of PS image quality remains a complex task due to the need to simultaneously preserve both spectral fidelity and spatial detail. Existing metrics exhibit distinct biases, prioritizing either visual structural fidelity, mathematical signal preservation, or semantic utility. This section provides a critical synthesis of IQA metrics, focusing on their inherent limitations, evaluation biases, and failure conditions.

To provide a clear reference framework, Table 3 summarizes the reviewed metrics according to the established structural categorization, detailing their prerequisites, context, advantages, and limitations.

Table 3. Representative pan-sharpening IQA metrics.

Metric	Characteristics & Evaluation Profile
ERGAS	**Category:** Full-Reference **Prerequisites:** Reference MS **Context:** Global radiometric error **Advantages:** Normalized across bands; intuitive **Limitations:** Ignores local structural changes
$Q_4/Q2^n$	**Category:** Full-Reference **Prerequisites:** Reference MS **Context:** Spectral and spatial consistency **Advantages:** Evaluates multivariate vector consistency **Limitations:** Complex formulation; window size sensitive

Table 3. *Cont.*

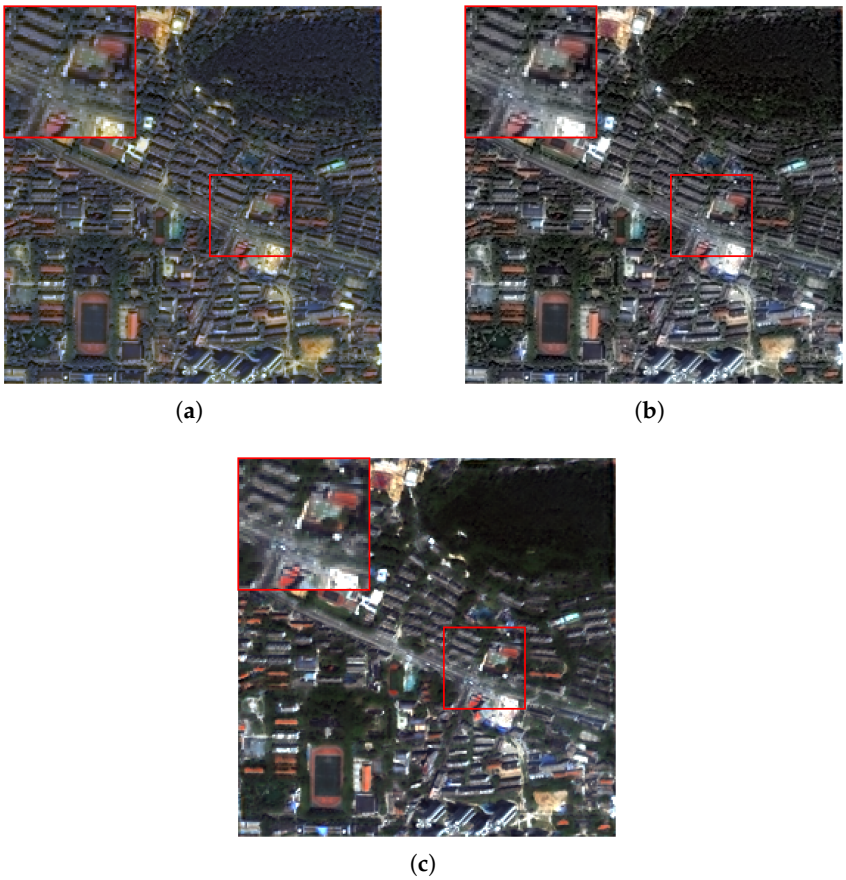
Metric	Characteristics & Evaluation Profile
SSIM	**Category:** Full-Reference **Prerequisites:** Reference image **Context:** Structural fidelity **Advantages:** High correlation with human vision **Limitations:** Not designed for multispectral data
SAM	**Category:** Full-Reference **Prerequisites:** Reference MS **Context:** Spectral angle **Advantages:** Insensitive to global illumination changes **Limitations:** Completely ignores spatial distortions
QNR	**Category:** Consistency No-Reference **Prerequisites:** PAN and MS inputs **Context:** Full-resolution global evaluation **Advantages:** No reference required; standard benchmark **Limitations:** Sensitive to arbitrary weights (α, β)
GQNR	**Category:** Consistency No-Reference **Prerequisites:** PAN and MS inputs **Context:** Full-resolution structure **Advantages:** Incorporates SSIM instead of UIQI **Limitations:** Dependent on decimation filtering steps
NIQE	**Category:** Statistical No-Reference **Prerequisites:** None (Pristine MVG model) **Context:** Blind quality estimation **Advantages:** No subjective training scores needed **Limitations:** Tuned for natural images, not satellite sensors
DIQA	**Category:** Deep Learning No-Reference **Prerequisites:** PAN, MS, and PS images **Context:** Deep feature representation **Advantages:** Captures complex non-linear dependencies **Limitations:** High computational cost; severe overfitting
MLPSIE	**Category:** Deep Learning No-Reference **Prerequisites:** Pre-trained CNN backbones **Context:** Perceptual prediction **Advantages:** Avoids end-to-end training via KPCA **Limitations:** Dependent on chosen pre-trained architecture
TBN-PSI	**Category:** Deep Learning No-Reference **Prerequisites:** Large training dataset **Context:** End-to-end multi-level representation **Advantages:** Bridges low-level details and high-level semantics **Limitations:** Computationally heavy; low interpretability

7.1. Reference-Based Metrics

Reference-based metrics exhibit a fundamental dichotomy between global radiometric assessment and local structural evaluation. ERGAS provides a normalized global distortion measure but completely masks localized structural degradations. A fusion algorithm producing severe but spatially isolated artifacts can still achieve favorable ERGAS scores. In contrast, metrics derived from UIQI, including Q_4 and $Q2^n$, evaluate correlation and luminance locally. However, their reliability is strictly bound to the arbitrary selection of sliding window sizes. The lack of standard window dimensions prevents objective cross-study standardization. Furthermore, specific metrics exhibit extreme domain biases.

SSIM prioritizes human visual perception by isolating luminance and contrast, rendering it insensitive to multispectral chromatic aberrations. It effectively evaluates structural injection but fails to penalize spectral degradation. Conversely, SAM measures spectral angle vectors, ignoring spatial contrast entirely. Similarly, EFM evaluates spatial fidelity strictly through MTF analysis; while this physical grounding makes it effective for structural assessment in urban scenes, it completely omits radiometric preservation. The primary limitation of all FR metrics remains their dependence on reduced-scale protocols. High performance on spatially degraded datasets does not guarantee equivalent fusion fidelity at full sensor resolution.

To illustrate the application of these metrics across different algorithmic paradigms, Figure 2 presents a visual and quantitative comparison of image fusion outputs generated by the IHS [29], ATWT (à Trous Wavelet Transform) [33], and A-PNN [118] methods for a structurally complex urban scene. The IHS method yields the highest error metrics (SAM = 5.7580, ERGAS = 3.9037), indicating severe spectral distortion, although it maintains spatial detail alongside localized artifacts. The ATWT method produces lower error values (SAM = 4.5573, ERGAS = 3.0990), reflecting improved spectral and structural preservation relative to IHS. The A-PNN architecture achieves the lowest values for both metrics (SAM = 3.6494, ERGAS = 2.5244), demonstrating the highest spectral consistency with the reference image among the analyzed algorithms.



(**a**) (**b**)

(**c**)

Figure 2. Visual and quantitative comparison of pan-sharpening outputs (IHS, ATWT, and A-PNN) for an urban scene. The corresponding full-reference metrics (SAM and ERGAS) illustrate the variation in spectral distortion across the component substitution, multiresolution analysis, and deep learning paradigms. (**a**) IHS (SAM 5.7580, ERGAS 3.9037). (**b**) ATWT (SAM 4.5573 ERGAS 3.0990). (**c**) A-PNN (SAM 3.6494 ERGAS 2.5244).

7.2. Consistency-Based No-Reference Metrics

Consistency-based methods evaluate full-resolution products by isolating spatial and spectral properties. QNR and GQNR provide hybrid evaluations by combining spatial and spectral distortion indices.

A methodological limitation in these consistency indices is their strict reliance on scale-invariance assumptions. QNR assumes the inter-band relationships remain constant across spatial scales. When the spectral response of the broad PAN band does not fully encompass the specific ranges of the MS channels, this assumption is physically invalidated. Consequently, the metric penalizes mathematically valid fusion outputs. Furthermore, hybrid metrics depend on arbitrary weighting parameters (α and β) to combine distinct spatial and spectral distortion terms. Adjusting these parameters alters the performance rankings of different algorithms.

As illustrated in Figure 3, the empirical results indicate a significant dependence of the final quality score on the value of the exponent β. Although all analyzed methods exhibit a decrease in QNR values as β increases, the rate of decline varies across algorithms. Modifying the β parameter affects both the absolute QNR values and the relative ranking of the evaluated methods. For instance, the scores for Gram–Schmidt Adaptive (GSA) [119], EXP, and Brovey Transform with Haze correction (BT-H) [27] converge at $\beta = 1.0$, whereas at $\beta = 1.5$, their ranking is altered. This demonstrates that the QNR outcome is highly sensitive to parameterization, which severely limits the objective utility of the metric if default coefficients are adopted without justification.

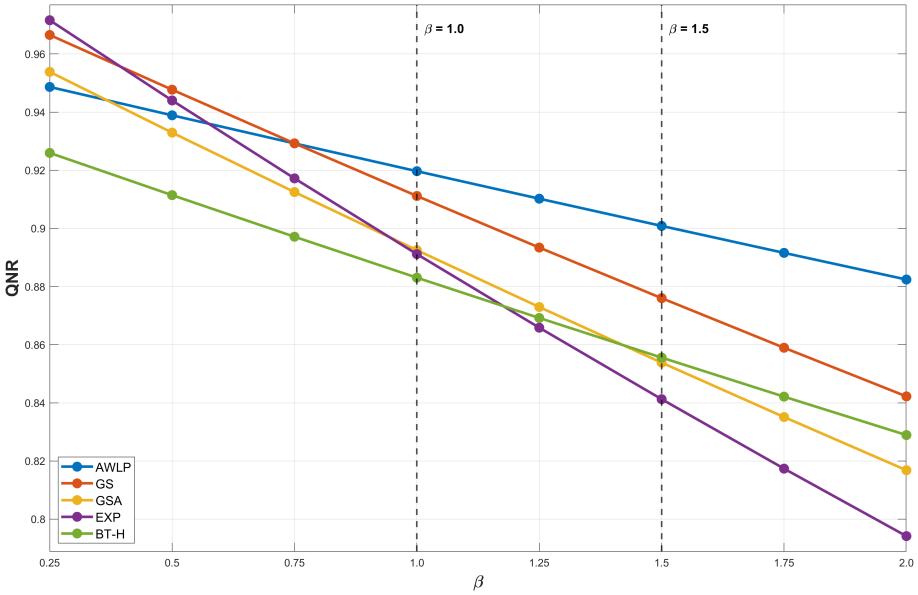


Figure 3. Impact of the β parameter on QNR evaluation scores. The results demonstrate rank volatility among exemplary algorithms resulting from changes in the spectral weighting exponent.

7.3. Statistical No-Reference Metrics

Statistical NSS models eliminate the need for reference imagery by measuring deviations from a pristine feature distribution. The base NIQE model [100] assumes natural scene statistics, which diverge significantly from the radiometric profiles of optical satellite imagery [120]. Adapted models require retraining on domain-specific high-resolution datasets. Such a process introduces strong sensor bias. Statistical frameworks trained on one specific satellite constellation (e.g., WorldView-3) often fail to generalize to data with different radiometric characteristics (e.g., Gaofen-2) due to sensor-specific feature distributions that are not transferable across platforms [121]. Furthermore, extracting complex multiband NSS features significantly increases computational overhead [100] while

providing limited spatial localization of distortions due to the global nature of statistical feature aggregation [95]. Global statistical evaluators identify widespread anomalies but struggle to pinpoint discrete structural artifacts introduced by specific fusion techniques.

7.4. Deep Learning-Based Methods

Deep learning frameworks transition IQA from mathematical signal processing to data-driven feature extraction. Architectures such as DIQA, MDFS, MLPSIE, and TBN-PSI map complex non-linear distortions more effectively than statistical models. They generally achieve higher correlation with human visual perception.

However, this representational capacity introduces severe cross-sensor overfitting. DL models internalize the specific artifacts and noise profiles of their training datasets. When evaluated on unseen sensors or disparate geographical features, their predictive accuracy drops sharply, revealing a fundamental limitation in generalization. End-to-end models also lack interpretability. They function as black boxes, obscuring whether a poor quality score results from spectral leakage, spatial blurring, or simple radiometric shifting. Finally, supervised opinion-aware models depend on large annotated datasets with DMOS labels, which are resource-intensive to produce and inherently subject to human observer bias.

7.5. Object-Based and Task-Driven Assessment

Object-based evaluation segments images to assess localized structural boundaries. Its reliability depends entirely on the accuracy of the underlying watershed or segmentation algorithms. Segmentation errors directly corrupt the subsequent quality assessment.

Task-driven assessment shifts the validation paradigm from mathematical pixel errors to semantic utility. Mathematical fidelity does not strictly correlate with performance in downstream automated applications. A pan-sharpened image may achieve optimal scores on traditional IQA metrics while introducing subtle spectral distortions that degrade land cover classification or object detection frameworks. Validating fusion methods based on their operational impact highlights the perceptual gap. It proves that algorithms prioritizing visual sharpness for human interpretation often disrupt the spectral signatures required by machine-driven feature extraction protocols.

7.6. Analysis of Objective vs. Subjective Rankings

The limitations of existing metrics are visible in the qualitative assessment of competing algorithms, as illustrated in Figure 4 [8]. The figure compares objective metric values with the geometric mean of DMOS scores (DMOS-GM) obtained from subjective expert evaluations of the true and pseudo-color versions of PS images, which serve as the reference criterion [8]. The agreement between objective measures and subjective perception varies across spatial scales and resolution configurations of the same scene. For severely distorted images, most IQA methods differentiate quality levels effectively. However, when performance differences are marginal, objective assessment becomes challenging. In the analyzed examples, subjective DMOS-GM values identify PCA, BDSD, IHS, MTF-GLP-CBD, HPF, ATWT-M2, and the EXP baseline as the preferred outputs for specific scale configurations. Among the evaluated metrics, the Q-index and MLPSIE exhibit the highest consistency with subjective scores, matching the human-preferred solution in five out of seven cases. ERGAS aligns with the subjective choice in four cases, whereas $Q2^n$ and Q_{oa} (the opinion-aware quality index) are consistent in only three. For instance, in the first row, both Q and MLPSIE correctly identify PCA as the preferred output. Conversely, in the second row, the subjective ranking prefers BDSD, whereas Q, $Q2^n$, MLPSIE, and Q_{oa} indicate alternative solutions.



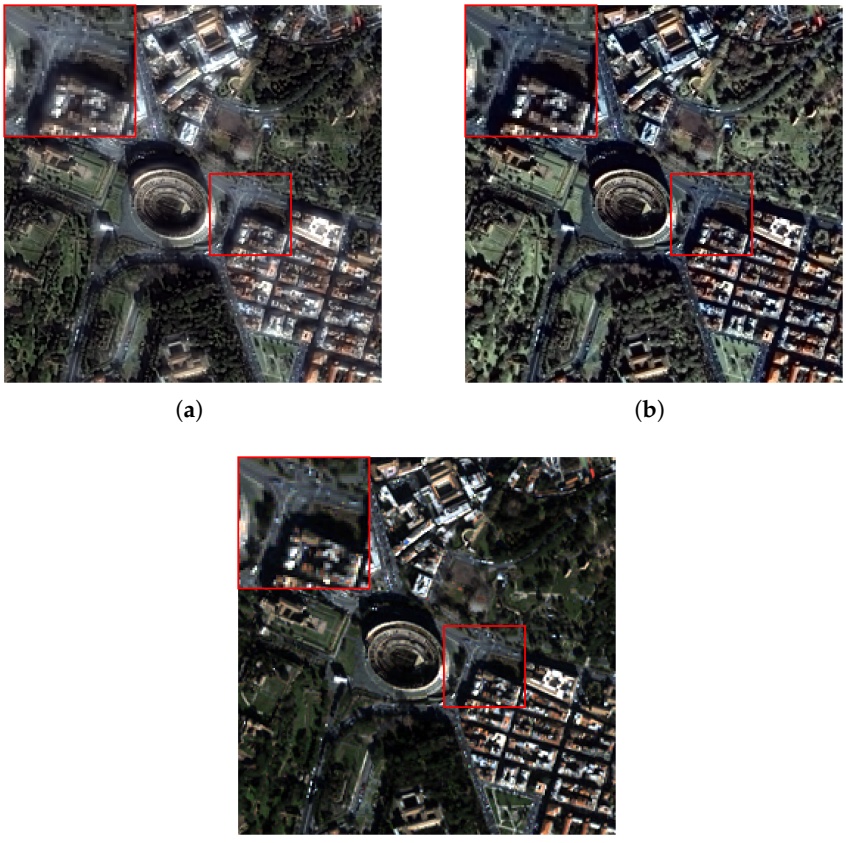
Figure 4. Comparison of pan-sharpening method rankings obtained using objective quality metrics and subjective DMOS-GM evaluations across different spatial and spectral resolution configurations of a specific scene. In each subfigure, the method highlighted in red denotes the top-performing algorithm according to the respective objective metric. The subjective DMOS-GM scores serve as the perceptual reference baseline against which the objective rankings are evaluated.

The subjective DMOS-GM scores for the top-evaluated techniques in these configurations exhibit minimal statistical variance. This marginal variance explains the high rank volatility observed in objective metrics when differentiating among the best-performing algorithms. Conversely, in the last three rows of Figure 4, the rankings generated by

objective metrics and subjective evaluations show strong agreement in identifying the poorest-performing methods. This alignment occurs because human experts and standard structural metrics both severely penalize spatial blur at degraded scales. Consequently, significant structural degradation consistently triggers both high mathematical error and unanimous perceptual rejection.

No single metric consistently reproduces subjective rankings across all spatial configurations. Nevertheless, MLPSIE maintains agreement with subjective values in challenging scenarios, such as correctly identifying the preference for MTF-GLP-CBD where conventional metrics select alternative solutions. These observations highlight the limitations of traditional full-reference quality measures, demonstrating that high objective scores do not guarantee alignment with expert perceptual evaluations. This discrepancy, defined as the perceptual gap, requires a shift in assessment focus from mathematical error minimization toward perceptually oriented protocols that capture human visual preferences and task-specific utility.

The analysis reveals a discrepancy between classical reference-based metrics and perceptual evaluation (Figure 5). The HPF method yields lower SAM (8.7952) and ERGAS (13.8718) values than the PCA method (10.7881 and 15.4725, respectively), indicating reduced spectral distortion. Conversely, PCA achieves a lower DMOS score (35.5742) compared to HPF (47.2789). Because lower DMOS values correspond to higher perceived quality, human observers preferred the PCA output over HPF, despite PCA's inferior performance across SAM and ERGAS metrics. This instance confirms that mathematical error minimization does not consistently align with human visual preference.



(**a**) (**b**)

(**c**)

Figure 5. Comparison of objective metrics (SAM, ERGAS) and subjective evaluation (DMOS) for the HPF and PCA methods, illustrating the discrepancy between mathematical error and perceptual preference. (**a**) HPF (SAM 8.7952, ERGAS 13.8718, DMOS 47.2789). (**b**) PCA (SAM 10.7881, ERGAS 15.4725, DMOS 35.5742). (**c**) Reference.

To further quantify the perceptual gap and evaluate the limitations of objective quality assessment, a comparison against subjective benchmarks is conducted. The evaluation utilizes a benchmark database containing subjective evaluations (DMOS) [8]. While metrics such as MLPSIE, ERGAS, and Q_{oa} were explicitly designed for fused multi-band imagery, the remaining general-purpose and deep learning architectures (including BRISQUE, TRES, MANIQA, ARNIQA, AGAIQA, and WAQNIQA) were originally developed for standard spatial images. To accommodate the remote sensing data, these latter methods were adapted by computing the metric independently for each spectral channel and averaging the resulting scores. Table 4 presents a comparative evaluation using SRCC, KRCC, PLCC, and RMSE to measure the alignment between objective metrics and human visual perception.

Table 4. Performance evaluation of selected quality assessment methods based on SRCC, KRCC, PLCC, and RMSE metrics.

Method	SRCC	KRCC	PLCC	RMSE
BRISQUE^{2012} [100]	0.9007	0.7473	0.9655	1.9523
Q_{oa}^{2019} [8]	0.8971	0.7307	0.9381	2.7947
ERGAS^{2000} [52]	0.9265	0.7796	0.9658	1.9560
MLPSIE^{2022} [112]	0.9548	0.8295	0.9803	1.5797
TRES^{2022} [122]	0.9308	0.7965	0.9750	1.7363
MANIQA^{2022} [123]	0.9354	0.7962	0.9696	2.0628
ARNIQA^{2024} [124]	0.6630	0.4819	0.7488	5.1797
AGAIQA^{2024} [125]	0.9237	0.7792	0.9717	1.8111
WAQNIQA^{2026} [126]	0.9442	0.8195	0.9810	0.0398

The data demonstrates variability in prediction performance across different assessment paradigms. The highest correlation with subjective scores is achieved by MLPSIE (SRCC = 0.9548, PLCC = 0.9803) and WAQNIQA (SRCC = 0.9442, PLCC = 0.9810). The WAQNIQA method registers the lowest prediction error (RMSE = 0.0398). These results indicate that specific learning architectures correlate closely with human perceptual evaluations. However, the data also highlights the generalization limitations of certain deep learning approaches. The ARNIQA model yields the lowest correlation metrics in the comparison set (SRCC = 0.6630, PLCC = 0.7488) alongside the highest error (RMSE = 5.1797). This discrepancy provides empirical evidence that learning-based models exhibit reduced reliability when processing data distributions that deviate from their training domains. Traditional and statistical metrics exhibit varying degrees of consistency. The classical ERGAS metric yields a rank correlation of SRCC = 0.9265, exceeding the values obtained by the general-purpose statistical BRISQUE evaluator (SRCC = 0.9007) and the opinion-aware Q_{oa} index (SRCC = 0.8971). The observed performance spread, with SRCC values ranging from 0.6630 to 0.9548, provides validation for the perceptual gap hypothesis, as mathematical error minimization does not consistently translate to perceptual alignment.

7.7. Summary of IQA Limitations

The critical analysis of existing IQA frameworks reveals that no single metric provides a universally reliable assessment of pan-sharpened imagery. Full-reference metrics are fundamentally constrained by scale-reduction protocols and exhibit rigid biases toward either spatial structure or spectral fidelity. Consistency-based no-reference indices mitigate the need for ground truth but remain highly sensitive to arbitrary parameterization and invalid scale-invariance assumptions, leading to rank volatility. Conversely, deep learning and statistical approaches demonstrate superior alignment with subjective human perception. However, their operational usability is restricted by cross-sensor overfitting, domain

shift vulnerabilities, and a lack of interpretability. The empirical validation confirms a perceptual gap, in which mathematical error minimization frequently contradicts both human visual preference and downstream task utility. Addressing these algorithmic vulnerabilities requires shifting the evaluation paradigm from isolated, pixel-wise distortion calculations toward perceptually robust, task-driven frameworks optimized for specific sensor environments.

8. Future Directions

Future research directions in pan-sharpening image quality assessment address structural, operational, and generalization limitations. Key objectives include improving prediction accuracy for heterogeneous scenes through multiscale, context-aware frameworks and generating high-resolution spatial quality maps using adaptive windows or object-based segments. Benchmark databases require expansion to incorporate realistic operational conditions, including spatial misregistration and non-linear spectral mixing. In full-resolution assessment, consistency-based metrics must relax strict scale-invariance assumptions to accommodate complex physical environments. Furthermore, advancing Human Visual System modeling requires capturing non-linear perceptual responses and automating evaluation procedures.

To mitigate the generalization bottleneck of learning-based models, research should investigate alternative label acquisition strategies, such as crowdsourcing, synthetic data generation, domain adaptation, and unsupervised transfer learning. Incorporating uncertainty quantification and explainable AI (XAI) techniques is necessary to report prediction confidence and ensure operational trust. Real-time data delivery demands lightweight IQA architectures optimized for edge computing and on-board satellite processing under strict hardware constraints. Finally, task-driven evaluation should link IQA metrics directly to downstream performance in automated land cover classification, target detection, and change detection. These generalized principles can extend to other multisensor fusion domains, including hyperspectral-multispectral and SAR-optical integration.

In summary, future developments should shift from pixel-wise error calculations toward comprehensive evaluation frameworks prioritizing consistency-based generalization, on-board execution, spatial mapping, and task-driven benchmarking.

9. Conclusions

In this paper, a systematic review of automatic FR and NR quality assessment approaches for PS images is presented. The main contribution is the establishment of a comprehensive PS IQA taxonomy detailing structural dependencies, parameter sensitivities, and failure conditions. Furthermore, the analysis quantifies the perceptual gap and validates the cross-sensor generalization limits of emerging DL metrics. Despite recent progress, PS IQA still struggles to balance spectral fidelity, perceptual quality, and practical utility. Many existing methods simply adapt general IQA techniques using arbitrary parameters, such as window sizes or weighting coefficients, which prevents cross-study standardization. Moreover, the lack of full-resolution ground truth remains the primary obstacle, necessitating the use of consistency-based and NR approaches.

While DL models effectively capture complex spatial-spectral dependencies, they introduce new methodological challenges. High accuracy on specific benchmarks often comes at the cost of overfitting, poor cross-sensor generalization, and a lack of interpretability. To ensure reliability in diverse remote sensing applications, future research must incorporate domain adaptation, uncertainty quantification, and XAI.

Additionally, reducing image quality to a single global score is insufficient for modern operational demands. PS IQA should shift towards task-driven assessment, evaluating

fusion methods based on their actual impact on downstream applications like land cover classification, object detection, or change detection. However, implementing task-driven assessment is difficult because results depend heavily on label quality, classifier architecture, and scene characteristics, complicating objective comparisons. Since no single metric works universally across all sensors and scenes, future frameworks must adopt hybrid strategies. Benchmark datasets need to reflect real operational satellite conditions, including misregistration, nonlinear spectral mixing, and partial spectral overlaps. Finally, developing high-resolution spatial quality maps and lightweight models for on-board processing will be crucial to bridge the gap between theoretical IQA research and practical Earth observation.

Author Contributions: Conceptualization, I.S. and M.O.; methodology, I.S. and M.O.; investigation, I.S. and M.O.; writing and editing, I.S. and M.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study are openly available. The data analyzed in Section 7 can be accessed from the GitHub repository "Pansharpening-IQA" at https://github.com/oscaragudelom/Pansharpening-IQA (accessed on 11 June 2026).

Acknowledgments: During the preparation of this manuscript/study, the author(s) used Gemini Pro 3.1 for the purposes of proofreading and text editing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

References

1. Goetz, A.F. Three Decades of Hyperspectral Remote Sensing of the Earth: A Personal View. *Remote Sens. Environ.* **2009**, *113*, S5–S16. [CrossRef]
2. Yilmaz, C.S.; Yilmaz, V.; Gungor, O. A Theoretical and Practical Survey of Image Fusion Methods for Multispectral Pansharpening. *Inf. Fusion* **2022**, *79*, 1–43. [CrossRef]
3. Lillesand, T.; Kiefer, R.W.; Chipman, J. *Remote Sensing and Image Interpretation*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
4. Kpalma, K.; Chikr El-Mezouar, M.; Taleb, N. Recent Trends in Satellite Image Pan-Sharpening Techniques. In Proceedings of the 1st International Conference on Electrical, Electronic and Computing Engineering, Vrniacka Banja, Serbia, 2–5 June 2014; HAL Id: hal-01075703.
5. Du, Q.; Younan, N.H.; King, R.; Shah, V.P. On the Performance Evaluation of Pan-Sharpening Techniques. *IEEE Geosci. Remote Sens. Lett.* **2007**, *4*, 518–522. [CrossRef]
6. Souza, C.; Firestone, L.; Silva, L.M.; Roberts, D. Mapping Forest Degradation in the Eastern Amazon from SPOT 4 through Spectral Mixture Models. *Remote Sens. Environ.* **2003**, *87*, 494–506. [CrossRef]
7. Mohammadzadeh, A.; Tavakoli, A.; Valadan Zoej, M.J. Road Extraction Based on Fuzzy Logic and Mathematical Morphology from Pan-Sharpened IKONOS Images. *Photogramm. Rec.* **2006**, *21*, 44–60. [CrossRef]
8. 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]
9. Barbato, M. Multimodal Artificial Intelligence Strategies for Remote Sensing Earth Observation. Ph.D. Thesis, Università degli Studi di Milano-Bicocca, Milano, Italy, 2024.
10. Ducay, R.; Messinger, D.W. Radiometric Assessment of Four Pan-Sharpening Algorithms as Applied to Hyperspectral Imagery. In *Proceedings of the Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imagery XXVI*; SPIE: Bellingham, WA, USA, 2020; Volume 11392, p. 1139203.
11. Guo, J.; Kong, Z.; Xu, Q. A Multi-Level Supervised Network for Pansharpening to Reduce Color Distortion. In *Proceedings of the 2023 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2023)*; IEEE: New York, NY, USA, 2023; pp. 6811–6814. [CrossRef]
12. Javan, F.D.; Samadzadegan, F.; Mehravar, S.; Toosi, A.; Khatami, R.; Stein, A. A Review of Image Fusion Techniques for Pan-Sharpening of High-Resolution Satellite Imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 101–117. [CrossRef]
13. Gao, X.; Lu, W.; Tao, D.; Li, X. Image Quality Assessment and Human Visual System. In *Proceedings of the Visual Communications and Image Processing 2010*; SPIE: Bellingham, WA, USA, 2010, Volume 7744, pp. 316–325.

14. Mantiuk, R.; Tomaszewska, A.; Mantiuk, R. Comparison of Four Subjective Methods for Image Quality Assessment. *Comput. Graph. Forum* **2012**, *31*, 2273–2282. [CrossRef]
15. He, L.; Gao, F.; Hou, W.; Hao, L. Objective Image Quality Assessment: A Survey. *Int. J. Comput. Math.* **2013**, *91*, 2374–2388. [CrossRef]
16. Ciotola, M.; Guarino, G.; Mazza, A.; Poggi, G.; Scarpa, G. A Comprehensive Benchmarking Framework for Sentinel-2 Sharpening: Methods, Dataset, and Evaluation Metrics. *Remote Sens.* **2025**, *17*, 1983. [CrossRef]
17. Aghapour Maleki, S.; Ghassemian, H.; Imani, M. Nonreference Object-Based Pansharpening Quality Assessment. *Egypt. J. Remote Sens. Space Sci.* **2024**, *27*, 227–241. [CrossRef]
18. Song, W.; Gao, M.; Chehri, A.; Zhai, W.; Li, Q.; Jeon, G. Dual-Branch and Triple-Attention Network for Pan-Sharpening. *Appl. Intell.* **2024**, *54*, 8041–8058. [CrossRef]
19. Gao, H.; Li, S.; Li, J.; Dian, R. Multispectral Image Pan-Sharpening Guided by Component Substitution Model. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5406413. [CrossRef]
20. Choi, J.; Park, H.; Seo, D. Pansharpening Using Guided Filtering to Improve the Spatial Clarity of VHR Satellite Imagery. *Remote Sens.* **2019**, *11*, 633. [CrossRef]
21. Menon, A.S.; Aravinth, J.; Veni, S. Pan-Sharpening of Multi-Spectral Remote Sensing Data Using Multi-Resolution Analysis. In *Proceedings of the Machine Intelligence Techniques for Data Analysis and Signal Processing: Proceedings of the 4th International Conference MISP 2022, Volume 1*; Lecture Notes in Electrical Engineering; Springer Nature: Singapore, 2023, Volume 997, pp. 697–705. [CrossRef]
22. Meng, X.; Shen, H.; Li, H.; Zhang, L.; Fu, R. Review of the Pansharpening Methods for Remote Sensing Images Based on the Idea of Meta-Analysis: Practical Discussion and Challenges. *Inf. Fusion* **2019**, *46*, 102–113. [CrossRef]
23. Xing, Y.; Wang, M.; Yang, S.; Jiao, L. Pan-Sharpening via Deep Metric Learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 165–183. [CrossRef]
24. Choudhary, G.; Sethi, D. From Conventional Approach to Machine Learning and Deep Learning Approach: An Experimental and Comprehensive Review of Image Fusion Techniques. *Arch. Comput. Methods Eng.* **2023**, *30*, 1267–1304. [CrossRef]
25. Kaur, H.; Koundal, D.; Kadyan, V. Image Fusion Techniques: A Survey. *Arch. Comput. Methods Eng.* **2021**, *28*, 4425–4447. [CrossRef]
26. Tu, T.M.; Su, S.C.; Shyu, H.C.; Huang, P.S. A New Look at IHS-Like Image Fusion Methods. *Inf. Fusion* **2001**, *2*, 177–186. [CrossRef]
27. Vivone, G.; Alparone, L.; Chanussot, J.; Dalla Mura, M.; Garzelli, A.; Licciardi, G.A.; Restaino, R.; Wald, L. A Critical Comparison Among Pansharpening Algorithms. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2565–2586. [CrossRef]
28. Kwarteng, P.; Chavez, A. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens.* **1989**, *55*, 339–348.
29. Tu, T.M.; Huang, P.S.; Hung, C.L.; Chang, C.P. A Fast Intensity-Hue-Saturation Fusion Technique with Spectral Adjustment for IKONOS Imagery. *IEEE Geosci. Remote Sens. Lett.* **2004**, *1*, 309–312. [CrossRef]
30. Laben, C.A.; Brower, B.V. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening. US Patent 6,011,875, 4 January 2000.
31. Vivone, G. Robust Band-Dependent Spatial-Detail Approaches for Panchromatic Sharpening. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6421–6433. [CrossRef]
32. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A.; Selva, M. An MTF-Based Spectral Distortion Minimizing Model for Pan-Sharpening of Very High Resolution Multispectral Images of Urban Areas. In *Proceedings of the 2003 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*; IEEE: New York, NY, USA, 2003; pp. 90–94. [CrossRef]
33. Vivone, G.; Marano, S.; Chanussot, J. Pansharpening: Context-Based Generalized Laplacian Pyramids by Robust Regression. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6152–6167. [CrossRef]
34. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A.; Selva, M. MTF-Tailored Multiscale Fusion of High-Resolution MS and Pan Imagery. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 591–596. [CrossRef]
35. Chen, C.; Li, Y.; Liu, W.; Huang, J. SIRF: Simultaneous Satellite Image Registration and Fusion in a Unified Framework. *IEEE Trans. Image Process.* **2015**, *24*, 4213–4224. [CrossRef]
36. Zhu, X.X.; Bamler, R. A Sparse Image Fusion Algorithm With Application to Pan-Sharpening. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2827–2836. [CrossRef]
37. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O. A New Pansharpening Algorithm Based on Total Variation. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 318–322. [CrossRef]
38. Serifoglu Yilmaz, C.; Yilmaz, V.; Güngör, O. On the Use of the SOS Metaheuristic Algorithm in Hybrid Image Fusion Methods to Achieve Optimum Spectral Fidelity. *Int. J. Remote Sens.* **2020**, *41*, 3993–4021. [CrossRef]
39. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]

40. Yuan, Q.; Wei, Y.; Meng, X.; Shen, H.; Zhang, L. A Multiscale and Multidepth Convolutional Neural Network for Remote Sensing Imagery Pan-Sharpening. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 978–989. [CrossRef]
41. Wei, Y.; Yuan, Q.; Shen, H.; Zhang, L. Boosting the Accuracy of Multispectral Image Pansharpening by Learning a Deep Residual Network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1795–1799. [CrossRef]
42. Ma, J.; Yu, W.; Chen, C.; Liang, P.; Guo, X.; Jiang, J. Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. *Inf. Fusion* **2020**, *62*, 110–120. [CrossRef]
43. Zhou, H.; Liu, Q.; Wang, Y. PanFormer: A Transformer Based Model for Pan-sharpening. In *Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME)*; IEEE: New Yok, NY, USA, 2022; pp. 1–6. [CrossRef]
44. Meng, Q.; Shi, W.; Li, S.; Zhang, L. PanDiff: A Novel Pansharpening Method Based on Denoising Diffusion Probabilistic Model. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5611317. [CrossRef]
45. Moore, G. *USAF Summer Research Program–1995 Graduate Student Research Program, Final Reports, Phillips Laboratory*; Technical report; Phillips Laboratory: Albuquerque, NM, USA, 1995; Volume 8.
46. Jasim, B.S.; Al-Bayati, Z.M.K.; Obaid, M.K. Accuracy of Horizontal Coordinates of Cadastral Maps after Geographic Regression and Their Modernization Using GIS Techniques. *Int. J. Civ. Eng. Technol. (IJCIET)* **2018**, *9*, 1395–1403.
47. Kadhum, Z.; Jasim, B.; Al-Saedi, A.S.J. Improving the Spectral and Spatial Resolution of Satellite Image Using Geomatics Techniques. *AIP Conf. Proc.* **2023**, *2776*, 040011. [CrossRef]
48. Ciotola, M.; Guarino, G.; Vivone, G.; Poggi, G.; Chanussot, J.; Plaza, A.; Scarpa, G. Hyperspectral Pansharpening: Critical Review, Tools, and Future Perspectives. *IEEE Geosci. Remote Sens. Mag.* **2024**, *13*, 311–338. [CrossRef]
49. Schott, J.R. *Remote Sensing: The Image Chain Approach*; Oxford University Press: New York, NY, USA, 2007.
50. Palubinskas, G. On the Validation of Pansharpening Methods. *arXiv* **2021**, arXiv:2111.07625. [CrossRef]
51. Ghassemian, H. Multi-Sensor Image Fusion Using Multirate Filter Banks. In *Proceedings of the Proceedings 2001 International Conference on Image Processing*; IEEE: New Yok, NY, USA, 2001; Volume 1, pp. 846–849. [CrossRef]
52. Wald, L. Quality of High Resolution Synthesised Images: Is There a Simple Criterion? In *Proceedings of the Third Conference "Fusion of Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images"*; SEE/URISCA: Nice, France, 2000; pp. 99–103.
53. Jiao, J.; Wu, L. Image Restoration for the MRA-Based Pansharpening Method. *IEEE Access* **2020**, *8*, 13694–13709. [CrossRef]
54. Zhang, J.; Su, R.; Fu, Q.; Ren, W.; Heide, F.; Nie, Y. A Survey on Computational Spectral Reconstruction Methods from RGB to Hyperspectral Imaging. *Sci. Rep.* **2022**, *12*, 11905. [CrossRef]
55. Ranchin, T.; Wald, L. Fusion of High Spatial and Spectral Resolution Images: The ARSIS Concept and Its Implementation. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 49–61.
56. Alparone, L.; Wald, L.; Chanussot, J.; Thomas, C.; Gamba, P.; Bruce, L.M. Comparison of Pansharpening Algorithms: Outcome of the 2006 GRS-S Data-Fusion Contest. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3012–3021. [CrossRef]
57. Wald, L.; Ranchin, T.; Mangolini, M. Fusion of Satellite Images of Different Spatial Resolutions: Assessing the Quality of Resulting Images. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 691–699; HAL Id: hal-00365304.
58. Deng, L.J.; Vivone, G.; Jin, C.; Chanussot, J. Detail Injection-Based Deep Convolutional Neural Networks for Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 6995–7010. [CrossRef]
59. Ranchin, T.; Aiazzi, B.; Alparone, L.; Baronti, S.; Wald, L. Image Fusion—The ARSIS Concept and Some Successful Implementation Schemes. *ISPRS J. Photogramm. Remote Sens.* **2003**, *58*, 4–18. [CrossRef]
60. Hao, X.; Li, X.; Wu, J.; Wei, B.; Song, Y.; Li, B. A No-Reference Quality Assessment Method for Hyperspectral Sharpened Images via Benford's Law. *Remote Sens.* **2024**, *16*, 1167. [CrossRef]
61. Thomas, C.; Ranchin, T.; Wald, L.; Chanussot, J. Synthesis of Multispectral Images to High Spatial Resolution: A Critical Review of Fusion Methods Based on Remote Sensing Physics. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1301–1312. [CrossRef]
62. Zhou, J.; Civco, D.L.; Silander, J.A. A Wavelet Transform Method to Merge Landsat TM and SPOT Panchromatic Data. *Int. J. Remote Sens.* **1998**, *19*, 743–757. [CrossRef]
63. Khan, M.M.; Alparone, L.; Chanussot, J. Pansharpening Quality Assessment Using the Modulation Transfer Functions of Instruments. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3880–3891. [CrossRef]
64. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A. Quality Assessment of Pansharpening Methods and Products. *IEEE Geosci. Remote Sens. Soc. Newsl.* **2011**, *161*, 10–18.
65. Ghassemian, H. A Review of Remote Sensing Image Fusion Methods. *Inf. Fusion* **2016**, *32*, 75–89. [CrossRef]
66. Pohl, C.; Moellmann, J.; Fries, K. Standardizing Quality Assessment of Fused Remotely Sensed Images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *XLII-2/W7*, 863–869. [CrossRef]
67. Bray, M.A.; Fraser, A.N.; Hasaka, T.P.; Carpenter, A.E. Workflow and Metrics for Image Quality Control in Large-Scale High-Content Screens. *J. Biomol. Screen.* **2012**, *17*, 266–274. [CrossRef]
68. Wang, Z.; Bovik, A.C. Modern Image Quality Assessment. In *Synthesis Lectures on Image, Video, and Multimedia Processing*; Springer: Cham, Switzerland, 2006; Volume 2, pp. 1–156. [CrossRef]

69. Grimson, W.E.L.; Grimson, W. *From Images to Surfaces: A Computational Study of the Human Early Visual System*; MIT Press: Cambridge, MA, USA, 1981; Volume 4.
70. Karimi, M.; Entezami, E. Quality Assessment for Retargeted Images: A Review. In *Proceedings of the 2020 International Conference on Machine Vision and Image Processing (MVIP)*; IEEE: New Yok, NY, USA, 2020; pp. 1–6. [CrossRef]
71. Hartling, S.; Sagan, V.; Sidike, P.; Maimaitijiang, M.; Carron, J. Urban Tree Species Classification Using a WorldView-2/3 and LiDAR Data Fusion Approach and Deep Learning. *Sensors* **2019**, *19*, 1284. [CrossRef]
72. Ustin, S.L.; Middleton, E.M. Current and Near-Term Advances in Earth Observation for Ecological Applications. *Ecol. Processes* **2021**, *10*, 1. [CrossRef]
73. Tong, X.-Y.; Lu, Q.; Xia, G.-S.; Zhang, L. Large-Scale Land Cover Classification in Gaofen-2 Satellite Imagery. In *Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium*; IEEE: New Yok, NY, USA, 2018; pp. 3599–3602. [CrossRef]
74. Campana, S. Archaeological Site Detection and Mapping: Some Thoughts on Differing Scales of Detail and Archaeological 'Non-Visibility'. In *Seeing the Unseen: Geophysics and Landscape Archaeology*; Campana, S.; Piro, S., Eds.; CRC Press/Taylor & Francis: Boca Raton, FL, USA, 2008; pp. 31–52.
75. Peled, A.; Gilichinsky, M. GIS-Driven Classification of Land Use Using IKONOS Data and a Core National Spatial Information Database. *Appl. Geomat.* **2013**, *5*, 109–117. [CrossRef]
76. Alcaras, E.; Della Corte, V.; Ferraioli, G.; Martellato, E.; Palumbo, P.; Parente, C.; Rotundi, A. Comparison of Different Pan-Sharpening Methods Applied to IKONOS Imagery. *Geogr. Tech.* **2021**, *16*, 198–210. [CrossRef]
77. Zhang, Y. Methods for Image Fusion Quality Assessment—A Review, Comparison and Analysis. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, *37*, 1101–1109.
78. Alparone, L.; Baronti, S.; Garzelli, A.; Nencini, F. A Global Quality Measurement of Pan-Sharpened Multispectral Imagery. *IEEE Geosci. Remote Sens. Lett.* **2004**, *1*, 313–317. [CrossRef]
79. Basaeed, E.; Bhaskar, H.; Al-Mualla, M. Comparative Analysis of Pan-Sharpening Techniques on DubaiSat-1 Images. In *Proceedings of the 16th International Conference on Information Fusion (FUSION 2013), Istanbul, Turkey*; IEEE: New Yok, NY, USA, 2013; pp. 227–234.
80. Wang, Z.; Bovik, A.C. A Universal Image Quality Index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]
81. Guan, X.; Li, F.; Zhang, X.; Ma, M.; Mei, S. Assessing Full-Resolution Pansharpening Quality: A Comparative Study of Methods and Measurements. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 6860–6875. [CrossRef]
82. Garzelli, A.; Nencini, F. Hypercomplex Quality Assessment of Multi/Hyperspectral Images. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 662–665. [CrossRef]
83. Javan, F.D.; Samadzadegan, F.; Reinartz, P. Spatial Quality Assessment of Pan-Sharpened High Resolution Satellite Imagery Based on an Automatically Estimated Edge-Based Metric. *Remote Sens.* **2013**, *5*, 6539–6559. [CrossRef]
84. Li, H.; Nie, R.; Cao, J.; Jin, B.; Han, Y. MPEFNet: Multilevel Progressive Enhancement Fusion Network for Pansharpening. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 9358–9368. [CrossRef]
85. Boreman, G.D. *Modulation Transfer Function in Optical and Electro-Optical Systems*; Tutorial Texts in Optical Engineering; SPIE Press: Bellingham, WA, USA, 2001; Volume TT52, p. 120. [CrossRef]
86. Yuhas, R.H.; Goetz, A.F.H.; Boardman, J.W. Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm. In *Summaries of the Third Annual JPL Airborne Geoscience Workshop* ; JPL: Pasadena, CA, USA, 1992; pp. 147–149.
87. 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]
88. Li, C.; Bovik, A.C. Three-Component Weighted Structural Similarity Index. In *Proceedings of the Image Quality and System Performance VI*; Proceedings of SPIE; SPIE: Bellingham, WA, USA, 2009; Volume 7242, pp. 252–260. [CrossRef]
89. Okarma, K. Current Trends and Advances in Image Quality Assessment. *Elektron. Elektrotechnika* **2019**, *25*, 77–84. [CrossRef]
90. Bakurov, I.; Buzzelli, M.; Schettini, R.; Castelli, M.; Vanneschi, L. Structural Similarity Index (SSIM) Revisited: A Data-Driven Approach. *Expert Syst. Appl.* **2022**, *189*, 116087. [CrossRef]
91. Palubinskas, G. Joint Quality Measure for Evaluation of Pansharpening Accuracy. *Remote Sens.* **2015**, *7*, 9292–9310. [CrossRef]
92. Makarau, A.; Richter, R.; Müller, R.; Reinartz, P. Haze Detection and Removal in Remotely Sensed Multispectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5895–5905. [CrossRef]
93. Palubinskas, G. Quality Assessment of Pan-Sharpening Methods. In *Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, QC, Canada*; IEEE: New Yok, NY, USA, 2014; pp. 2526–2529. [CrossRef]
94. Palubinskas, G. Fast, Simple, and Good Pan-Sharpening Method. *J. Appl. Remote Sens.* **2013**, *7*, 073526. [CrossRef]
95. Alparone, L.; Aiazzi, B.; Baronti, S.; Garzelli, A.; Nencini, F.; Selva, M. Multispectral and Panchromatic Data Fusion Assessment Without Reference. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 193–200. [CrossRef]

96. Kwan, C.; Budavari, B.; Bovik, A.C.; Marchisio, G. Blind Quality Assessment of Fused WorldView-3 Images by Using the Combinations of Pansharpening and Hypersharpening Paradigms. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1835–1839. [CrossRef]
97. Arienzo, A.; Vivone, G.; Garzelli, A.; Alparone, L.; Chanussot, J. Full-Resolution Quality Assessment of Pansharpening: Theoretical and Hands-On Approaches. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 168–201. [CrossRef]
98. Xiong, Y.; Shao, F.; Meng, X.; Jiang, Q.; Sun, W.; Fu, R.; Ho, Y.S. A Large-Scale Remote Sensing Database for Subjective and Objective Quality Assessment of Pansharpened Images. *J. Vis. Commun. Image Represent.* **2020**, *73*, 102947. [CrossRef]
99. Zhang, X.; Wang, J. A Deep Learning-Based Blind Full Resolution Assessment Method for Pansharpened Images. In *Proceedings of the 2025 6th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*; IEEE: New Yok, NY, USA, 2025; pp. 603–611. [CrossRef]
100. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [CrossRef]
101. Xu, S.; Jiang, S.; Min, W. No-Reference/Blind Image Quality Assessment: A Survey. *IETE Tech. Rev.* **2017**, *34*, 223–245. [CrossRef]
102. Athar, S.; Wang, Z. A Comprehensive Performance Evaluation of Image Quality Assessment Algorithms. *IEEE Access* **2019**, *7*, 140030–140070. [CrossRef]
103. Bosse, S.; Maniry, D.; Müller, K.R.; Wiegand, T.; Samek, W. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Trans. Image Process.* **2018**, *27*, 206–219. [CrossRef]
104. Zhou, B.; Shao, F.; Meng, X.; Fu, R.; Ho, Y.S. No-Reference Quality Assessment for Pansharpened Images via Opinion-Unaware Learning. *IEEE Access* **2019**, *7*, 40388–40401. [CrossRef]
105. Zhang, L.; Zhang, L.; Bovik, A.C. A Feature-Enriched Completely Blind Image Quality Evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. [CrossRef]
106. Ghadiyaram, D.; Bovik, A.C. Perceptual Quality Prediction on Authentically Distorted Images Using a Bag of Features Approach. *J. Vis.* **2017**, *17*, 32. [CrossRef]
107. Badal, N.; Soundararajan, R.; Garg, A.; Patil, A. No Reference Pansharpened Image Quality Assessment Through Deep Feature Similarity. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 7235–7247. [CrossRef]
108. Ni, Z.; Liu, Y.; Ding, K.; Yang, W.; Wang, H.; Wang, S. Opinion-Unaware Blind Image Quality Assessment Using Multi-Scale Deep Feature Statistics. *IEEE Trans. Multimed.* **2024**, *26*, 9133–9146. [CrossRef]
109. Bao, K.; Meng, X.; Chai, X.; Shao, F. A Blind Full Resolution Assessment Method for Pansharpened Images Based on Multistream Collaborative Learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5410311. [CrossRef]
110. Napoletano, P.; Piccoli, F.; Schettini, R. Anomaly Detection in Nanofibrous Materials by CNN-Based Self-Similarity. *Sensors* **2018**, *18*, 209. [CrossRef]
111. Pires de Lima, R.; Marfurt, K. Convolutional Neural Network for Remote-Sensing Scene Classification: Transfer Learning Analysis. *Remote Sens.* **2019**, *12*, 86. [CrossRef]
112. Stępień, I.; Oszust, M. No-Reference Quality Assessment of Pan-Sharpening Images with Multi-Level Deep Image Representations. *Remote Sens.* **2022**, *14*, 1119. [CrossRef]
113. Wang, W.; Zhang, M.; Wang, D.; Jiang, Y. Kernel PCA Feature Extraction and the SVM Classification Algorithm for Multiple-Status, Through-Wall, Human Being Detection. *EURASIP J. Wirel. Commun. Netw.* **2017**, *2017*, 151. [CrossRef]
114. Stępień, I.; Oszust, M. Three-Branch Neural Network for No-Reference Quality Assessment of Pan-Sharpened Images. *Eng. Appl. Artif. Intell.* **2025**, *139*, 109594. [CrossRef]
115. Rodríguez-Esparragón, D.; Marcello, J.; Eugenio, F.; García-Pedrero, A.; Gonzalo-Martín, C. Object-Based Quality Evaluation Procedure for Fused Remote Sensing Imagery. *Neurocomputing* **2017**, *255*, 40–51. [CrossRef]
116. DadrasJavan, F.; Samadzadegan, F. An Object-Level Strategy for Pan-Sharpening Quality Assessment of High-Resolution Satellite Imagery. *Adv. Space Res.* **2014**, *54*, 2286–2295. [CrossRef]
117. Aghapour Maleki, S.; Ghassemian, H. A Pansharpened Image Quality Assessment Using Segmentation Procedure. *Int. J. Remote Sens.* **2021**, *42*, 4157–4176. [CrossRef]
118. Scarpa, G.; Vitale, S.; Cozzolino, D. Target-Adaptive CNN-Based Pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5443–5457. [CrossRef]
119. Aiazzi, B.; Baronti, S.; Selva, M. Improving Component Substitution Pansharpening Through Multivariate Regression of MS + Pan Data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3230–3239. [CrossRef]
120. Li, S.; Yang, Z.; Li, H. Statistical Evaluation of No-Reference Image Quality Assessment Metrics for Remote Sensing Images. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 133. [CrossRef]
121. Wang, H.; Zhang, H.; Tian, X.; Ma, J. Zero-Sharpen: A universal pansharpening method across satellites for reducing scale-variance gap via zero-shot variation. *Inf. Fusion* **2024**, *101*, 102003. [CrossRef]
122. Golestaneh, S.A.; Dadsetan, S.; Kitani, K.M. No-Reference Image Quality Assessment via Transformers, Relative Ranking, and Self-Consistency. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; IEEE: New Yok, NY, USA, 2022; pp. 1220–1230. [CrossRef]

123. Yang, S.; Wu, T.; Shi, S.; Lao, S.; Gong, Y.; Cao, M.; Wang, J.; Yang, Y. MANIQA: Multi-dimension Attention Network for No-Reference Image Quality Assessment . In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Los Alamitos, CA, USA, 19–20 June 2022*; IEEE: New Yok, NY, USA, 2022; pp. 1190–1199. [CrossRef]
124. Agnolucci, L.; Galteri, L.; Bertini, M.; Del Bimbo, A. ARNIQA: Learning Distortion Manifold for Image Quality Assessment. In *Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; IEEE: New Yok, NY, USA, 2024; pp. 188–197. [CrossRef]
125. Wang, H.; Liu, J.; Tan, H.; Lou, J.; Liu, X.; Zhou, W.; Liu, H. Blind Image Quality Assessment via Adaptive Graph Attention. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 10299–10309. [CrossRef]
126. Huo, Y.; Huang, G.; Yu, Z.; Yuan, X.; Pun, C.M.; Cheng, L.; Shi, H.; Chen, X.; Chen, Z. WAQNIQA: Wavelet-Augmented Quaternion Network for No-Reference Image Quality Assessment. *IEEE Trans. Syst. Man Cybern. Syst.* **2026**, *56*, 4521–4534. [CrossRef]

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