

A Comprehensive Survey of Robust Image Quality Metrics for Satellite Imagery

Pooja Pandya^{1,*}, Bhargav Rajyagor²

Abstract

Satellite imagery is essential for applications like environmental monitoring, urban development, precision agriculture, defence surveillance, and disaster response. The reliability of these applications is closely tied to the quality of the captured images, which may be compromised by atmospheric effects, sensor imperfections, compression artifacts, and transmission noise. As a result, accurate image quality assessment (IQA) is essential to ensure trustworthy analysis and informed decision-making in satellite-based systems. The distinctive properties of remote sensing data—including high spatial resolution, multispectral and hyperspectral characteristics, and application-specific requirements—introduce additional complexity to conventional IQA methods.

This paper delivers a detailed review and analytical evaluation of IQA techniques tailored for satellite and remote sensing imagery. Its primary aim is to systematically classify, compare, and assess existing quality assessment methods, highlighting their advantages, limitations, and practical relevance. The reviewed approaches are grouped into three main categories: full-reference (FR), reduced-reference (RR), and no-reference (NR) metrics. Within these groups, the study examines traditional statistical measures such as mean squared error (MSE) and peak signal-to-noise ratio (PSNR), structural similarity-based metrics, information-theoretic models, perceptual quality indicators, and recent deep learning-based approaches. The paper also addresses task-oriented evaluation frameworks and emerging methods designed for multispectral, hyperspectral, and super-resolution satellite imagery. A comparative analysis is performed based on criteria such as robustness to various distortions, computational efficiency, scalability, and suitability for operational satellite systems. The findings reveal that although classical FR metrics are computationally simple and widely used, they often show limited correlation with perceptual and application-driven quality in remote sensing contexts. Structural and perceptual metrics improve correlation but may not fully capture spectral complexity. Deep learning-based NR methods demonstrate strong adaptability and prediction accuracy in challenging distortion conditions; however, they demand substantial annotated data and computational resources. Task-driven evaluation strategies appear promising in aligning image quality assessment with mission-specific performance objectives.

In conclusion, no single IQA method can comprehensively address all the requirements of satellite imagery analysis. Future research should emphasize the development of adaptive, intelligent, and data-driven frameworks capable of handling multimodal satellite data and real-time operational constraints. The integration of domain expertise with advanced learning techniques, along with the creation of standardized benchmark datasets, will be crucial for advancing robust and practical satellite image quality assessment solutions

*Author for Correspondence

Pooja Pandya
E-mail: poojapandyamca@gmail.com

¹Assistant Professor, Faculty of Computer Application, Noble University, Junagadh, India

²Principal and Associate Professor, Faculty of Computer Application, Noble University, Junagadh, India

Received Date: March 19, 2026

Accepted Date: March 23, 2026

Published Date: April 24, 2026

Citation: Pooja Pandya, Bhargav Rajyagor. A Comprehensive Survey of Robust Image Quality Metrics for Satellite Imagery. Research & Reviews: Journal of Space Science & Technology. 2026; 15(1): 7–20p.

Keywords: Satellite imagery, Image Quality Assessment (IQA), Remote sensing, Robust metrics, Full-reference metrics, No-reference metrics, Deep learning, Multispectral imaging, Hyperspectral imaging.

INTRODUCTION

Satellite imagery has become a vital asset for both scientific exploration and commercial operations. Applications such as climate observation, disaster response, urban development, defense monitoring, and precision agriculture increasingly depend on high-resolution, high-quality satellite data. Contemporary Earth observation platforms continuously produce enormous volumes of multispectral, hyperspectral, and panchromatic imagery. Because these images directly support critical analyses and policy-making processes, ensuring their accuracy and consistency is essential. As sensor technologies and satellite systems advance rapidly, the demand for reliable and comprehensive image quality evaluation frameworks has grown correspondingly.

Despite these advancements, satellite imagery is frequently affected by multiple forms of degradation, including atmospheric disturbances, sensor noise, radiometric variations, compression artifacts, geometric distortions, cloud contamination, and transmission errors. Such issues can significantly hinder downstream applications like land-cover mapping, object detection, image segmentation, and change analysis. Studies have shown that environmental factors such as illumination variability and atmospheric conditions can distort spectral characteristics, thereby reducing model robustness [4]. In specialized domains, such as solar EUV remote sensing, domain-specific distortions may compromise scientific interpretation [2]. Furthermore, in super-resolution reconstruction, certain artifacts can inflate quantitative performance metrics while simultaneously diminishing perceptual image quality [1]. These limitations indicate that conventional evaluation methods may not adequately capture the practical effectiveness and usability of satellite imagery.

To overcome these limitations, the development of robust image quality assessment (IQA) metrics is essential for objectively evaluating satellite imaging systems and image enhancement techniques. Traditional full-reference measures, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), are commonly employed; however, they often inadequately represent perceptual and spectral distortions unique to remote sensing imagery. Consequently, reduced-reference and no-reference methods have gained increasing relevance, especially in real-world satellite operations where pristine ground-truth images are typically unavailable. Recent advances in visual quality assessment for aerial imagery highlight the growing significance of perceptual models and human-visual-system-inspired approaches for accurate quality prediction [5]. Furthermore, deep learning-based IQA frameworks have emerged as powerful alternatives, offering adaptive and data-driven evaluation strategies that exhibit stronger correlation with subjective quality perception [3].

Recent research efforts have addressed various dimensions of satellite image quality assessment. Hybrid deep learning architectures have been introduced for remote sensing image super-resolution, achieving improved spatial detail reconstruction while preserving spectral fidelity [1]. Robustness analyses of multispectral satellite imagery under environmental variability have underscored the value of spectral-sensitive metrics such as Spectral Angle Mapper (SAM) and ERGAS for more reliable evaluation [4]. In addition, specialized quality assessment frameworks have been developed for domain-specific systems, including solar EUV remote sensing applications [2]. Broader surveys in the IQA field further document the progression from conventional statistical measures toward advanced transformer-based and perceptual learning methodologies [3]. Nevertheless, despite these developments, a unified and comprehensive review focusing specifically on robustness in satellite image quality metrics—particularly under real-world environmental and sensor-induced distortions—remains insufficiently explored.

This review seeks to address the identified gap by delivering a systematic and in-depth examination of robust image quality assessment (IQA) metrics for satellite imagery. Building upon recent advances in perceptual modeling [5], deep learning-based evaluation frameworks [3], super-resolution quality analysis [1], spectral-aware robustness metrics [4], and domain-specific assessment strategies [2], this paper provides a consolidated perspective on current methodologies and emerging trends.

Specifically, the contributions of this review are as follows:

Systematic Classification: Satellite image quality assessment techniques are organized into full-reference, reduced-reference, and no-reference categories, highlighting their theoretical foundations and practical relevance.

Comparative Analysis: The robustness, computational complexity, and operational suitability of these metrics are compared across multispectral, hyperspectral, and high-resolution satellite imagery, with particular attention to spectral-aware measures such as SAM and ERGAS [4].

Review of Modern Approaches: Recent perceptual and deep learning-driven IQA models are examined, emphasizing their improved correlation with subjective quality perception and their adaptability to complex distortion scenarios [3], [5].

Identification of Research Gaps: Key challenges are discussed, including limited benchmark datasets, cross-sensor generalization issues, domain-specific distortions [2], and the mismatch between numerical performance and perceptual quality in enhancement tasks such as super-resolution [1].

Future Directions: The paper outlines promising research pathways toward adaptive, task-oriented, and real-time satellite image quality evaluation frameworks capable of operating reliably under real-world environmental and sensor-induced variations.

FUNDAMENTALS OF IMAGE QUALITY ASSESSMENT

Image Quality Assessment (IQA) seeks to objectively measure the visual fidelity of an image using computational models. In satellite and remote sensing applications, IQA is essential for evaluating imaging sensors, compression schemes, super-resolution techniques, and image enhancement algorithms. Depending on the availability of reference data, IQA approaches are generally classified into three categories: Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR) methods [3].

Categories of IQA Metrics

Full-Reference (FR) Metrics

Full-reference metrics require complete access to the original, undistorted image. These methods compare the reference and distorted images by calculating pixel-level or structural differences and are commonly used to benchmark reconstruction and super-resolution algorithms in remote sensing applications [1], [3].

- **Mean Squared Error (MSE)**

Mean Squared Error (MSE) computes the average squared difference between corresponding pixel values in the reference and degraded images. It is mathematically straightforward and computationally efficient. However, because it ignores structural content and perceptual characteristics, it often shows weak correlation with human visual perception [3]. Despite this limitation, MSE remains widely used in satellite image restoration studies due to its simplicity and ease of implementation.

- **Peak Signal-to-Noise Ratio (PSNR)**

PSNR is derived from MSE and expresses the ratio between the maximum possible pixel intensity and the distortion noise, typically measured in decibels (dB). It is extensively employed in evaluating super-resolution and compression methods for remote sensing imagery [1]. Although PSNR provides a convenient quantitative indicator, it does not reliably reflect perceptual quality or spectral distortions, particularly in multispectral satellite data [4].

- **Structural Similarity Index (SSIM)**

SSIM measures similarity by comparing luminance, contrast, and structural information rather than relying solely on pixel-wise errors. It demonstrates stronger alignment with human perception than MSE or PSNR [3]. In satellite image analysis, SSIM is frequently applied to assess spatial reconstruction quality. Nevertheless, it may not adequately represent spectral fidelity in multispectral and hyperspectral imagery [4].

Reduced-Reference (RR) Metrics

Reduced-reference metrics rely on partial information extracted from the original image, such as statistical descriptors or transform-domain features. Instead of transmitting the entire reference image, only selected quality-relevant parameters are shared for comparison. This approach is particularly advantageous in satellite communication systems with limited bandwidth. RR methods aim to achieve a balance between evaluation accuracy and transmission efficiency, making them suitable for remote sensing compression and data transmission scenarios [3]. However, designing robust RR metrics for multispectral satellite imagery is challenging due to the complexity and high dimensionality of spectral information [4].

No-Reference (NR) Metrics

No-reference metrics evaluate image quality without requiring access to the original image. These approaches are especially valuable in operational satellite systems where ground-truth references are unavailable or impractical to obtain. NR-IQA methods typically leverage natural scene statistics, distortion-specific characteristics, or learned perceptual representations [3].

- **BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator)**

BRISQUE is a distortion-agnostic NR metric that uses spatial-domain natural scene statistical features to predict image quality. It does not require prior identification of distortion types and has been widely applied in general-purpose image quality evaluation tasks [3]. However, its effectiveness in multispectral satellite imagery may be limited because remote sensing data often differ significantly from conventional natural image statistics.

- **NIQE (Natural Image Quality Evaluator)**

NIQE is an unsupervised NR metric that quantifies deviations from the statistical regularities of natural images. Unlike BRISQUE, it does not rely on training with human-rated images [3]. Although computationally efficient, NIQE may struggle to capture domain-specific degradations in satellite imagery, such as atmospheric scattering effects and spectral inconsistencies [4].

- **Deep Learning-Based NR-IQA**

Recent developments in deep learning have introduced data-driven NR-IQA models capable of automatically extracting perceptual quality features using convolutional neural networks (CNNs) and transformer-based architectures [3]. These models generally demonstrate stronger correlation with subjective human assessment and improved robustness to complex distortions. In remote sensing contexts, learning-based IQA frameworks are increasingly utilized to evaluate super-resolution results and environmental robustness [1], [4]. However, their effectiveness largely depends on the availability of high-quality annotated datasets, which remain scarce in the satellite imaging domain.

SATELLITE-SPECIFIC IMAGE QUALITY CHALLENGES

Satellite imagery introduces distinctive quality assessment challenges that differ markedly from conventional natural image processing. Unlike ground-based photography, satellite imaging systems operate under dynamic atmospheric conditions, variable illumination environments, and complex sensor configurations. These factors generate distortions that are often inadequately represented by traditional image quality metrics. The following subsections outline the primary satellite-specific challenges that complicate reliable image quality evaluation.

Atmospheric Scattering

Atmospheric scattering is a major source of degradation in satellite imagery. As electromagnetic waves propagate through the atmosphere, interactions with air molecules, aerosols, and suspended particles produce haze, contrast reduction, and spectral alterations. Rayleigh and Mie scattering predominantly affect shorter wavelengths, leading to color shifts and diminished clarity. These distortions disrupt spectral consistency and reduce classification reliability. Investigations into multispectral robustness under environmental variability indicate that spatial similarity metrics such as PSNR and SSIM do not sufficiently capture atmospheric-induced spectral distortions, underscoring the importance of spectral-sensitive quality measures [4]. Moreover, domain-oriented quality assessment in solar EUV imaging demonstrates how radiative and atmospheric interferences can influence scientific interpretation [2].

Cloud Cover

Cloud presence represents a significant obstacle in satellite imaging, frequently causing partial or complete occlusion of surface features. Dense cloud formations obstruct reflected radiation, while thin clouds introduce semi-transparent distortions that alter spectral characteristics. Such contamination complicates both manual interpretation and automated tasks including land-cover classification and object detection. Conventional full-reference metrics may yield high similarity scores even when critical semantic content is obscured. Robustness analyses under environmental disturbances suggest that cloud-related degradations require perceptually adaptive and task-aware evaluation strategies [4]. Additionally, in super-resolution applications, cloud-affected regions may generate artificial reconstruction artifacts that increase PSNR values without improving meaningful image content [1].

Multispectral and Hyperspectral Complexity

Satellite sensors commonly acquire imagery across multiple spectral bands spanning visible, infrared, and shortwave infrared regions, while hyperspectral systems capture hundreds of narrow spectral channels. Although this spectral richness enhances analytical potential, it introduces substantial complexity in quality evaluation. Traditional spatial metrics such as MSE and SSIM primarily measure intensity-based similarity and often overlook inter-band spectral relationships. Research on multispectral robustness highlights that spectral distortion indicators like Spectral Angle Mapper (SAM) and ERGAS more effectively characterize cross-band quality variations [4]. Nevertheless, developing unified metrics that simultaneously assess spatial structure and spectral fidelity remains a significant research challenge, as emphasized in contemporary IQA surveys [3].

Spatial Resolution Variability

Satellite imagery exhibits wide variations in spatial resolution depending on sensor design and orbital altitude. High-resolution images provide detailed spatial information but are more prone to noise amplification and enhancement artifacts. Conversely, lower-resolution imagery may experience blurring and information loss. Super-resolution algorithms aim to mitigate these limitations; however, traditional evaluation metrics such as PSNR may not accurately reflect perceptual realism or task-based performance improvements [1]. Resolution heterogeneity across datasets also complicates cross-sensor evaluation, reinforcing the need for adaptive IQA frameworks capable of accommodating multi-resolution satellite data [3].

Radiometric Distortion

Radiometric distortion arises from inconsistencies in pixel intensity values due to atmospheric absorption, sensor limitations, and illumination variability. Changes in solar angle, seasonal conditions, and surface reflectance properties produce fluctuations in brightness and contrast, complicating consistent quality assessment. Such radiometric shifts can adversely affect downstream analyses including change detection and vegetation monitoring. Studies examining environmental robustness in multispectral imagery indicate that purely spatial similarity metrics often fail to maintain reliability under radiometric variations [4]. Perceptually driven quality assessment approaches recommend

incorporating contrast-aware and human-visual-system-inspired modeling to better represent radiometric inconsistencies [5].

Sensor Calibration Errors

Calibration inaccuracies result from detector aging, electronic interference, spectral band misalignment, and platform instability. These issues may manifest as striping artifacts, geometric distortions, and band misregistration, degrading both spatial coherence and spectral alignment. Subtle calibration defects may go undetected by conventional full-reference metrics if pixel-wise deviations are minimal. Consequently, robust quality assessment frameworks must consider cross-sensor variability and band alignment consistency. Research in environmental robustness and modern IQA methodologies highlights the importance of calibration-aware and learning-based evaluation models to ensure dependable satellite system assessment [3], [4].

REVIEW OF EXISTING METHODS

This section reviews existing image quality assessment (IQA) approaches in satellite and remote sensing imagery, focusing on super-resolution evaluation, spectral-aware quality metrics, deep learning-based methods, and robustness analysis. The discussion synthesizes findings from recent literature to highlight methodological trends, strengths, and limitations [1]–[5].

Super-Resolution Based Quality Evaluation

Super-resolution (SR) has become a central research area in remote sensing, aiming to enhance spatial resolution while preserving structural and spectral fidelity. Most remote sensing SR studies evaluate reconstruction performance using full-reference metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [1], [3]. These metrics are computationally efficient and provide standardized benchmarks for algorithm comparison.

However, reliance on PSNR and SSIM presents limitations. Although higher PSNR values often indicate reduced pixel-wise error, they do not necessarily correspond to improved perceptual realism or task-level performance. In many cases, reconstruction artifacts can artificially inflate PSNR without enhancing meaningful spatial details [1]. Similarly, SSIM primarily measures structural similarity but may not fully account for spectral consistency in multispectral and hyperspectral datasets [4]. As a result, recent research emphasizes the need for perceptually aligned and spectral-aware evaluation strategies to complement traditional full-reference metrics.

Multispectral and Hyperspectral Quality Assessment

Satellite sensors frequently capture imagery across multiple spectral bands, necessitating evaluation methods that account for both spatial and spectral fidelity. Conventional spatial metrics alone are insufficient for representing inter-band spectral distortions. To address this limitation, spectral-aware quality measures such as Spectral Angle Mapper (SAM) and ERGAS have been widely adopted in multispectral robustness analysis [4].

SAM evaluates the angular similarity between spectral vectors, making it effective for detecting spectral distortion independent of illumination magnitude. ERGAS provides a global relative error measure that balances spatial and spectral discrepancies. These metrics better capture cross-band fidelity compared to purely spatial measures like PSNR and SSIM. Nevertheless, robustness challenges remain, particularly under environmental variability, atmospheric scattering, and sensor noise [4]. Designing unified frameworks that simultaneously evaluate spatial structure and spectral integrity continues to be an open research problem, as highlighted in broader IQA surveys [3].

Deep Learning-Based Quality Metrics

The evolution of IQA has shifted from traditional statistical measures toward data-driven and perceptually motivated models. Early approaches relied on handcrafted features and natural scene

statistics to estimate distortion levels [3]. While effective for common distortions, these methods often struggle with the complex degradation patterns observed in satellite imagery.

Recent advancements incorporate convolutional neural networks (CNNs) and transformer-based architectures to automatically learn perceptual quality representations from large datasets [3]. These learning-based models demonstrate stronger correlation with human subjective assessment and improved adaptability to diverse distortion types. In remote sensing contexts, deep learning-based IQA methods have been applied to evaluate super-resolution outputs, environmental robustness, and cross-sensor performance [1], [4]. Additionally, task-aware quality evaluation frameworks integrate downstream objectives—such as classification or detection accuracy—into the assessment process, bridging the gap between numerical fidelity and practical usability.

This progression from pixel-level similarity metrics to AI-driven perceptual models reflects a broader trend toward intelligent and adaptive quality evaluation systems capable of handling high-dimensional satellite data [3], [5].

Robustness Evaluation in Satellite Imagery

Robustness evaluation examines how IQA metrics perform under various distortions including noise, blur, compression artifacts, and adverse environmental conditions. Satellite imagery is particularly susceptible to such degradations due to atmospheric interference, sensor instability, and transmission constraints [4]. Traditional metrics often show inconsistent reliability when exposed to combined or non-uniform distortions.

Research in environmental robustness indicates that spectral-aware and learning-based metrics provide improved stability under illumination shifts and atmospheric scattering [4]. Similarly, deep learning frameworks demonstrate enhanced resilience to complex distortions compared to handcrafted measures [3]. However, cross-dataset generalization remains a significant challenge. IQA models trained on specific sensors or environmental conditions may not perform consistently across different satellite platforms or geographic regions.

Addressing robustness under noise, blur, compression, adversarial perturbations, and cross-sensor variability is critical for developing dependable satellite image quality assessment systems. Emerging adaptive and data-driven approaches show promise in achieving more consistent and operationally reliable performance across diverse real-world scenarios [1]–[4].

COMPARATIVE ANALYSIS OF IMAGE QUALITY METRICS FOR SATELLITE IMAGERY

A systematic comparison of image quality assessment (IQA) metrics is necessary to evaluate their robustness, suitability for multispectral data, computational complexity, and inherent limitations within satellite imaging environments. Drawing on recent studies [1]–[5], Table 1 provides a consolidated overview of widely used traditional, spectral-aware, and learning-based quality evaluation methods applied to remote sensing imagery.

Comparative Discussion

Traditional full-reference metrics such as MSE and PSNR remain popular in satellite image evaluation due to their simplicity, reproducibility, and low computational overhead, particularly in super-resolution benchmarking tasks [1]. However, extensive research indicates that these metrics exhibit weak alignment with perceptual quality and are inadequate for representing spectral distortions in multispectral and hyperspectral imagery [3], [4]. Although SSIM improves structural similarity assessment by incorporating luminance and contrast comparisons, it still lacks comprehensive spectral sensitivity.

Table 1. Comparative Analysis of Satellite Image Quality Metrics

Metric	Type	Robust to Noise	Suitable for Multispectral Data	Computational Cost	Major Limitation
MSE	Full-Reference (FR)	Low	No	Very Low	Weak perceptual correlation; ignores structural and spectral information [3]
PSNR	Full-Reference (FR)	Low–Moderate	Limited	Very Low	Pixel-error sensitive; ineffective for perceptual and spectral distortions [1], [4]
SSIM	Full-Reference (FR)	Moderate	Limited	Low	Emphasizes spatial structure; limited spectral awareness [3], [4]
SAM	Full-Reference (FR)	Moderate	Yes	Moderate	Overlooks spatial texture and perceptual attributes [4]
ERGAS	Full-Reference (FR)	Moderate	Yes	Moderate	Sensitive to radiometric shifts; lacks perceptual modeling [4]
BRISQUE	No-Reference (NR)	Moderate	Limited	Low	Based on natural scene statistics; limited generalization to satellite imagery [3]
NIQE	No-Reference (NR)	Moderate	Limited	Low	Domain mismatch for remote sensing data [3]
CNN-based NR-IQA	No-Reference (NR)	High	Potentially Yes	High	Requires labeled datasets; risk of overfitting [1], [3]
Transformer-based IQA	No-Reference (NR)	High	Potentially Yes	High–Very High	Computationally intensive; data-dependent [3]
Hybrid Spectral–Spatial Metrics	FR / Learning-Based	High	Yes	Moderate–High	Increased complexity; lack of standardized benchmarks [4], [5]

Spectral-oriented metrics such as SAM and ERGAS offer improved suitability for evaluating multispectral data by preserving spectral signature integrity, which is essential for applications including vegetation monitoring and land-cover classification [4]. Despite their effectiveness in representing spectral fidelity, these measures often neglect perceptual spatial characteristics such as texture realism and contextual structure.

No-reference approaches, including BRISQUE and NIQE, provide practical advantages in operational satellite systems where reference images are unavailable [3]. Nevertheless, because these models rely on natural scene statistical assumptions derived from terrestrial imagery, their generalization capability to remote sensing data remains limited.

Recent progress in deep learning-based IQA frameworks demonstrates enhanced robustness against distortions such as noise, blur, compression artifacts, and environmental variability [1], [3]. CNN-based and transformer-based architectures learn high-level spatial and contextual representations that better correlate with subjective human perception. Moreover, hybrid spectral–spatial models show promising performance in multispectral robustness evaluation by jointly modeling spectral consistency and spatial structure [4]. However, these approaches demand substantial computational resources and annotated training data, posing challenges for real-time onboard satellite implementation.

Overall, the comparative evaluation suggests that no single metric simultaneously achieves optimal robustness, spectral fidelity, perceptual alignment, and computational efficiency. Consequently, contemporary research trends advocate for hybrid, task-aware, and domain-adaptive IQA frameworks specifically designed for satellite imaging systems [3], [4].

PERFORMANCE EVALUATION CRITERIA

Assessing the effectiveness of image quality assessment (IQA) metrics requires well-defined statistical and computational evaluation standards. In the context of satellite imagery, performance validation must account for perceptual relevance, statistical agreement with subjective evaluations, computational practicality, and robustness under varying environmental and sensor conditions. Drawing on recent studies [1]–[4], the following criteria are commonly employed in IQA research for remote sensing applications.

Alignment with Human Perception

A central goal of IQA models is to approximate human visual judgment. Subjective quality scores, often expressed as Mean Opinion Score (MOS), represent the perceived level of degradation as evaluated by human observers. Objective metrics are therefore validated based on how strongly they correlate with these subjective assessments. Conventional measures such as PSNR frequently demonstrate limited perceptual alignment, particularly in super-resolution and multispectral satellite imagery scenarios [1], [3]. Recent research highlights the importance of perceptually motivated and learning-based models that better capture structural and contextual distortions affecting human interpretation [3]. In remote sensing applications, perceptual fidelity is especially critical when imagery supports manual analysis, mapping, and decision-making processes [5].

Spearman Rank Correlation Coefficient (SRCC)

The Spearman Rank Correlation Coefficient (SRCC) evaluates the monotonic relationship between predicted objective quality scores and subjective human ratings. Specifically, it measures how consistently an IQA model preserves the ranking order of images compared to human perception. Because SRCC does not assume a linear relationship between predicted and subjective scores, it is widely adopted in IQA benchmarking studies [3]. In satellite image evaluation, SRCC is frequently reported for deep learning-based frameworks to demonstrate robustness against complex distortions such as atmospheric interference and environmental variability [4].

Pearson Linear Correlation Coefficient (PLCC)

The Pearson Linear Correlation Coefficient (PLCC) measures the linear agreement between objective predictions and MOS values. Unlike SRCC, which focuses on ranking consistency, PLCC quantifies how closely predicted scores approximate the actual magnitude of perceived degradation. High PLCC values indicate strong predictive accuracy in absolute terms. In remote sensing super-resolution and enhancement studies, PLCC is commonly used to validate both traditional and learning-based IQA approaches [1], [3].

Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) calculates the average deviation between predicted objective scores and corresponding subjective ratings. Lower RMSE values indicate better predictive precision. RMSE is typically reported alongside SRCC and PLCC to provide a comprehensive statistical assessment of IQA performance [3]. In multispectral environmental robustness studies, RMSE analysis reveals how certain metrics degrade in reliability under illumination changes and atmospheric disturbances [4].

Computational Efficiency

Computational efficiency refers to the processing time and hardware resources required by an IQA metric. In operational satellite systems—particularly onboard processing environments—low complexity and real-time capability are critical. Traditional full-reference metrics such as MSE and

PSNR are computationally lightweight and well-suited for rapid evaluation tasks [1]. In contrast, deep learning-based IQA models, including CNN- and transformer-based architectures, demand substantial computational resources and memory, limiting their feasibility for real-time onboard deployment [3]. Consequently, practical implementation often involves balancing robustness with computational constraints.

Generalization Capability

Generalization capability reflects how effectively an IQA model performs across diverse datasets, sensor platforms, spatial resolutions, and environmental conditions. Satellite imagery exhibits considerable variability in spectral composition, geographic coverage, and acquisition parameters. Metrics trained on conventional natural image datasets may not transfer reliably to remote sensing domains due to distributional differences [3]. Robustness analyses of multispectral imagery underscore the necessity of cross-sensor validation and domain-adaptive modeling frameworks [4]. IQA methods with strong generalization ability are particularly valuable for large-scale Earth observation systems operating under heterogeneous and dynamic environmental conditions [6].

Overall, these evaluation criteria collectively ensure that IQA models for satellite imagery are not only statistically reliable but also perceptually meaningful, computationally feasible, and operationally robust across real-world scenarios [1]–[4].

CURRENT RESEARCH GAPS

Although substantial progress has been made in developing image quality assessment (IQA) techniques for satellite imagery, several important challenges remain unresolved. A review of recent studies [1]–[5] reveals persistent limitations that hinder the development of robust, perceptually aligned, and operationally practical quality evaluation frameworks tailored to remote sensing data.

Absence of Standardized Satellite IQA Datasets

A fundamental limitation in advancing satellite-specific IQA models is the lack of comprehensive, standardized benchmark datasets containing large-scale imagery with reliable subjective quality annotations. While numerous natural image IQA datasets exist, they do not reflect the complex spectral characteristics, environmental variability, and sensor-specific properties of satellite imagery. Contemporary IQA surveys emphasize that data-driven and learning-based models depend heavily on domain-specific annotated datasets to achieve reliable performance and generalization [3]. Moreover, robustness analyses in multispectral environments demonstrate that quality perception is strongly influenced by environmental conditions, reinforcing the need for dedicated satellite IQA benchmarks [4]. In the absence of standardized datasets, objective comparison and fair evaluation of competing quality metrics remain challenging.

Weak Perceptual Alignment in Spectral Imagery

Conventional metrics such as MSE, PSNR, and SSIM predominantly assess spatial similarity and often fail to capture perceptual distortions inherent in multispectral and hyperspectral data. In remote sensing super-resolution research, improvements in PSNR do not consistently translate into perceptually meaningful or semantically accurate reconstructions [1]. Additionally, studies on multispectral robustness indicate that spectral distortions caused by environmental factors may not substantially affect spatial similarity scores, potentially leading to misleading quality evaluations [4]. Recent IQA reviews highlight the necessity of integrating perceptual modeling with joint spectral–spatial consistency measures to enhance correlation with human interpretation and downstream task performance [3].

Limited Robustness to Atmospheric Effects

Atmospheric phenomena—including haze, scattering, and illumination changes—introduce nonlinear distortions that challenge traditional quality metrics. Environmental robustness investigations

reveal that spatial similarity measures often lose reliability under atmospheric variability [4]. Domain-specific research in solar EUV imaging further demonstrates that radiative interference and environmental noise require specialized evaluation frameworks beyond standard statistical approaches [2]. Although deep learning-based IQA methods show promise in modeling complex distortions [3], achieving consistent robustness across diverse atmospheric conditions and geographic regions remains an open research problem.

Inadequate Cross-Sensor Validation

Satellite imagery is captured by heterogeneous platforms with differing spatial resolutions, spectral configurations, and calibration characteristics. Despite this diversity, many IQA models are validated on limited datasets associated with specific sensors. General IQA research indicates that models trained within one data distribution may exhibit reduced performance when applied to different domains due to distributional shifts [3]. Similarly, multispectral robustness studies emphasize the importance of cross-sensor validation to ensure consistent metric performance across varying acquisition conditions [4]. The lack of systematic cross-platform benchmarking restricts the operational reliability of many proposed satellite IQA approaches.

High Computational Demands of Deep Learning Models

Learning-based IQA frameworks, including convolutional neural networks and transformer-based architectures, demonstrate enhanced perceptual alignment and robustness compared to traditional metrics [3]. However, these models require substantial computational resources, extensive training data, and significant processing time. In remote sensing super-resolution tasks, hybrid deep learning models introduce considerably higher computational complexity than conventional metrics such as PSNR and SSIM [1]. Given that real-time onboard satellite systems operate under strict hardware and power constraints, achieving an effective balance between robustness, predictive accuracy, and computational efficiency remains a critical research challenge.

FUTURE RESEARCH DIRECTIONS

Drawing from an extensive review of recent studies [1]–[18], several promising research avenues emerge for advancing robust and application-oriented image quality assessment (IQA) frameworks in satellite imagery.

Hybrid Spectral–Spatial Quality Metrics

Conventional spatial similarity measures such as PSNR and SSIM continue to dominate super-resolution and pan-sharpening evaluation tasks [1], [8], [12]. However, these metrics are primarily intensity-based and insufficient for capturing spectral distortions present in multispectral and hyperspectral imagery. Research on spectral-aware evaluation for pan-sharpened and hyperspectral data underscores the importance of preserving inter-band relationships and spectral consistency [7], [12], [18]. Furthermore, robustness analyses indicate that environmental variability and structural distortions jointly influence spatial and spectral fidelity [4], [9].

Future investigations should therefore emphasize hybrid spectral–spatial frameworks that integrate structural similarity, spectral angle-based metrics, modulation transfer function (MTF)-based assessment [13], and perceptually motivated modeling into unified evaluation strategies [1], [4], [12], [18]. Such integrated approaches could provide a more comprehensive representation of satellite image quality.

Lightweight Deep Learning-Based IQA Models

Deep learning-driven IQA approaches—including neural network-based full-reference models [11] and AI-enhanced super-resolution evaluation techniques [1], [8]—have demonstrated superior perceptual alignment and robustness compared to traditional metrics. Nevertheless, these architectures typically require significant computational resources, limiting their feasibility in real-time satellite

systems [16]. Broader surveys on IQA evolution highlight the increasing architectural complexity of CNN- and AI-based methods [3], [8], [17].

To facilitate practical deployment in operational satellite environments, future research should focus on lightweight network architectures, model compression techniques, knowledge distillation, and efficient feature extraction strategies specifically tailored to remote sensing data characteristics [1], [11], [16].

Task-Oriented Quality Evaluation

Most conventional IQA frameworks evaluate pixel-level similarity without accounting for downstream application performance. However, quality inspection research in land-use mapping indicates that image degradations can significantly impair classification and cartographic accuracy even when numerical quality metrics show minimal variation [9]. Similarly, multispectral robustness analyses demonstrate that structural and spectral degradations influence vertical structure estimation and environmental modeling tasks [4]. Pan-sharpening evaluation studies further emphasize the need for application-aware validation protocols [12].

Future IQA systems should therefore adopt task-driven evaluation paradigms, where image quality is assessed based on its impact on classification accuracy, segmentation reliability, object detection performance, and other mission-specific outcomes [4], [9], [12], [17]. Integrating analytical task performance into quality metrics can bridge the gap between numerical fidelity and operational relevance.

Domain-Adaptive IQA Frameworks

Satellite imagery exhibits substantial heterogeneity across sensors, spectral resolutions, acquisition geometries, and geographic regions. Investigations involving ZY3 datasets [15], multi-satellite pan-sharpening systems [7], and high-resolution MTF-based evaluations [13] reveal sensor-dependent variability in metric performance. Additionally, statistical analyses of no-reference IQA methods demonstrate limited cross-dataset generalization capability [14]. Broader IQA literature identifies domain shift as a critical limitation of learning-based models [3], [17].

To enhance generalization and operational reliability, future research should prioritize domain-adaptive IQA approaches leveraging transfer learning, self-supervised learning, and cross-sensor validation frameworks [3], [7], [14], [15]. Such strategies can improve robustness under varying acquisition and environmental conditions.

Real-Time Onboard Satellite IQA Systems

The growing demand for near-real-time Earth observation necessitates automated onboard quality monitoring systems. Comparative studies of real-time satellite IQA highlight the need for efficient yet reliable evaluation techniques [16]. While traditional measures such as MSE and PSNR offer computational simplicity [1], [8], they lack robustness to complex distortions. Conversely, neural-network-based metrics provide improved accuracy but impose significant computational overhead [11]. Emerging combined no-reference approaches optimized for remote sensing imagery present a promising direction for operational deployment [10].

Future research should therefore focus on developing adaptive, low-latency IQA frameworks capable of detecting blur, noise, compression artifacts, and calibration anomalies during image acquisition and transmission processes [10], [16]. Achieving a balance between computational efficiency and robustness will be essential for practical onboard satellite implementations.

CONCLUSION

This survey has provided a structured and comprehensive overview of robust image quality assessment (IQA) methodologies for satellite imagery, encompassing traditional statistical metrics,

spectral-aware evaluation techniques, and emerging deep learning-based approaches. The analysis shows that classical full-reference measures such as MSE, PSNR, and SSIM continue to be widely utilized because of their simplicity and low computational requirements. However, these metrics often fall short in accurately representing perceptual realism, spectral distortions, and task-specific relevance in multispectral and hyperspectral remote sensing data.

Spectral-oriented metrics, including SAM, ERGAS, and pan-sharpening quality indices, offer improved evaluation of spectral fidelity but generally lack integrated modeling of perceptual characteristics and spatial structural consistency. Reduced-reference and no-reference approaches provide practical solutions in scenarios where reference images are unavailable, particularly in operational satellite systems. Nonetheless, their effectiveness is frequently constrained by domain dependence and variability across datasets and sensors.

Advances in deep learning-based IQA frameworks—such as CNN-driven and neural network-based full-reference models—have demonstrated stronger alignment with human perception and enhanced robustness against complex distortions including noise, blur, compression artifacts, and atmospheric effects. Despite these improvements, several critical challenges persist. These include the absence of standardized satellite-specific IQA benchmarks, limited cross-sensor validation, vulnerability to environmental variability, and the substantial computational demands of advanced AI-based models.

Overall, the survey findings suggest that no single metric currently satisfies all essential criteria, including perceptual consistency, spectral integrity, robustness, computational efficiency, and generalization capability. Future research should therefore prioritize hybrid spectral–spatial evaluation frameworks, lightweight and domain-adaptive deep learning architectures, and task-oriented quality assessment strategies. Additionally, the development of real-time, onboard IQA systems with adaptive and interpretable decision-making capabilities will be pivotal for next-generation satellite platforms.

In summary, the continued advancement of satellite-specific IQA methodologies is vital for ensuring reliable Earth observation, enhancing downstream analytical accuracy, and supporting informed decision-making across applications such as environmental monitoring, urban development, disaster management, and defense surveillance.

REFERENCES

1. Lu R, Miao H, Hai X. Efficient remote sensing image super-resolution with residual-enhanced wavelet and key-value adaptation. *Frontiers in Remote Sensing*. 2025;6:1718058.
2. Dai S, He L, Xu S, Sun L, Chen H, Yu S, Wu K, Wang Y, Xuan Y. Quality Assessment of Solar EUV Remote Sensing Images Using Multi-Feature Fusion. *Sensors*. 2025 Oct 14;25(20):6329.
3. Ma C, Shi Z, Lu Z, Xie S, Chao F, Sui Y. A survey on image quality assessment: Insights, analysis, and future outlook. *arXiv preprint arXiv:2502.08540*. 2025 Feb 12.
4. Yadav BK, Lucieer A, Jordan GJ, Baker SC. Assessing the Robustness of Multispectral Satellite Imagery with LiDAR Topographic Attributes and Ancillary Data to Predict Vertical Structure in a Wet Eucalypt Forest. *Remote Sensing*. 2025 May 15;17(10):1733.
5. Anikeeva I, Chibunichev A. Methodology for Visual Quality Estimation of Aerial and Satellite Images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2024 Dec 16;48:21-6.
6. Antonel LG. A Novel No-Reference Image Quality Metric For Assessing Sharpness In Satellite Imagery. *arXiv preprint arXiv:2410.10488*. 2024 Oct 14.
7. Pinheiro G, Rather IH, Raj A, Minz S, Kumar S. Image Quality Assessment of Multi-Satellite Pan-Sharpener Approach: A Case Study using Sentinel-2 Synthetic Panchromatic Image and Landsat-8. *EAI Endorsed Transactions on Scalable Information Systems*. 2024 Oct 1;11(6).
8. Arabboev M, Begmatov S, Rikhsivoev M, Nosirov K, Saydiakbarov S. A comprehensive review of image super-resolution metrics: classical and AI-based approaches. *Acta IMEKO*. 2024 Mar 12;13(1):1-8.

9. Dong S, Lu C, Gao W, Liu C, Bai J. Quality inspection and problem analysis of satellite image data in land use survey. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2024 May 10;48:129-34.
10. Rubel A, Ieremeiev O, Lukin V, Fastowicz J, Okarma K. Combined no-reference image quality metrics for visual quality assessment optimized for remote sensing images. *Applied Sciences*. 2022 Feb 14;12(4):1986.
11. Ieremeiev O, Lukin V, Okarma K, Egiazarian K. Full-reference quality metric based on neural network to assess the visual quality of remote sensing images. *Remote Sensing*. 2020 Jul 22;12(15):2349.
12. Dadras Javan F, Samadzadegan F, Mehravar S, Toosi A. A review on spatial quality assessment methods for evaluation of pan-sharpened satellite imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2019 Oct 18;42:255-61.
13. Wu Z, Luo Z, Zhang Y, Guo F, He L. Image quality assessment of high-resolution satellite images with MTF-based fuzzy comprehensive evaluation method. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2018 Apr 30;42:1907-14.
14. Li S, Yang Z, Li H. Statistical evaluation of no-reference image quality assessment metrics for remote sensing images. *ISPRS International Journal of Geo-Information*. 2017 Apr 28;6(5):133.
15. Zhang Y, Cui WH, Yang F, Wu ZC. No-reference image quality assessment for zy3 imagery in urban areas using statistical model. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2016 Jun 10;41:949-54.
16. Rajkumar S, Malathi G. A comparative analysis on image quality assessment for real time satellite images. *Indian J. Sci. Technol*. 2016 Sep;9(34):1-1.
17. Mohammadi P, Ebrahimi-Moghadam A, Shirani S. Subjective and objective quality assessment of image: A survey. *arXiv preprint arXiv:1406.7799*. 2014 Jun 30.
18. Singh AK, Kumar HV, Kadambi GR, Kishore JK, Shuttleworth J, Manikandan J. Quality metrics evaluation of hyperspectral images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2014 Nov 28;40:1221-6.