

Review

A Survey and Tutorial on Image Quality Assessment with a Contrast-Weighted Structural Similarity Framework

Sos S. Agaian ¹, Artyom M. Grigoryan ^{2,*} and Hrach Ayunts ³

¹ Department of Computer Science, School of Engineering, City University of New York (CUNY), New York, NY 10031, USA; sos.agaian@csi.cuny.edu

² Department of Electrical and Computer Engineering, The University of Texas at San Antonio, San Antonio, TX 78249, USA

³ Department of Informatics and Applied Mathematics, Yerevan State University, Yerevan 0025, Armenia

* Correspondence: artyom.grigoryan@utsa.edu

Abstract

Objective Image Quality Assessment (IQA) is a fundamental pillar of computer vision, essential for optimizing tasks ranging from supervised machine learning to real-time video streaming. While IQA aims to quantify image degradation caused by noise and artifacts, a persistent gap remains between technical objective measurements and subjective human perception. Objective IQA has advanced significantly through full-reference (FR) metrics designed to approximate human judgment. Standard measures such as the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and root mean square error (RMSE) provide established benchmarks; however, they frequently fail to capture nuanced human visual preferences, often penalizing perceptually insignificant shifts or favoring overly smoothed images. Conversely, modern deep-learning metrics like LPIPS offer better perceptual alignment but remain computationally prohibitive for real-time, resource-constrained environments. This paper addresses these challenges through a dual-purpose approach. First, it provides a comprehensive survey and tutorial of the IQA landscape, offering self-contained mathematical derivations of classical error sensitivity measures, including MSE, RMSE, MAE, Euclidean distance, RMSLE, and Huber loss, as well as artificial neural network (ANN) approaches. This foundational review ensures a rigorous understanding of the field's mathematical evolution. We introduce the Adaptive Contrast-Weighted Structural Similarity (ACSSIM) framework. ACSSIM is a lightweight hybrid metric that enhances classical FR-IQA by incorporating local weighting derived from human visual system (HVS) properties. Specifically, it targets Weber's Law-based contrast and entropy, which are key elements of our hybrid quality assessment logic and key components of non-reference image quality metrics. Extensive numerical experiments on the TID2013 and KADID-10k benchmark show that ACSSIM improves correlation with human subjective judgments compared with the baseline PSNR and SSIM. Our results confirm that ACSSIM maintains low computational overhead, bridging the gap between efficiency and accuracy for practical deployment. We made our code publicly available to facilitate future research in efficient perceptual modeling.

Academic Editor: Xinming Zhang

Received: 22 March 2026

Revised: 15 June 2026

Accepted: 25 June 2026

Published: 27 June 2026

Copyright: © 2026 by the authors.

Submitted for possible open access

publication under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

Keywords: similarity metrics; full reference; structural similarity index

1. Introduction

Modern computer vision has shifted from purely signal-based approaches to those that explicitly incorporate the intricacies of human visual perception (HVP). A key takeaway from this shift is that ignoring the biological limitations of the human observer results in less effective image processing and evaluation. Metrics that mirror subjective human judgment, such as the mean opinion scores (MOS) and differential mean opinion scores (DMOS), continue to serve as the benchmark for validating objective algorithms [1–3].

This human-centric paradigm is critical in high-stakes domains such as medical diagnostics (e.g., mammography), automated industrial quality control, and pattern recognition, where visual fidelity directly affects clinical or operational outcomes. However, in real-world scenarios such as cloud streaming, IoT surveillance, and clinical imaging, a “pristine” reference image is often unavailable. This need has led to no-reference (NR) methods that assess quality or similarity using perceptual principles, information theory (entropy), and natural-scene statistics (NSS).

A. Taxonomy of Image Similarity Approaches

Pixel-based similarity [4–14]: These full-reference (FR) metrics, such as the PSNR and the mean squared error (MSE), are widely used for their computational simplicity. However, they lack a perceptual basis and often fail to capture spatial dependencies. The SSIM shifted focus toward evaluating luminance, contrast, and structural information. While the SSIM is more perceptually aligned, it remains sensitive to geometric shifts and requires perfectly aligned reference pairs.

Histogram-Based Similarity [15–17]: This approach uses histograms to represent the overall statistical distributions of color or intensity. Objects are described using histograms of visual features like SIFT [18], SURF [19], Spatial Envelope [20], or HOG [21], which provide some invariance to scale and rotation. These methods are effective for quick database searches and matching probability distributions, such as with the chi-square distance. However, they do not account for spatial relationships and can be affected by quantization effects [22].

Feature-Based Similarity [23–28]: By extracting high-level primitives such as edges, corners, and shapes using algorithms like ORB (Oriented FAST and Rotated BRIEF), these methods are robust to illumination changes and affine transformations.

Deep learning-based similarity [29–34]: Semantic similarity learning uses convolutional neural networks (CNNs) to embed images in a structured feature space. Metrics such as LPIPS (learned perceptual image patch similarity) correlate strongly with human judgment but incur high computational costs, offer low interpretability (a “black-box” nature), and are vulnerable to adversarial perturbations.

The summary of the above-mentioned metrics is given in Table 1.

Table 1. Summary of representative metric families.

Metric	Goal/Intuition	Key Limitations
MSE/PSNR	Pixel-wise error; Signal fidelity	No perceptual basis; sensitive to misalignment.
SSIM/MS-SSIM	Structural, luminance, and contrast modeling	Requires pixel-perfect alignment; tends to favor over-smoothing.
Histogram/Chi-Square	Distributional matching; Global stats	Ignores spatial structure; sensitive to bin quantization.
Deep Metrics (LPIPS)	Semantic feature space distances	High compute; lack of transparency; hardware-intensive.

This work aims to present (i) a survey and tutorial on image quality assessment; the background provides the essential mathematical scaffolding for readers to understand the evolution of IQA, and (ii) the adaptive contrast-weighted structural similarity (ACSSIM) framework lies in its ability to be integrated with various standard measures discussed in the background, such as PSNR, SSIM, and RMSE. By establishing these metrics early, we provide the necessary context for how ACSSIM enhances them.

B. The ACSSIM Framework: A Conceptual Preview

To bridge the gap between computationally expensive deep learning models and overly simplistic traditional metrics, this paper also introduces the ACSSIM framework.

Unlike traditional static metrics, ACSSIM acts as a perceptual wrapper that can be integrated with any baseline metric (e.g., PSNR, SSIM, or RMSE). By leveraging local entropy and Weber-law-based contrast weighting, ACSSIM dynamically adjusts the significance of image regions according to the human visual system sensitivity. It effectively “masks” noise in high-visual-complexity areas while emphasizing structural degradation in salient regions.

The ACSSIM framework formalizes this by decomposing the reference image f and distorted image g into non-overlapping blocks $B_{i,j}$. For each block, a baseline similarity score $S(B_{i,j}, \hat{B}_{i,j})$ is calculated using a standard FR metric. The final global score is then derived as a perceptually-weighted average:

$$ACSSIM(f, g) = \frac{\sum_{i,j} w_{i,j} S(B_{i,j}, \hat{B}_{i,j})}{\sum_{i,j} w_{i,j}}, \quad (1)$$

where the weight $w_{i,j}$ is a function of the following:

- (a) Local contrast (alpha): Derived from a modified Weber-law or the Agaian–Panetta contrast measures, ensuring that high-contrast areas, where the HVS is more sensitive to structural changes, are prioritized.
- (b) Local entropy (eta): Acting as a proxy for information content; blocks with higher entropy (higher complexity) are weighted to account for the HVS’s reduced sensitivity to noise in busy textures.

C. Primary Contributions

Unified Theoretical Review: A cohesive analysis of classical FR and modern NR methods.

Methodological Innovation: The introduction of the lightweight ACSSIM framework, which enhances sensitivity to perceptually significant distortions without the overhead of deep learning.

Empirical Validation: Rigorous testing on the TID2013 and KADID-10k benchmarks shows that ACSSIM greatly surpasses standard SSIM in matching human subjective scores. The results also confirm that low-complexity, perceptually motivated extensions can close the gap between efficiency and accuracy in resource-constrained environments.

The remainder of this paper is organized as follows. Section 2 reviews established full-reference (FR) metrics, focusing on the mathematical properties of PSNR, SSIM, and RMSE, as well as the robust characteristics of the Huber loss function. Section 3 explores distance metrics in feature space, providing a comparative analysis of Euclidean and Chebyshev distances and the application of the Minkowski distance across multimodal data types, including image, audio, and text. Section 4 discusses the discrepancy between subjective and objective assessment, detailing the MOS/DMOS standards used for validation. Section 5 establishes the perceptual foundations of vision, extending beyond classical Weber’s and Michelson’s laws to incorporate modern psychophysics. This section also introduces local information entropy and its intrinsic relationship with visual contrast. Section 6 examines no-reference (NR) measures, specifically the EME, EMEE, and AMEE contrast measures developed by Agaian and Panetta. Section 7 introduces the core contribution of this work: the ACSSIM framework, which hybridizes FR and NR

assessment strategies. This section also presents the experimental results and discussion, using the TID2013 and KADID-10k benchmarks for validation. Finally, Section 8 provides concluding remarks and suggests directions for future research.

2. Full-Reference Metrics

Recent advances in computer vision and image processing have moved away from traditional methods that treat image signals as physical quantities interacting with observers via mathematical models. This shift reveals an important insight: ignoring the human observer can impede progress in image processing research. Understanding and measuring this subjective experience is crucial for developing better image processing techniques, improving visual quality, benchmarking algorithms, and monitoring image quality in real-time applications. Since humans are the ultimate users of visual content, their subjective experience is vital. Image similarity measures and assessments are specific computer vision methods that offer a numerical, quantitative evaluation of how alike two images or image patches are. These techniques are essential for automating tasks like the content-based image retrieval (CBIR), duplicate detection, and medical image registration. The primary objectives of the image similarity measures are as follows:

- (1) To translate the subjective visual similarity between two images into an accurate, objective, and comparable numerical value;
- (2) To identify near-duplicate images, even those with varying resolutions, formats, or minor modifications such as cropping or brightness adjustments, within extensive databases;
- (3) To provide a metric that assists image analysis algorithms, such as those used in medical imaging, to align images of the same subject captured at different times or modalities, and to emulate human perception when evaluating the quality of restored or compressed images compared to the original.

In recent years, significant progress has been made in objective image quality assessment, especially in full-reference metrics designed to approximate human subjective quality judgments. These measures—such as the PSNR, SSIM, and RMSE—are widely used in computer vision, medical imaging, and multimedia processing [1–3]. However, despite their popularity, traditional FR metrics often fail to accurately capture perceptual similarity. They are sensitive to small geometric misalignments, illumination changes, and content-preserving transformations, and they frequently rank images in ways that conflict with human preference, for example, assigning higher scores to overly smooth images than to realistically textured ones.

Full-reference, or pixel-based, metrics are essential tools for measuring image similarity and quality. These metrics compare pixel values between a test image and a perfect reference image, providing numerical measures of image quality and resemblance. While they are computationally efficient and mathematically precise, they face challenges in capturing how humans perceive image quality. In machine learning, pixel-based metrics are often used as cost or loss functions to guide model training. Defining and measuring image similarity involves considering several important factors. First, we need to identify and quantify the similarities by choosing suitable metrics, such as the Euclidean distance for direct pixel differences or the cosine similarity for comparing vectors based on angles. Second, the choice of a suitable metric depends on the characteristics of the data, such as type, scale, distribution, and dimensionality. Third, when implementing this, it is necessary to analyze various image transformations, variations, and noise. Additionally, practical issues such as missing data, outliers, and normalization must be carefully addressed.

The human visual system can distinguish between original images and distorted ones, such as blurred, rotated, or shifted images. Figure 1 illustrates the challenges associated with traditional pixel-wise image similarity measures. Computational methods

that rely on pixel-based metrics struggle to match this ability. Ideally, these metrics should stay consistent despite differences in amplitude or intensity, while still detecting structural differences that affect perceived quality. These metrics should remain unchanged when the amplitude or intensity changes, ensuring that similarity calculations focus on relative differences rather than absolute pixel values. For example, a strong similarity measure should identify two identical images as very similar, except for a brightness adjustment.

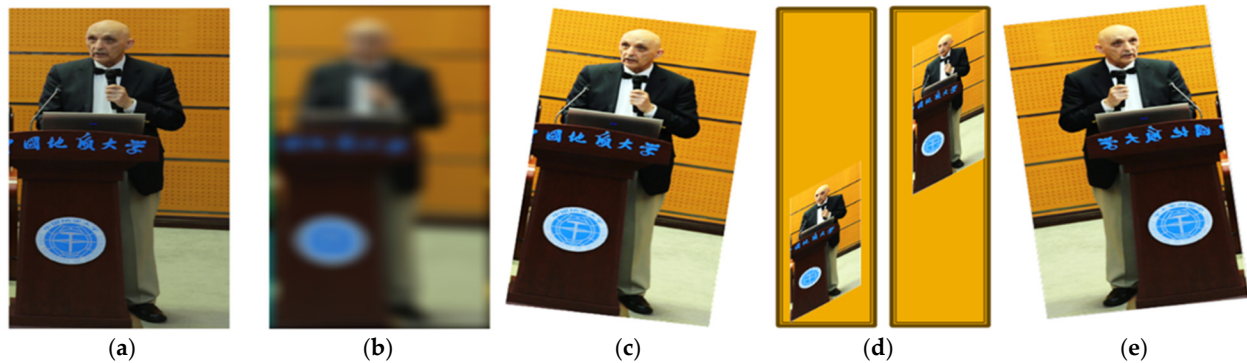


Figure 1. (a) The original image; (b) a blurred image; (c) an enhanced image; (d) translated images; and (e) a rotated image. The non-English term appearing in the figure is "China University of Geosciences".

2.1. Image Similarity Measurement PSNR and SSIM

Pixel-based measures are the most straightforward method for comparing images, as they involve examining each pixel individually, much like analyzing every brush stroke in two paintings. This approach treats each image as a collection of numerical values that represent color and brightness at every point.

Definition 1: The MSE between two images $X = \{X_{i,j}\}$ and $Y = \{Y_{i,j}\}$ of size $N \times M$ pixels each is defined as

$$MSE = MSE(X, Y) = \frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (X_{i,j} - Y_{i,j})^2. \quad (2)$$

We also consider a full-reference metric for predicting image perceived quality, namely the structural similarity index measure (SSIM). Introduced in 2003, this measure has changed the approach to evaluating image quality using machines [3,4]. This measure addresses the limitations of pixel-based methods; it more effectively reflects human perception of images. Instead of analyzing individual pixels, SSIM assesses the structural relationships within images, such as pixel connections and pattern organization. It emphasizes three key elements: luminance (brightness), contrast (variations in light), and structure. By focusing on these aspects, SSIM produces scores that generally align more closely with human judgments than traditional pixel-based comparison techniques.

Definition 2: The SSIM between two images X and Y with the dynamic range $L = (2^l - 1)$, $l > 1$, is calculated as

$$SSIM(X, Y) = \frac{2\mu_X\mu_Y + c_1}{\mu_X^2 + \mu_Y^2 + c_1} \cdot \frac{2\sigma_{XY} + c_2}{\sigma_X^2 + \sigma_Y^2 + c_2}, \quad (3)$$

where μ_X and μ_Y are the means and σ_X^2 and σ_Y^2 are the sample variances of X and Y , respectively. σ_{XY} is the covariance of X and Y . Two coefficients c_1 and c_2 are considered to be $c_1 = (0.01L)^2$ and $c_2 = (0.03L)^2$.

As an example, we consider the RGB color image $f_{n,m} = (r_{n,m}, g_{n,m}, b_{n,m})$ of size 1146×1369 pixels and range $[9, 255]$ which is shown in Figure 2a with its color components in parts (b–d). The values of the similarity metrics SSIM and MSE between the color components are given in Table 2. For this flower image, the highest similarity is between the red and green colors, and the lowest between the red and blue colors.

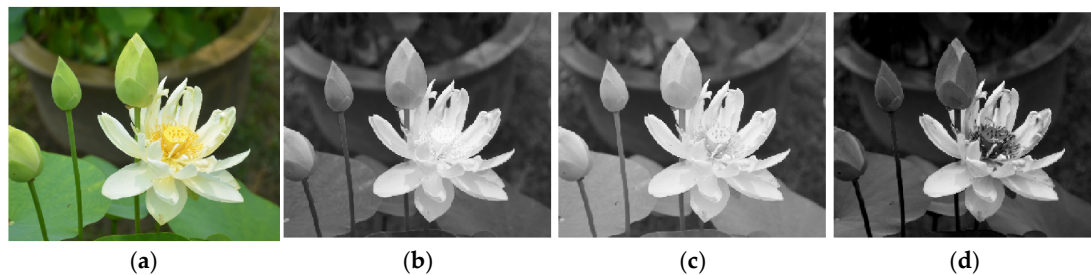


Figure 2. (a) The color ‘flower’ image and its (b) red, (c) green, and (d) blue components.

Table 2. Four metrics for the color components of the ‘flower’ image.

Image	SSIM	MSE	PSNR	MAE
Red and Green	0.9265	831.65	18.90	21.51
Green and Blue	0.7710	3301.31	12.91	49.91
Red and Blue	0.1942	6073.46	10.26	59.61

We also consider the gray component of the color image. This component can be calculated as the average of the colors, which is $a_{n,m} = (r_{n,m} + g_{n,m} + b_{n,m})/3$. The intensity of the image can also be used for the grayscale component, which is calculated by $i_{n,m} = 0.30r_{n,m} + 0.59g_{n,m} + 0.11b_{n,m}$. The gray and intensity images of the color image are shown in Figure 3 in parts (a) and (b), respectively. The difference in these images (in absolute scale and multiplied by 9) is shown in part (c).

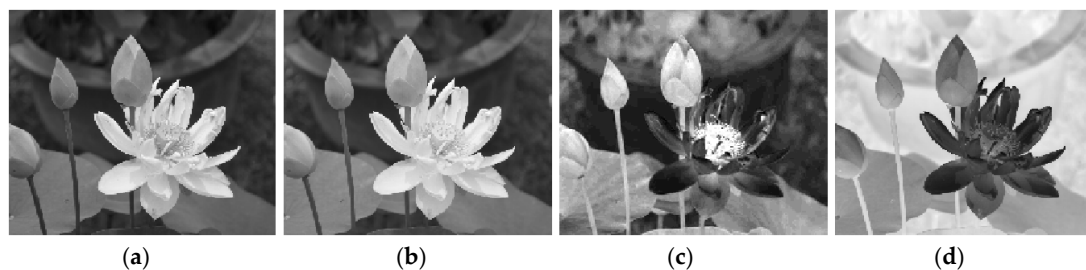


Figure 3. (a) The gray, (b) the intensity, (c) the difference in these two images, and (d) the negative image.

The data of the similarity metrics for these images are given in Table 3. One can see from this table that the high similarity, 0.9871, has a place for the gray and intensity components of the image. The color red is similar to gray, and the green is similar to the intensity.

Table 3. Four metrics for components of the ‘flower’ image.

Image	SSIM	MSE	PSNR	MAE
Gray and Intensity	0.9871	182.08	25.49	11.70
Gray and difference	0.1743	5988.32	10.32	58.28
Intensity and difference	0.2816	5268.30	10.88	52.26
Red and Gray	0.9767	179.55	25.55	9.19
Green and Gray	0.9500	738.10	19.42	23.11
Blue and Gray	0.9089	1002.50	18.09	26.80
Red and Intensity	0.9720	252.24	24.08	11.50
Green and Intensity	0.9857	200.17	25.08	11.41
Blue and Intensity	0.8402	2013.08	15.06	38.50

The negative gray flower image, $g_{n,m} = 255 + 9 - a_{n,m}$, is shown in Figure 3d. Should the negative image be considered similar to the original? The answer is that it should not. The coefficient of similarity of these images $SSIM(g, a) = -0.8583$.

It should be noted that our visual perception of images may not quite fit into the above metrics. We consider the example of images shown in Figure 4. The original image X in part (a) was processed with the 8-direction compass gradient and after rounding this gradient image $G(X)$ is shown in part (b). Its binary image Y , after using the threshold $T = 28$, is shown in part (c). The same binary image in the negative format is shown in part (d).

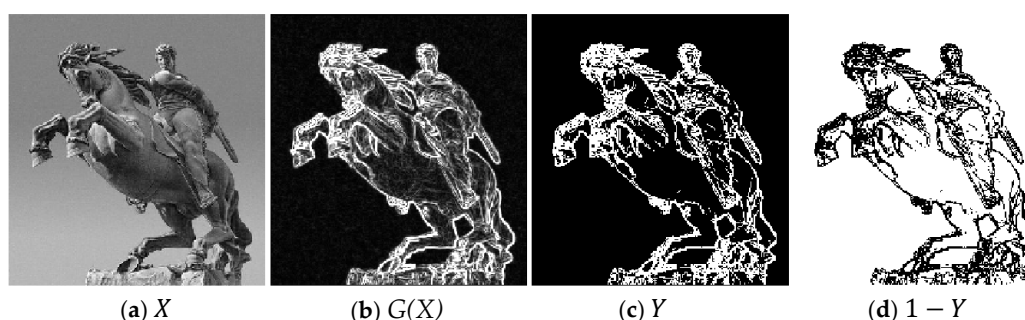


Figure 4. (a) The grayscale image, (b) the gradient image, (c) the threshold gradient image, and (d) the negative image of the gradient image in (c).

The data of similarities of these four images are given in Table 4. The score of similarity of the images X and $G(X)$ is negative and small, $SSIM(X, G(X)) = -0.0457$. Also, we have a small score of $SSIM(Y, G(X)) = 0.1223$. The pair of ‘exactly similar’ images Y and $(1 - Y)$ are estimated with the negative score of $SSIM(Y, 1 - Y) = -0.4020$ and minimum $MSE = 1$. According to the data in this table, these four images bear no resemblance whatsoever.

Table 4. Four metrics for the image in Figure 4.

Image	SSIM	MSE	PSNR	MAE
X and $G(X)$	−0.0457	15,967.21	6.10	114.68
X and Y	−0.0018	18,013.22	5.57	125.18
$G(X)$ and Y	0.1223	551.47	20.72	12.79
Y and $(1 - Y)$	−0.4020	1.00	48.13	1.00

Figure 5 demonstrates the limitations of the MSE and the advantages of the SSIM. The original image of a bird is shown in part (a) together with five distorted versions of the image in parts (b–f), for comparison. All five of these images have the same $MSE = 396$, but they demonstrate noticeably different visual quality and structural fidelity. Human observers perceive significant differences; some images are only slightly fuzzy, while others appear severely damaged or misaligned. This example illustrates why relying solely on the MSE can be misleading for assessing image quality.

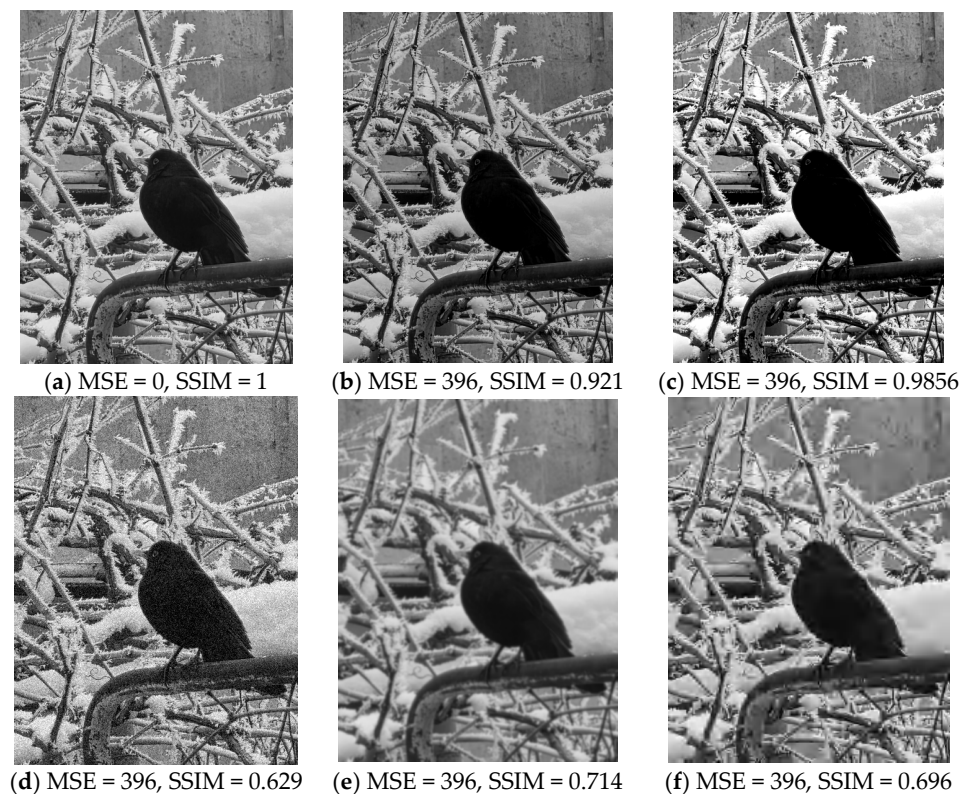


Figure 5. (a) The original image, (b) mean-shifted image, (c) contract-enhanced image, (d) salt-pepper impulsive noise-contaminated images, (e) blurred image, and (f) JPEG compressed image.

The PSNR measures how well the original signal (image or video) is preserved when it is altered or compressed. It compares the maximum possible signal power to the noise or distortion power. This is done using a logarithmic scale in decibels. The PSNR is primarily used in image compression techniques to evaluate the quality of an image or video reconstruction.

Definition 3: The PSNR of the original image X and a distorted one Y is calculated by

$$PSNR(X, Y) = 20 \log_{10} \max(X) - 10 \log_{10} MSE(X, Y) = -10 \log_{10} \frac{MSE(X, Y)}{L^2}. \quad (4)$$

When calculating the PSNR metric, it is also assumed that the noise in the signal is independent, which is not always the case in real-world scenarios, where the noise can be structured or correlated in ways that the PSNR cannot detect. Due to these shortcomings, alternative metrics, such as SSIM, are often used to assess image quality more accurately from a human perspective.

2.2. Root Mean Squared Error

In this section, we consider other metrics between images of $N \times M$ pixel each and with the range $L = 2^r - 1$.

Root mean square error in AI: It is a key metric used across artificial intelligence and data analysis to assess the accuracy of a model's predictions against actual data. Essentially, the RMSE is the square root of the MSE, making it more interpretable since it provides error in the same units as the data being predicted.

Definition 4: The RMSE of two images X and Y is calculated by

$$RMSE(X, Y) = \sqrt{MSE(X, Y)} = \sqrt{\frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (X_{i,j} - Y_{i,j})^2}. \quad (5)$$

Here, $X_{i,j}$ and $Y_{i,j}$ represent the predicted and actual values, respectively.

Root mean square logarithmic error (RMSLE) is a common evaluation metric in regression tasks, which operates on the logarithms of predicted $(p_1, p_2, \dots, p_n)$ and actual $(q_1, q_2, \dots, q_n)$ values, similar to the RMSE.

Definition 5: The RMSLE is calculated by

$$d_2(p, q) = \sqrt{\frac{1}{n} \sum_{i=1}^n ((\ln(1 + p_i) - \ln(1 + q_i))^2)}. \quad (6)$$

This formulation helps dampen the impact of large errors by compressing the value scale. The RMSLE is particularly effective when the target values exhibit exponential growth or large range variability.

Mean absolute error is another widely used loss function, especially in image processing and regression applications. It quantifies the average magnitude of absolute differences between corresponding elements of datasets without considering direction.

Definition 6: The MAE between two images $X = \{X_{i,j}\}$ and $Y = \{Y_{i,j}\}$ of size $N \times M$ pixels each is calculated by

$$MAE(X, Y) = \frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} |X_{i,j} - Y_{i,j}|. \quad (7)$$

Figure 6 shows the original grayscale image of the flower, X , of size 3024×2268 pixels in part (a). The range of this image is 255. The smoothed image, Y , is shown in part (b). This image was calculated by using the mean operation with window 3×3 . In part (c), the image X is shown after thresholding as

$$X_1(n, m) = \{X(n, m), \text{ if } X(n, m) \geq 100; 0, \text{ otherwise}\}. \quad (8)$$

Also, the following two images are shown after calculating with the threshold $T = 255/2$:

$$X_2(n, m) = \begin{cases} X(n, m), & \text{if } X(n, m) \geq T; \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad X_3(n, m) = 255 \times \begin{cases} 1, & \text{if } X(n, m) \geq T; \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

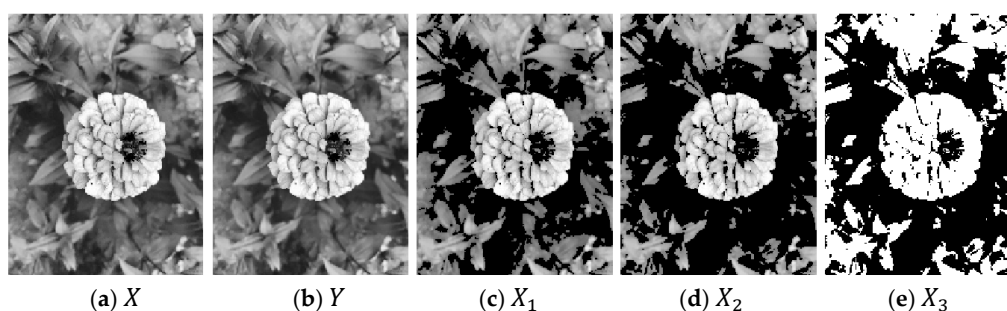


Figure 6. (a) The original ‘pirate’ image, (b) the smooth image, (c) and (d) two threshold images, and (e) the binary image.

Table 5 shows the metrics between the image X and images Y , X_1 , X_2 , and X_3 . The metrics of similarity of the image X_2 with its binary representation X_3 are also given. As expected, the best similarity is observed between images X and Y . We also note the high values of the metrics between image X_3 and X_2 compared with the metrics between X_3 and X .

Table 5. Four metrics for the flower images.

Image (from Figure 6)	SSIM	PSNR	MSE	MAE
X and Y	0.9996	43.51	2.90	1.00
X and X_1	0.8554	15.27	1932.52	26.35
X and X_2	0.7377	11.92	4182.24	45.98
X and X_3	0.6244	9.50	7304.11	79.55

X_2 and X_3	0.8605	13.19	3121.87	33.57
-----------------	--------	-------	---------	-------

Figure 7 shows the original grayscale image, X , ‘flowers.tiff’ of size 362×500 pixels in part (a). The intensities of this image are in the interval $[9, 255]$ and the mean $\mu(X) = 86.84$. The smoothed image, Y , is shown in part (b). The images in parts (c–e) are also calculated by Equations (8) and (9) (with the threshold $T = 132$). It should be noted the following. The image X_3 in part (e) is the binary copy of the image X_2 shown in part (d). However, the similarity metric SSIM between images X and X_2 is smaller than SSIM between image X and X_3 , that is, $SSIM(X, X_2) < SSIM(X, X_3)$. Table 6 shows the metrics between the ‘flowers’ images of Figure 7.



Figure 7. (a) The original ‘flowers’ image, (b) the smooth image, (c,d) two-threshold images, and (e) the binary image.

Table 6. Four metrics for ‘flowers’ images.

Image (from Figure 7)	SSIM	PSNR	MSE	MAE
X and Y	0.9849	29.29	76.49	5.00
X and X_1	0.7277	13.96	2614.77	38.64
X and X_2	0.5243	11.68	4415.87	55.13
X and X_3	0.5532	10.56	5721.63	69.37
X_2 and X_3	0.8631	16.97	1305.77	14.23

Figure 8 shows the grayscale thermal panel image ‘dji-0008.jpg’ of size 256×336 pixels in part (a) which was enhanced by using the histogram equalization and Fourier transform-based alpha-rooting method [35]. The histogram equalization is shown in part (b). The images of 0.70- and 0.76-rooting are given in parts (c) and (d), respectively. The method of α -rooting by the Fourier transform of the image $f_{n,m}$ is described as follows:

$$f_{n,m} \rightarrow 2D \text{ DFT} \rightarrow F_{ps} = |F_{ps}| e^{i\varphi(n,m)} \rightarrow |F_{ps}|^\alpha e^{i\varphi(n,m)} \rightarrow \text{inverse } 2D \text{ DFT} \rightarrow g_{n,m}. \quad (10)$$

Here, $\alpha \in (0,1)$ and F_{ps} is the 2D discrete Fourier transform (DFT) calculated by

$$F_{ps} = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f_{n,m} \exp\left(-\frac{2\pi i}{N} np\right) \exp\left(-\frac{2\pi i}{M} ms\right), p, s = 0: (N-1), (M-1). \quad (11)$$

The inverse transform is calculated by

$$f_{n,m} = \frac{1}{NM} \sum_{p=0}^{N-1} \sum_{s=0}^{M-1} F_{ps} \exp\left(\frac{2\pi i}{N} np\right) \exp\left(\frac{2\pi i}{M} ms\right), n, m = 0: (N-1), (M-1). \quad (12)$$

In the α -rooting, the module of the transform is processed exponentially $|F_{ps}| \rightarrow |F_{ps}|^\alpha$, and the phase function is preserved.

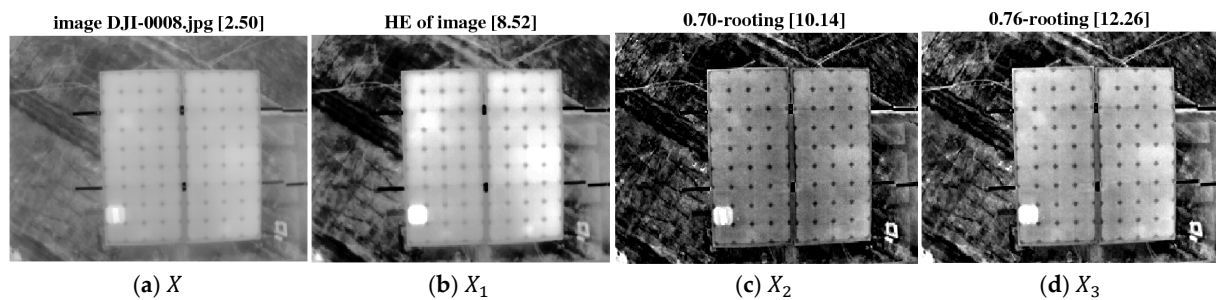


Figure 8. (a) The original image, (b) the histogram equalization, and the images enhanced by (c) 0.70-rooting and (d) 0.76-rooting.

The numbers in the titles of these images show the corresponding values of the enhancement measure EME which was calculated by using the blocks of size 3×5 . The EME measure of the image $f_{n,m}$ is calculated by averaging the max/min ratios of the image values inside the blocks W_k on which the image is divided. Thus, the following formula is used for this measure [35,36]:

$$EME(f) = \frac{1}{K} \sum_{k=1}^K \frac{\max_{n,m \in W_k} f_{n,m}}{\min_{n,m \in W_k} f_{n,m} + c} \quad (13)$$

The constant c can be considered equal to 1. This measure has been used for image enhancement together with many other measures (which are discussed in detail in Section 6).

Here, we want to see how the EME measure correlates with scores from the above mentioned metrics. The original image of low quality with $EME(X) = 2.50$, and for enhanced images $EME(X_1) = 8.52$, $EME(X_2) = 10.14$, and $EME(X_3) = 12.26$. At point $\alpha = 0.76$ the EME function as the function of α has the maximum value, as shown in Figure 9. Therefore, we can consider the 0.76-rooting image, X_3 , as a reference image.

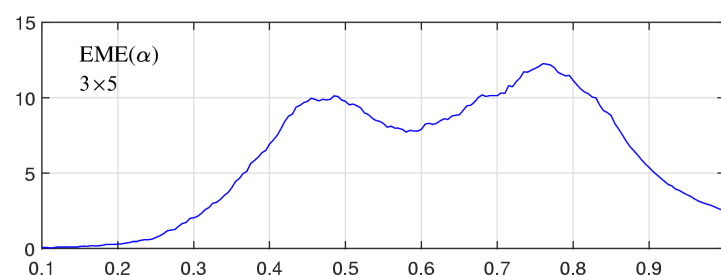


Figure 9. The graph of the EME function for image enhancement by the α -rooting.

Next, we analyze the above similarity metric for the images shown in Figure 8. Table 7 shows the metrics between all these images. According to these metrics, the greatest similarity is observed between two images which are processed by using the 0.70-rooting and the histogram equalization method, that is, X_3 and X_1 . For these two images, the PSNR is higher and MSE and MAE are smaller, than for other images. However, our personal perception suggests a strong similarity between images of 0.70-rooting and 0.76-rooting with high values of EME.

Table 7. Four metrics for the thermal panel images.

Image (from Figure 8)	SSIM	PSNR	MSE	MAE	Image	EME
X_3 and X_2	0.9202	17.01	1294.80	32.44	X_3	12.26
X_3 and X_1	0.9241	18.37	945.43	24.49	X_2	10.14
X_3 and X	0.7626	14.90	2105.05	35.71	X_1	8.52
X_2 and X_1	0.7785	12.76	3441.68	50.00	X	2.50

X_2 and X	0.6979	11.81	4284.79	59.52
X_1 and X	0.8093	16.49	1458.64	30.47

Figure 10 illustrates how the MAE responds to outliers in a regression scenario. The MAE metric is frequently used to evaluate the performance of regression algorithms, including linear regression, decision trees, and deep learning models. While both MAE and MSE have their merits and limitations, the choice between them often depends on the nature of the task and the importance of penalizing large errors.

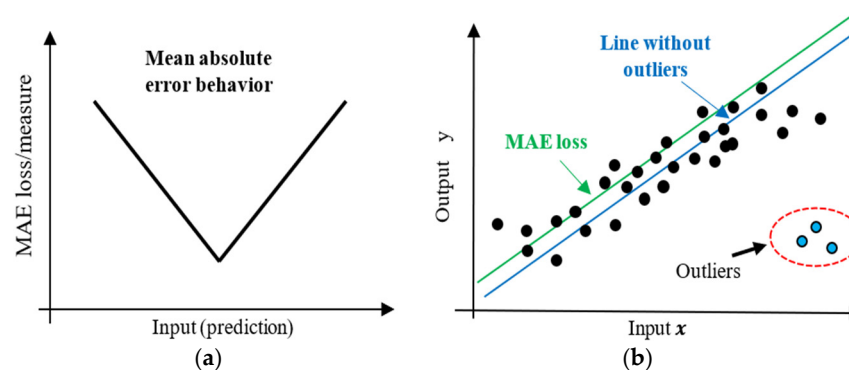


Figure 10. (a) MAE loss function, and (b) the data with outliers and MAE loss function.

Table 8 summarizes the main features of the MAE, MSE, and RMSLE metrics.

Table 8. The key characteristics of the MAE, MSE, and RMSLE measures.

Metric	Formula	Interpretability	Main Feature	Common Use Cases
MAE	$\frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} X_{i,j} - Y_{i,j} $	Low	Treats all deviations linearly	High (same units)
MSE	$\frac{1}{NM} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (X_{i,j} - Y_{i,j})^2$	Moderate	Amplifies larger errors; susceptible to outliers	Regression with large error penalties
RMSLE	$\sqrt{\frac{1}{n} \sum_{i=1}^n ((\ln(1 + p_i) - \ln(1 + \text{Low}))^2)}$ (log space)	Low (log space)	Compresses differences, downplaying both large and small errors while preserving their relative scale	Growth prediction, relative error evaluation

A. Histogram-based similarity is a technique in computer vision and data analysis that measures how alike two datasets are, often involving image color, intensity, or statistical distributions, by comparing their histograms. Comparison of histograms of two images, H_1 and H_2 (e.g., intersection, chi-square distance, see below definition), is invariant to small spatial changes but fails when spatial composition differs significantly. The similarity is calculated by using the following concepts.

Histogram intersection:

$$\text{Intersection}(H_1, H_2) = \sum_{i=0}^{L-1} \min(H_1(i), H_2(i)), \quad (14)$$

Chi-square distance:

$$\chi^2(H_1, H_2) = \sum_{i=0}^{L-1} \frac{(H_1(i) - H_2(i))^2}{H_1(i) + H_2(i)}. \quad (15)$$

This is a non-parametric, non-linear distance metric commonly used to compare histograms, probability distributions, or categorical data. It quantifies the difference between observed and expected frequencies, assigning greater importance to less expected values. In machine learning and computer vision, it is often employed to compare feature histograms, like color histograms, and performs well with sparse data.

In general, histogram-based similarity metrics offer numerous advantages, including low computational cost, which makes them ideal for large databases. Because histograms represent feature distributions rather than precise positions, they are robust against changes in object placement or orientation. These metrics are particularly effective at detecting shifts in overall image composition or lighting conditions. Additionally, histograms consume significantly less storage compared to raw pixel data. They are adaptable and can be applied across various color spaces, such as RGB, HSV, and CIELAB, and can be combined with other features, such as texture, to enhance accuracy.

2.3. The Huber Loss Function

The Huber loss function is a widely used metric that combines the strengths of the MSE and MAE. It behaves like the MSE for small errors, promoting smooth convergence, and like the MAE for larger errors, enhancing robustness against outliers. This hybrid approach makes Huber loss particularly suitable for regression problems involving noisy or anomalous data. It lessens MSE's sensitivity to outliers while remaining differentiable, facilitating gradient-based optimization [6].

Definition 7: The Huber loss function is the difference between the true value y and the predicted value $f(x)$ and is calculated by

$$L_{\delta}(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2, & \text{for } |y - f(x)| \leq \delta \\ \delta(|y - f(x)| - \frac{1}{2}\delta), & \text{for } |y - f(x)| > \delta \end{cases} \quad (16)$$

Here, δ (delta) is a threshold parameter that determines the point at which the function transitions from a quadratic to a linear behavior. This parameter is adjusted based on the specific data and error distribution.

- A small δ (e.g., 0.1) is desirable for datasets with heavy-tailed or skewed error distributions. This setting improves robustness to outliers;
- A large δ (e.g., 10.0) makes the loss function more sensitive to outliers but better suited for data with normally distributed errors.

Figure 11 shows the Huber loss function in part (a) and illustrates the relationship between outliers and the MAE and MSE losses in part (b). The graph of the Huber loss function versus prediction error demonstrates a smooth and continuous curve without sharp corners. This property ensures that the function is not only continuous but also differentiable everywhere, which is critical for gradient-based optimization algorithms.

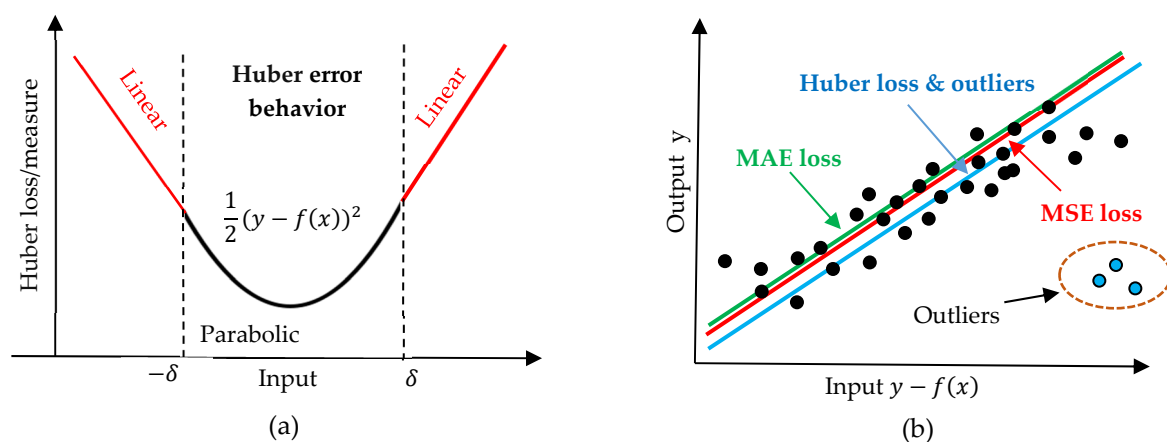


Figure 11. (a) The Huber loss function and (b) data with outliers and the MAE and MSE loss functions.

Although its piecewise structure makes the Huber loss slightly more complex than the MAE or MSE, it provides notable benefits in terms of training stability and model robustness. The Huber is commonly used in various machine learning and computer vision tasks, especially those involving noisy measurements or heavy-tailed error distributions. Typical applications include object detection, robust optimization, and time series forecasting, where accuracy and noise resistance are both critical.

3. The Euclidean and Chebyshev Distances

In this section, we describe different distances that are used in multi-dimensional spaces. These distances are well-known functions in the metric space, which play an important role in mathematics [37]. We consider an n -dimensional metric space X of vectors, or n -tuples, $x = (x_0, x_1, \dots, x_{n-1})$ with real numbers, when $n \geq 1$.

Definition 8: A metric space is a set X and a single-valued non-negative function $d(x, y)$ which is called a distance and satisfies the following conditions:

- (1) $d(x, y) = 0$, only if $x = y$,
- (2) $d(x, y) = d(y, x)$, where $x, y \in X$,
- (3) $d(x, y) + d(y, z) \geq d(x, z)$, for $x, y, z \in X$.

Consider n -dimensional spaces, or n -spaces, with the following examples:

- (a) The Euclidean space with the distance

$$d(x, y) = d_2(x, y) = \sqrt{\sum_{i=0}^{n-1} (x_i - y_i)^2}. \quad (17)$$

- (b) For $p \geq 1$, the distance (the Minkowski distance)

$$d(x, y) = d_p(x, y) = \left(\sum_{i=0}^{n-1} |x_i - y_i|^p \right)^{1/p}. \quad (18)$$

- (c) The Euclidean distance is the Minkowski distance for the $p = 2$ case. Another particular case, when $p = 1$, refers to the Manhattan distance,

$$d(x, y) = d_1(x, y) = \sum_{i=0}^{n-1} |x_i - y_i|. \quad (19)$$

- (d) The distance, also called the Chebyshev distance [7],

$$d(x, y) = d_\infty(x, y) = \{|y_i - x_i|; i = 0, 1, \dots, n - 1\}. \quad (20)$$

It is the limiting case of the Minkowski distance, namely, $d_\infty(x, y) = d_p(x, y)$.

- A. The Chebyshev distance measures the largest difference between components of the vectors. It is also known as the maximum metric. Imagine comparing two paintings and focusing only on the single brushstroke that shows the greatest difference. Unlike other metrics that add or average differences, the Chebyshev distance considers only the maximum discrepancy. This makes it especially useful in cases where even a single large difference is critical, such as in quality control, security, or anomaly detection [7].

Let us calculate the Chebyshev distance between the above images x and y and between x and z ,

$$\begin{aligned} d_\infty(x, y) &= \{|10 - 12|, |20 - 22|, |30 - 31|\} = \{2, 2, 1\} = 2, d_\infty(x, z) \\ &= \{|10 - 50|, |20 - 60|, |30 - 70|\} = \{40, 40, 30\} = 40. \end{aligned}$$

Table 9 provides an overview of the Chebyshev distance, highlighting its strengths and limitations for various use cases.

Table 9. The main advantages and disadvantages of the Chebyshev distance.

Category	Advantages	Category	Disadvantages
Computation	Simple to calculate using only comparisons and max operations.	Lacks sensitivity	It may not be accurate when all dimensions contribute equally.
Applicability	Ideal for grid-based systems (e.g., chess, robotics).	Limited application	Not suitable outside grid-based or constrained movement systems.
Dimension focus	Captures dominant, extreme variations.		Ignore smaller cumulative differences.
Scaling	Works well with normalized data.		Requires preprocessing for unnormalized data.
Smoothness	Produces clear, discrete results.	Not smooth	Unsuitable for gradient-based optimization.
Special applications	Useful for edge detection, pattern recognition, and pixel-based image operations.		Limited flexibility in continuous or non-grid systems.

Figure 12 shows the point $x = (1,2,2)$ in a 3D space and the distances between this point and the original point $y = (0,0,0)$, which are calculated by the above metrics d_1 , d_2 , and d_∞ in part (a). The graphs of the metric $M(p) = d_p(x, y)$, as the function of p , is shown in part (b).

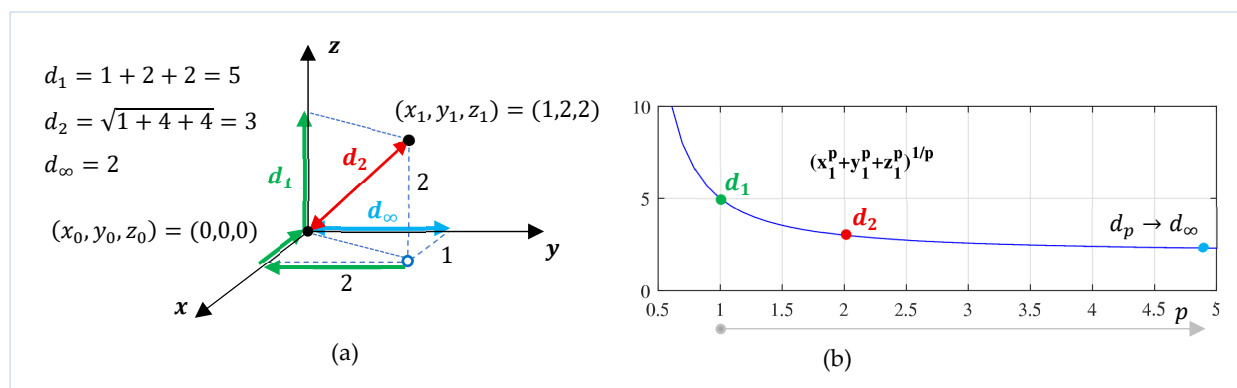


Figure 12. (a) Three-dimensional space with distances between two points $(0,0,0)$ and $(1,2,2)$, and (b) the function $M(p)$.

Figure 13 illustrates the shapes created by the Euclidean, Manhattan, and Chebyshev distances in a 2D space. Chebyshev distance provides AI with a precise focus on the worst-case difference. It prioritizes the most significant deviation over overall similarity, making it useful for applications that require high accuracy, safety, or threshold limits. Nonetheless, it is less suitable for general similarity tasks where total differences are more important. This helps demonstrate how AI systems “measure” similarity in multidimensional data.

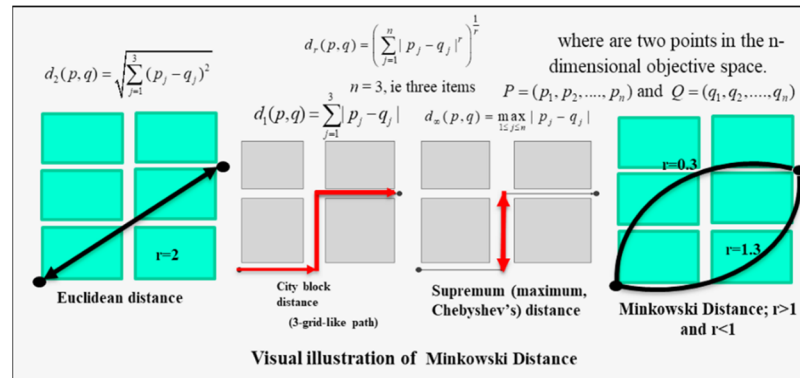


Figure 13. The Euclidean, Manhattan, Chebyshev, and Minkowski distances between two points.

B. *Cosine Similarity [8]: Measuring Aligned Patterns.* In the n -dimensional space with the metric (or the distance $d(x, y)$) the concept of the angle φ between two vectors x and y is defined similarly to the traditional 3D Euclidean space. The cosine of the angle φ is calculated by the dot product of vectors, $x \cdot y = x_0y_0 + x_1y_1 + \dots + x_{n-1}y_{n-1}$ as

$$S(x, y) = \cos \varphi = \frac{x \cdot y}{\sqrt{x \cdot x} \sqrt{y \cdot y}} = \frac{x_0y_0 + x_1y_1 + \dots + x_{n-1}y_{n-1}}{\sqrt{x_0^2 + x_1^2 + \dots + x_{n-1}^2} \sqrt{y_0^2 + y_1^2 + \dots + y_{n-1}^2}}. \quad (21)$$

Here, $\sqrt{x \cdot x}$ and $\sqrt{y \cdot y}$ are denoted as $\|x\|$ and $\|y\|$, respectively, and are called the norms, or magnitudes of the vectors.

The cosine similarity is scale-invariant; doubling the values in one vector does not change its similarity to another because the angle between them remains the same. It is essential to note that the cosine function measures the similarity of two vectors, not the distance between them. This function does not satisfy the triangle inequality in Definition 7 and is, therefore, not a distance.

Example 2: Consider the 4 D vector $x = (1, 2, 3, 5)$ and the noisy vector $y = x + (0, 0.1, -0.1, 0) = (1, 2.1, 2.9, 5)$. The cosine of the angle φ between these two vectors is calculated as follows:

$$S(x, y) = \cos \varphi = \frac{38.9}{\sqrt{39} \sqrt{38.82}} = \frac{1513.21}{1513.98} \approx 0.99949, \varphi = 1.827^\circ.$$

These vectors are not perfectly aligned, $S(x, y) \approx 0.99949 \neq 1$, but close to this; they are nearly identical vectors.

The cosine similarity metric is widely used in various areas, including natural language processing for comparing text documents, recommendation systems to identify similar user preferences, and clustering to group similar data points. Let us see how AI systems compare these images using two different methods: the Euclidean distance, which measures absolute differences, and the cosine similarity, which measures directional similarity.

Example 3: Consider the cosine similarity between the images $x = (10, 20, 30)$, $y = (12, 22, 31)$, and $z = (50, 60, 70)$ from Example 2. The calculation result is as follows :

$S(x, y) = 0.9989$	$S(x, z) = 0.9684$
$d_0(x, y) = 2$	$d_0(x, z) = 40$

Both results are relatively high, showing that x , y , and z share nearly identical patterns despite their brightness differences. The cosine similarity effectively recognizes that the image z is just a brighter version of the image x . This makes it extremely useful in situations where illumination changes, but the underlying structure remains constant, such as in facial recognition, where a person should be recognized in both shadowed and well-lit conditions.

C. Understanding the Difference

These two approaches show how AI can interpret images differently based on its goal. The Euclidean distance measures image similarity by considering absolute pixel values, making it sensitive to changes in brightness and exposure. The cosine similarity, on the other hand, assesses how closely the images follow the same overall pattern, ignoring their size differences. Imagine a photo of a cat, which was taken in dim light (the image x) and another in bright sunlight (the image y). In a Euclidean distance system, these photos look quite different because their pixel intensities vary. But with cosine similarity, they would appear remarkably similar; the structure, shape, and features of the cat remain the same. In real-world AI systems, both methods have their advantages. For example, the Euclidean distance can help detect changes in brightness or exposure in medical imaging. At the same time, the cosine similarity is effective for face recognition, text comparison, and pattern detection, where structure is more important than size. Ultimately, the choice between the distance and similarity depends on whether the AI needs to identify precise numeric differences or focus on consistent relational patterns.

Our method of measuring distances significantly influences an AI system's behavior, explaining why AI performance varies across applications. Techniques effective for medical image comparison may not be suitable for family photographs or satellite imagery. Developers must evaluate various approaches to identify the most appropriate one, considering data types and project objectives. Larger, detailed images require more computational power, which can cause devices like phones to heat up during processing. The curse of dimensionality, previously discussed, also indicates that simple distance comparisons become less dependable with high-resolution, complex data. To address this, modern AI often integrates basic distance measures with advanced deep learning methods better suited for complex data. Understanding the Minkowski distance clarifies how AI compares, classifies, and groups data, underpinning common systems such as photo management, music recommendations, and medical diagnostics. While straightforward, its adaptability makes it a vital tool for AI development.

Modern AI systems often combine these approaches, creating sophisticated models that can adapt their "vision" to different tasks. They might use distance-based methods for precise technical measurements while simultaneously employing similarity-based approaches for pattern recognition, creating a more nuanced and human-like understanding of visual information. As AI continues to advance, these basic ideas of similarity and distance remain essential, providing the mathematical foundation that enables machines to interpret and navigate our visual environment with greater sophistication and reliability. Whether you are interested in how your devices operate or in how AI might influence your career, understanding these key ideas gives important insight into the systems that drive our increasingly AI-integrated world.

4. Subjective and Objective Image Quality Assessment

Image quality assessment has emerged as a crucial element in image processing and computer vision, focusing on evaluating image quality to ensure it meets human perceptual expectations. Contrast, one of the most essential psycho-physical factors influencing human vision, plays a dual role in this context. In machine-based applications, contrast is a measurable physical quantity derived from the captured optical signal or the numerical representation of the visual signal. Various factors can affect image quality, including compression, transmission, display, and acquisition processes. Therefore, precise image quality measurement has become essential in many image-based applications. Image quality assessment uses two primary methodologies: subjective and objective.

Subjective image quality assessment (SIQA) relies on human observers' opinions of image quality. This process typically involves showing participants a set of images, gathering ratings based on predefined scales (e.g., a 5-point scale from "bad" to "excellent"), collecting comparative judgments through pairwise comparisons, and

averaging individual ratings to compute the mean opinion score (MOS). The key aspects evaluated in the subjective assessment include the following:

- Naturalness (the realistic appearance of enhanced images);
- Detail preservation (the retention of fine details, textures, and edges); and
- Visual appeal (the overall aesthetic improvement).

While SIQA is regarded the gold standard due to its direct link to human perception, it encounters several challenges: intrinsic subjectivity shaped by individual preferences and cultural backgrounds, variability stemming from viewing conditions, display devices, and lighting, a time-consuming and expensive implementation process, limited scalability for large datasets, and inconsistency in interpreting quality categories among non-specialist observers.

Objective image quality assessment uses computational models to automatically predict human perception of image quality. These models aim to mimic the human visual system and provide consistent, repeatable quality scores without relying on human observers. They typically extract relevant features from images, such as edges, textures, and color information, and analyze these to quantify various aspects of image quality, including sharpness, contrast, noise, and artifacts. They are quicker and less resource-intensive than subjective methods, more suitable for real-world applications due to their efficiency, and provide more standardized and reproducible results. Additionally, they are essential for a wide range of applications, including the following [10,11,38,39]: guiding the design of transformation functions, benchmarking image enhancement algorithms, optimizing image processing and transmission systems, and monitoring image quality in real-time applications.

Objective IQA methods are broadly categorized into three types based on the availability of a reference image:

Full-Reference (FR-IQA): FR-IQA methods require a pristine, undistorted reference image for comparison, making them suitable for applications where the original image is available, such as image restoration, denoising, and quality assessment of image reproduction systems. Popular FR-IQA metrics include PSNR, SSIM, and visual information fidelity (VIF) [4,8–12,34].

Reduced-Reference (RR-IQA): RR-IQA methods use partial information extracted from the reference image, such as features or statistical properties, to assess the quality of the distorted image [11]. This makes them suitable for scenarios where transmitting the full reference image is impractical due to bandwidth constraints or storage limitations. RR-IQA methods are often used in image and video communication systems, where only a limited amount of side information can be transmitted along with the compressed image data.

No-Reference (NR-IQA): NR-IQA methods, also called blind IQA, assess image quality without any reference image, making them the most challenging category of IQA methods [35,36,38–41]. NR-IQA is crucial for applications where a reference image is absent, such as images captured in uncontrolled environments, user-generated content analysis, and the quality assessment of historical or archival images. NR-IQA methods frequently rely on statistical models, natural scene statistics, and machine learning techniques to infer image quality from features extracted from the distorted image. Ongoing research in IQA, particularly in NR-IQA, focuses on developing more sophisticated models that better align with human perception and integrate machine learning techniques, continuously advancing the field and improving the visual quality of images across various industries, enhancing the visual experience for users across different applications and platforms.

Over the years, researchers have developed numerous IQA methods using traditional and deep learning approaches to achieve this goal, making IQA an expanding field of research. A notable innovation in this area is the entropy-based enhancement measure (EME) and its extended version, EMEE, introduced by Agaian in 1999 [36]. These

metrics combine Weber's law, a model of human perception related to just-noticeable differences in stimulus intensity, and Shannon entropy, a measure of information content, to evaluate the effectiveness of image enhancement algorithms. They assess the improvement in local contrast and information content achieved by an enhancement algorithm, providing valuable insights into its performance.

This no-reference metric evaluates image enhancement quality by combining Weber's law of local contrast with Shannon entropy, yielding an objective measure of enhancement effectiveness. The EMEE metric has proven particularly valuable for assessing the performance of image enhancement algorithms and has significantly advanced objective image quality assessment methods.

5. Perceptual Basis and Laws of Vision “Instead Of” Weber's and Mickelson's Laws

Weber's law describes the human visual system's sensitivity to changes in stimulus intensity. In image processing, it explains how we perceive local contrast. When a visual target is presented against a uniform background, its contrast is defined as the relative difference in luminance between the target and the background:

$$C = C(W) = \frac{\Delta_{n,m \in W} f_{n,m}}{\text{background } f_{n,m}} = \frac{\max_{n,m \in W} f_{n,m} - \min_{n,m \in W} f_{n,m}}{\text{background } f_{n,m}}, \quad (22)$$

considering the maximum, minimum, and background luminance of the image $f_{n,m}$ in a small, considered window W . The background is calculated as the mean of the block W . The Weber–Fechner law determines the value of Δf , known as the just noticeable difference, at which the target object becomes just visible. An excellent survey commonly used in this context is presented here. This mathematical formula represents the minimum threshold (through the constant c) and the relationship between the initial amount and our ability to perceive changes.

Michelson contrast of an image $f_{n,m}$ is calculated by

$$MC = MC(W) = \frac{\max_{n,m \in W} f_{n,m} - \min_{n,m \in W} f_{n,m}}{\max_{n,m \in W} f_{n,m} + \min_{n,m \in W} f_{n,m}}. \quad (23)$$

The existing definitions of contrast can be broadly divided into two main categories: global and local contrast measures. Global contrast measures are based on an image's global characteristics (e.g., the maximum and minimum image values). A global contrast measure is usually defined as the ratio of the darkest to the brightest pixel in an image. The Weber-Fechner and Michelson contrasts are the first global definitions widely used in many applications, such as CE, IQA, and quantization. These can be computed using spatial, frequency, and multi-resolution or multi-scale representations.

Several alternative formulations for extended Weber contrast were proposed. These formulations better match human visual perception and offer improved computational efficiency. The proposed contrast measures of the image $f_{n,m}$ in a small window W are calculated as

$$AC_1 = \ln \frac{f_{\max}}{f_{\min} + c}, AC_2 = \left(\frac{f_{\max} - a f_{\min}}{f_{\min} + c} \right)^\alpha, AC_3 = \left(\frac{f_{\max}}{f_{\min} + c} \right)^\alpha AC_1, \quad (24)$$

$$AC_4 = \frac{\ln(c + \ln(f_{\max} + c))}{\ln(c + \ln(f_{\min} + c))}, AC_5 = \left(\frac{f_{\max} - 2f_{n,m} + f_{\min}}{f_{\max} + 2f_{n,m} + f_{\min} + c} \right)^\alpha. \quad (25)$$

Here, $f_{\max} = f_{n,m}$ and $f_{\min} = f_{n,m}$, and c is a small positive constant (often close to 1) that prevents division by zero and stabilizes the logarithm, a is a constant, typically 0 or 1, $f_{n,m}$ image intensity at pixel (n, m) , and α is a positive number to adjust the contrast enhancement. These extended formulations provide more flexibility and robustness in practical image-processing scenarios [38–40].

The concept of the local contrast is also used for color images. We consider the RGB color model, wherein the image $f_{n,m} = (r_{n,m}, g_{n,m}, b_{n,m})$ is presented by three primary colors; red, green, and blue. The local contrast for color image is typically defined using a gray-level contrast applied to each color channel. Common approaches include [39]

$$CC1 = \frac{(r_{\max}(n, m), g_{\max}(n, m), b_{\max}(n, m))}{(r_{\max}(n, m), g_{\max}(n, m), b_{\max}(n, m)) + c} \text{ and } CC2 = (CC1) \ln(CC1). \quad (26)$$

Here, $(r_{\max}(n, m), g_{\max}(n, m), b_{\max}(n, m))$ represents the maximums of the prime colors of the image within a local block $W = W(n, m)$ centered at pixel (n, m) .

5.1. Preliminary Experimental Results

In this section, we describe a few examples of grayscale images and analyze the above measures. Figure 14 shows the grayscale thermal panel image ‘dji-0769.jpg’ of size of 256×336 pixels in part (a). The histogram equalization of this image is shown in part (b). The images of 0.80- and 0.75-rooting by the Fourier transform are given in parts (c) and (d), respectively.

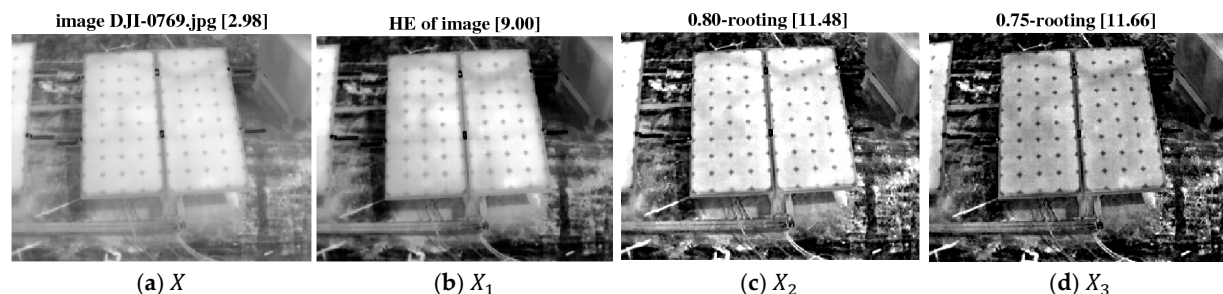


Figure 14. (a) The original image, (b) the histogram equalization, and the images enhanced by (c) 0.80-rooting and (d) 0.75-rooting.

The original image of low quality with EME measure $EME(X) = 2.98$, and for HE enhanced image $EME(X_1) = 9.00$. For the alpha-rooting enhancement, $EME(X_2) = 11.48$, and $EME(X_3) = 11.66$. At point $\alpha = 0.75$ the EME function as the function of α has the maximum value. All measures were calculated by using the blocks of size 3×5 . Together with the EME measure, we consider the above contrast measures. The values of these measures for the above four images have been calculated with parameters $c = a = \alpha = 1$ and are given in Table 10. One can note that the alpha-rooting image X_3 not only has the highest EME value but also surpasses other images in all other metrics.

Table 10. Eight measurements for thermal panel images ‘dji-0769.jpg’.

Image (from Figure 14)	EME	C	MC	AC1	AC2	AC3	AC4	AC5
original, X	2.98	0.1482	0.0738	0.7700	0.2091	1.0967	1.0148	0.0021
HE, X_1	9.0	0.5548	0.2323	1.0411	2.3215	8.4724	1.1217	0.0314
0.80-rooting, X_2	11.48	1.0545	0.3813	1.5649	15.5444	71.6057	1.3861	0.0720
0.75-rooting, X_3	11.66	1.5614	0.4820	1.9195	23.7760	109.1911	1.6071	0.1203

The 0.75-rooting image can be used as the reference image for images enhanced from the original image of low quality. The similarity metrics between the images are given in Table 11.

Table 11. SSIM table for the images in Figure 14.

	X	X_1	X_2	X_3
X	1	0.8894	0.8082	0.7776
X_1	0.8894	1	0.9339	0.9081
X_2	0.8082	0.9339	1	0.9358

X_3	0.7776	0.9081	0.9358	1
-------	--------	--------	--------	---

We also consider the above metrics for the image of size 510×661 pixels, which is shown in Figure 15 in part (a). The images processed by the HE and 0.85-rooting are shown in parts (b) and (c), respectively. The method of 0.85-rooting was also used on the negative image and is shown in part (d). In other words, the image was processed as follows: $X \rightarrow (Y = 255 - X) \rightarrow (Y' = 0.85\text{-rooting of } Y) \rightarrow X_3 = 255 - Y'$.

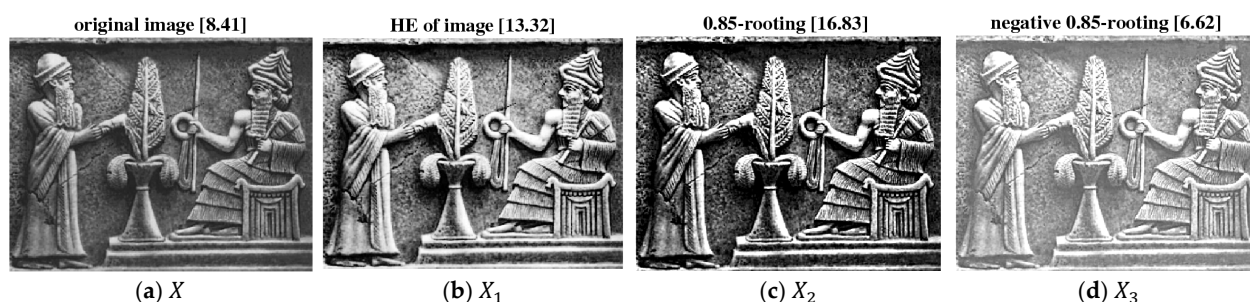


Figure 15. (a) The original image, (b) the histogram equalization, and the images enhanced by (c) 0.85-rooting and (d) 0.85-rooting.

The EME measure and other measures for all these four images are given in Table 12.

Table 12. Eight measurements for the images in Figure 15.

Image (from Figure 15)	EME	C	MC	AC1	AC2	AC3	AC4	AC5
The original, X	8.41	0.4112	0.2016	0.9265	0.5948	1.5816	1.0458	0.0097
HE, X_1	13.32	0.7923	0.3594	1.3465	7.8868	34.4506	1.2347	0.0367
0.85-rooting, X_2	16.83	1.7231	0.5969	2.1566	30.5462	146.9267	1.6803	0.1296
neg.-0.85-rooting, X_3	6.62	0.3226	0.1609	0.8749	0.4432	1.3501	1.0309	0.0024

The 0.85-rooting image has the highest values of all eight measurements. Therefore, this image can be used as the reference image for images enhanced from the original image of low quality. The similarity metrics between the images are given in Table 13. One can note that the highest similarity is between images X_1 and X_2 .

Table 13. SSIM table for the images in Figure 15.

	X	X_1	X_2	X_3
X	1	0.8432	0.7820	0.8528
X_1	0.8432	1	0.8832	0.8296
X_2	0.7820	0.8832	1	0.6911
X_3	0.8528	0.8296	0.6911	1

Figure 16 shows the 512×512 -point image 'Goldhill512.bmp' image in part (a) and HE in part (b). The image of the 0.89-rooting enhancement is shown in part (c). The method of 0.79-rooting on the negative image is shown in part (d); the image was processed as follows:

$$X \rightarrow (Y = 255 - X) \rightarrow (Y' = 0.79\text{-rooting of } Y) \rightarrow X_3 = 255 - Y'.$$



Figure 16. (a) The original image ‘Goldhill512.bmp’ 512 × 512, (b) the histogram equalization, and the images enhanced by (c) the 0.85-rooting and (d) the negative 0.85-rooting.

All eight measurements of these four images are given in Table 14. Two 0.85-rooting images have the highest scores in all eight measurements.

Table 14. Eight metrics for ‘Goldhill’ images.

Image (from Figure 16)	EME	C	MC	AC1	AC2	AC3	AC4	AC5
The original, X	6.60	0.3281	0.1605	0.8710	0.4310	1.2947	1.0359	0.0077
HE, X_1	13.16	0.7405	0.3175	1.1557	2.0989	6.0538	1.1327	0.0386
0.85-rooting, X_2	15.77	1.1098	0.4396	1.6168	13.6167	60.9416	1.3659	0.0689
neg.-0.85-rooting, X_3	27.52	1.0086	0.4635	1.6203	8.8506	34.0889	1.2396	0.0462

5.2. Perceptual Justification of Weber’s Law and Its Extensions

Weber’s law, a fundamental principle of psychophysics, states that the perceived change in a stimulus is proportional to its initial intensity. This is expressed mathematically as the Weber fraction $k = \Delta f / f$, where Δf is the change in intensity and f is the initial intensity. This law explains how our visual system responds to changes in brightness and why specific visual enhancements are more perceptible than others.

The core idea behind Weber’s law and its extensions, such as those proposed by Agaian, is that the absolute change in luminance does not determine the perceived change in brightness but rather the relative change compared to the background luminance (the $f_{max} / (f_{min} + c)$ or on a logarithmic scale $\ln[f_{max} / (f_{min} + c)]$ [36]. This principle has significant implications for image enhancement, as it aligns with the way the human visual system processes brightness variations. For example, consider two scenarios: a slight increase in brightness in a dark area and the same absolute increase in brightness in a very bright area. The change in the dark area will be much more noticeable to the human eye than in the bright area. This phenomenon is precisely what Weber’s law captures: our visual system is more sensitive to proportional changes than absolute ones. Also, this perceptual characteristic means that brighter pixels generally need a larger absolute change in luminance to be perceived differently from darker pixels. As a result, brighter regions require more aggressive enhancement than darker ones to achieve a visually uniform enhancement across an image. Understanding this principle enables the design of more effective image enhancement techniques that align with human perception.

Advantages of Agaian’s contrast measures given in Equations (24) and (25):

1. (*Emphasis on Edge Information*) The methods often focus on local contrast measurements, vital for highlighting edges and other visually significant features. Our visual system is highly attuned to contrast variations, especially at the edges, as these often convey crucial information about object boundaries and shapes. This improved edge definition is essential for accurate shape perception and object recognition. As an example, Figure 17 shows a contrast image generated by the extended Weber law on 3×3 block images at varying scales. Each pixel value in an

image is replaced with the $f_{max}/(f_{min} + c)$ and $\ln[f_{max}/(f_{min} + c)]$ value of its neighbors (see Equation (24)).

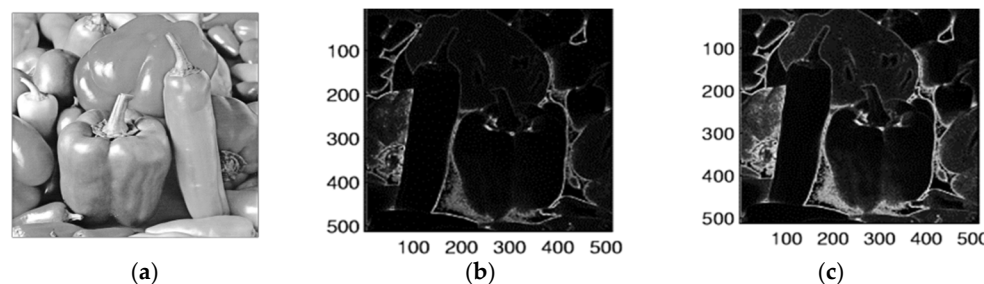


Figure 17. (a) The original image, (b) the contrast (visibility) image, and (c) the contrast image in a logarithm scale.

2. (*Accounting for Logarithmic Response*) The human visual system's response to light intensity is approximately logarithmic. The proposed contrast measures, particularly those involving logarithms, often account for this logarithmic response, leading to a more perceptually accurate contrast measure. This logarithmic transformation helps compress the wide range of light intensities in the real world into a range our visual system can effectively process. It also explains why ratios, rather than absolute differences, are so important in visual perception, as Weber's law highlights. This logarithmic scaling is crucial for accurately modeling how we perceive brightness differences.
3. (*Enhanced Texture and Pattern Perception*) By focusing on local contrast, these measures contribute to a better perception of texture and patterns within an image. These visual cues provide rich information about surfaces and materials, and their accurate representation is vital for image understanding. The localized nature of the proposed methods allows for a more sensitive analysis of how contrast variations contribute to the perception of these detailed surface characteristics.

5.3. Entropy for Images

As is known, the statistical definition of entropy was developed by Ludwig Boltzmann in the 1870s as the logarithmic measure for a system with n states, each of which has probability p_i of being occupied [42] :

$$S = -k_B \sum_{i=1}^n p_i \ln p_i, (k_B \text{ is the Boltzmann constant}). \quad (27)$$

In information theory, this fundamental concept is known as Shannon entropy (in 1948, [43]). It is a measure of the number of bits to be sent for a message composed from letters of the alphabet $A = \{a_0, a_1, \dots, a_{n-1}\}$:

$$H = - \sum_{i=0}^{n-1} p_i \log_2 p_i = \sum_{i=0}^{n-1} p_i \log_2 \frac{1}{p_i}. \quad (28)$$

It measures uncertainty associated with random variables, specifically quantifying the expected value of information contained in a message. For integer-valued images with scale 255, $n = 256$ and each intensity $i \in \{0, 1, \dots, 255\}$ is the state a_i , and p_i is the probability that a pixel has a gray level i .

A. Relationship between entropy and contrast

The contrast is the difference between light and dark areas in an image. Generally, a larger variance in gray intensities corresponds to higher contrast. While related, entropy and contrast are not directly interchangeable. An image with high contrast often has higher entropy, but this is not guaranteed. An image with minimal difference between its maximum and minimum intensities shows low contrast and likely low entropy. However,

creating images with high contrast but low entropy is possible, as is creating images with low contrast but high entropy. A higher entropy value, H , indicates more information (often, though not always, perceived as greater detail) in the image, reflecting greater uncertainty in pixel values.

Illustrative Example: Consider two chest X-ray images: one with clearly detailed lung structures and the other overexposed, appearing mostly homogeneous. The detailed image will have pixel values distributed across a broader range of gray levels, leading to higher entropy. For example, the probability distribution of pixel intensities might be $p(x_1) = 0.1$, $p(x_2) = 0.15$, $p(x_3) = 0.08$, and so on, resulting in a higher entropy of about 1.4 bits. The overexposed image, however, might have concentrated on most pixels in a narrow range of high-intensity values (e.g., $p(x_1) = 0.9$, $p(x_2) = 0.1$), yielding a much lower entropy 0.4690 bits. This demonstrates how entropy can quantify the difference in information content between two images.

B. Limitations of the entropy

Entropy has traditionally been used to assess image quality or complexity and identify potentially informative regions. However, a significant limitation of entropy is that it considers only the frequency of gray levels and completely ignores the spatial relationships between pixels. It does not account for structural information or local contrast variations crucial for human visual perception. As a result, two images with vastly different visual structures can have the same entropy value. Any rearrangement of the image does not change the entropy. Images might have identical entropy values despite appearing quite different to human observers. Figure 18 shows five images with very different structures that produce the same entropy value.



Figure 18. The images that can have identical entropy despite being perceptually distinct.

6. Non-Reference Image Quality Measures: EME and EMEE

The extended Weber entropy quality measure (EWEQ), known before as EMEE, was developed by Agaian to address the limitations of global entropy [38,39,44]. The EMEE is a non-reference image quality metric that measures improvements in local contrast and information content after image enhancement. Instead of calculating entropy over the entire image, it analyzes the image in small, localized blocks. This localized approach better reflects how the human visual system processes information. EMEE offers a more robust, perceptually relevant metric for evaluating image enhancement results by accounting for local variations. It combines the local entropy values to give an overall assessment of the image's quality and the effectiveness of the enhancement process. This allows EMEE to differentiate between images that may have similar global entropies but notably different local characteristics. Figure 19 shows the history of non-reference image quality measures:

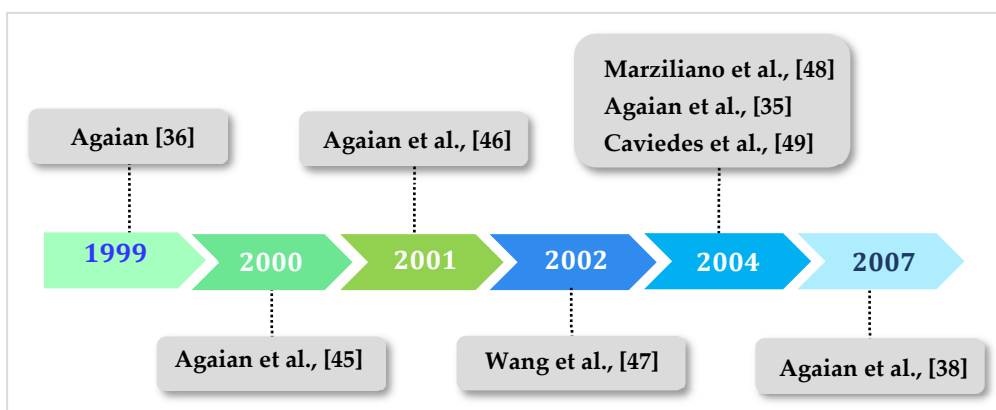


Figure 19. The earlier history of non-reference image quality measure development [35, 36, 38, 45–49].

We discuss the selected applications of contrast in computing and the representation of various forms of visual information. We also provide some guidelines for selecting the most appropriate contrast definition for each application or suggesting possible directions.

The EMEE is a non-reference image quality metric that measures the improvement in local contrast and information content after an image enhancement process. It examines the image in small, localized regions and combines the results. The EMEE is calculated as follows :

1. Dividing the image $f_{n,m}$ into blocks, or windows, W_k of $L_1 \times L_2$ pixel each, where $L_1, L_2 = 3, 5, 8, \text{ or } 16$.
2. Calculating the image contrast value in each window by combining an extended Weber law and the entropy,

$$AC_3(W_k) = \left(\frac{\max_{n,m \in W_k} f_{n,m}}{\min_{n,m \in W_k} f_{n,m} + c} \right)^\alpha \ln \frac{\max_{n,m \in W_k} f_{n,m}}{\min_{n,m \in W_k} f_{n,m} + c}, \quad (29)$$

Here,

- $W_k = W(x_k, y_k)$ are windows with centers (x_k, y_k) , $k = 1:K$, and K is the number of blocks in the image;
 - α is a constant that regulates the metric's sensitivity to changes in contrast. For instance, $\alpha = 0.98$;
 - c is a small constant added to prevent division by zero.
3. Averaging the $AC_3(W_k)$ values across all windows,

$$EMEE = \frac{1}{K} \sum_{k=1}^K AC_3(W_k) = \frac{1}{K} \sum_{k=1}^K \left(\frac{\max_{n,m \in W_k} f_{n,m}}{\min_{n,m \in W_k} f_{n,m} + c} \right)^\alpha \ln \frac{\max_{n,m \in W_k} f_{n,m}}{\min_{n,m \in W_k} f_{n,m} + c}.$$

This provides a global measure score for the entire image, based on local contrast information.

The EMEE provides a comprehensive evaluation by quantifying the extent to which the enhancement technique improves local contrast and overall information content. This metric has become an essential tool in image quality assessment. The effectiveness of the EMEE stems from its ability to analyze local image features, making it particularly valuable for applications that require detailed analysis of image structure and content. While the EMEE excels at capturing subtle details and textures that global metrics might overlook, it faces several significant limitations. The primary constraint lies in its exclusive focus on local features such as fine details, textures, and edges, potentially missing

broader image characteristics that contribute to perceived quality. While strong for detailed analysis, this localized approach may not fully capture the overall image quality that human observers perceive. Combining the EMEE with global image assessment (GIA) metrics has effectively addressed these limitations [50].

Agaian and Panetta Contrast Measures, AMEE

Panetta and Agaian developed various simple contrast enhancement evaluation (CEE) measures based on a global index computed from local measures of contrast and image gradient [36,44,50]. The Michelson and Weber–Fechner contrast measures mainly inspire these CEE measures. According to recent critical studies conducted on the performance analysis of CEE metrics, the most consistent metric related to the HVS, namely the absolute measure of enhancement by entropy (AMEE), is being discussed here.

The AMEE is a no-reference metric based on a contrast measure that uses the dynamic range (min-max) of pixel values within a block. The image $f_{n,m}$ on the rectangular lattice is divided into non-overlapping blocks $W_{i,j}$ of the same size. The local contrast at the pixel (i,j) is calculated by

$$C_{AMEE}(i,j) = \alpha \left(\frac{\max_{n,m \in W(i,j)} f_{n,m} - \min_{n,m \in W(i,j)} f_{n,m}}{\max_{n,m \in W(i,j)} f_{n,m} + \min_{n,m \in W(i,j)} f_{n,m} + \epsilon} \right)^\alpha \ln \frac{\max_{n,m \in W(i,j)} f_{n,m} - \min_{n,m \in W(i,j)} f_{n,m}}{\max_{n,m \in W(i,j)} f_{n,m} + \min_{n,m \in W(i,j)} f_{n,m} + \epsilon}. \quad (31)$$

Here, ϵ is the constant to control the division by zero, and α is an exponent that controls the enhancement effect. This local contrast is used to compute the global measure AMEE for CEE.

Figure 20 shows the image of size 750×440 pixels in part (a) with $EME = 6.55$ and $AMEE = 0.0288$. The histogram equalization of the image is shown in part (b). We also consider the α -rooting enhancement [34,35]. The images of the 0.74- and 0.84-rooting are shown in parts (c) and (d), respectively.

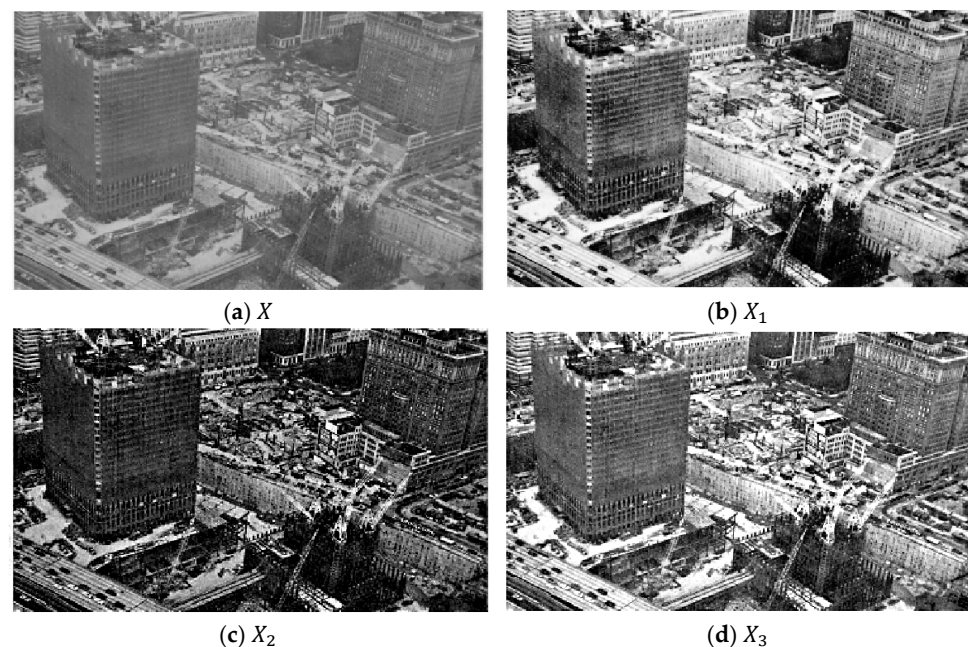


Figure 20. (a) The original image 750×440 , (b) the smooth image, (c) the histogram equalization, and the images enhanced by (c) 0.74-rooting and (d) 0.84-rooting.

The curves of the measurement functions EME and AMEE for the image are given in Figure 21. The maximum of EME is at point $\alpha = 0.84$ and for AMEE the maximum is at point $\alpha = 0.74$.

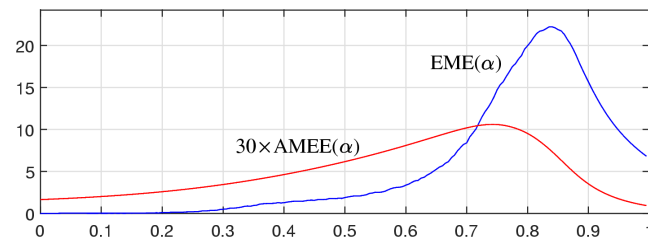


Figure 21. Graphs of the measures EME and AMEE.

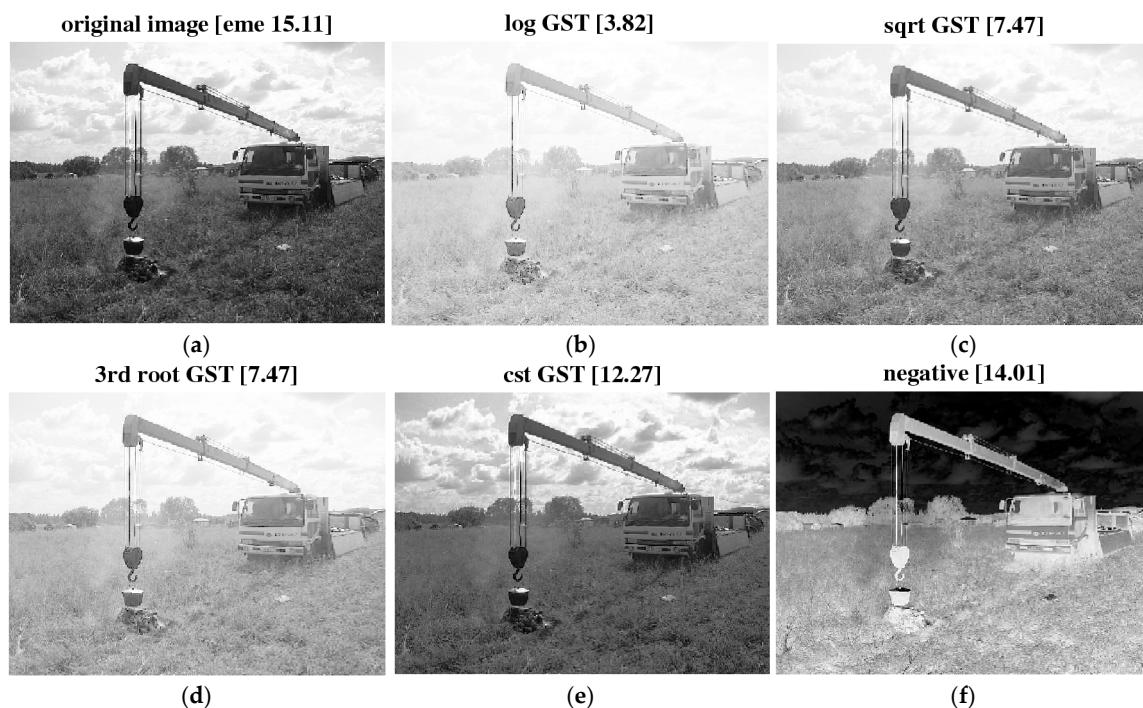
All measurements of these four images are given in Table 15. Two 0.85-rooting images have the highest scores in all eight measurements. One can note that except the EME values, the 0.74 and 0.84-rooting images have high scores for all measurements.

Table 15. Nine metrics for the images in Figure 20.

Image	EME	C	MC	AC1	AC2	AC3 *	AC4	AC5	AMEE
original, X	6.55	0.3244	0.1606	0.8688	0.4112	1.2527	1.0324	0.0054	0.0288
HE, X_1	24.78	1.0748	0.4801	1.5601	6.1438	21.3860	1.2150	0.0436	0.2222
0.74-rooting, X_2	13.90	2.7092	0.8483	3.6159	86.4983	434.3301	2.5130	0.2330	0.5210
0.84-rooting, X_3	22.24	1.2706	0.5740	2.0497	29.8071	149.7822	1.4891	0.0616	0.2909

* AC3 is the EME.

We also consider the grayscale image of size 463×604 , which is shown in Figure 22a. The images of the log, square root, the third root, and contrast stretching (CS) grayscale transformations (GST) are shown in parts (b–e), respectively. The negative image is shown in part (f). The enhanced images by the HE and the 0.9650- and 0.8150-rooting methods are shown in parts (g–i), respectively. The average of these two rooting images are given in part (j). The graphs of the EME and AMEE measures are given in part (k). The maximum of EME is at point 0.9650 and for the AMEE the maximum is at point 0.8150.



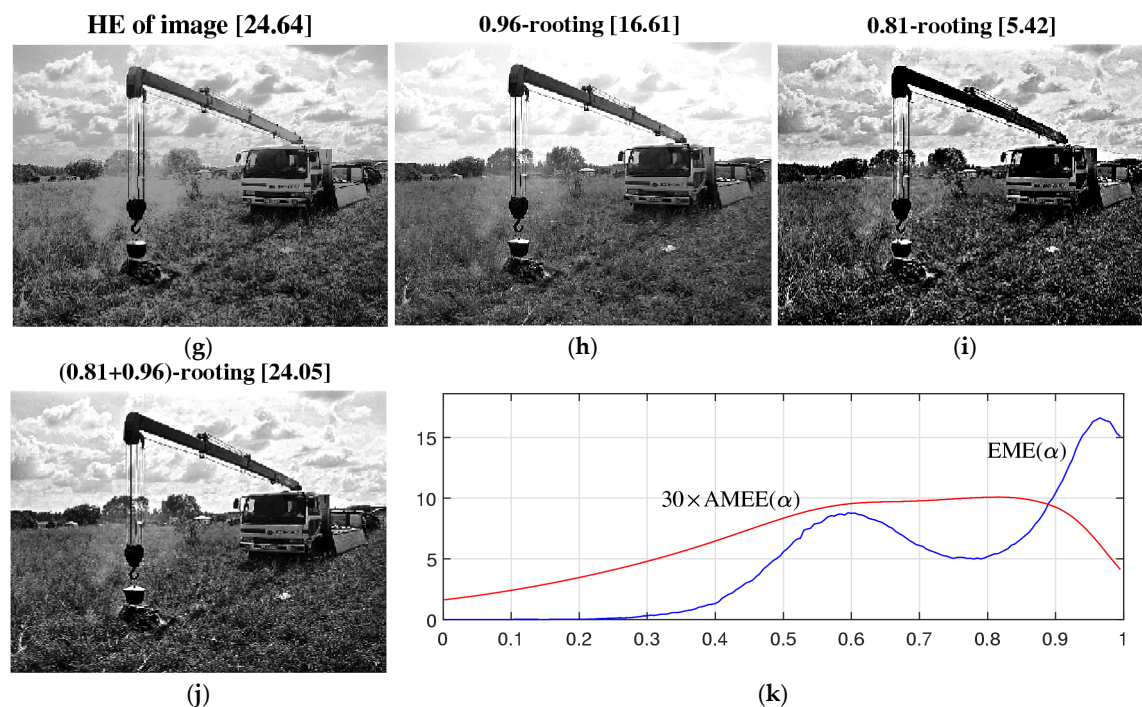


Figure 22. (a) The original image of size 453×604 pixels, (b) the log transformation, (c) square root, (d) the 3rd root, (e) the contrast stretching transform, (f) negative image, (g) HE, (h) 0.96-rooting, (i) 0.81-rooting, (j) average of rooting images, and (k) the graphs of the EME and AMEE.

All measurements of these ten images are given in Table 16.

Table 16. Nine metrics for the images in Figure 22.

Image	EME	C	MC	AC1	AC2	AC3	AC4	AC5	AMEE
original	15.11	0.65	0.31	1.19	2.83	9.56	1.11	0.08	0.13
log GST	3.82	0.18	0.09	0.81	0.83	4.12	1.02	0.02	0.02
sqrt GST	7.47	0.34	0.17	0.91	0.61	1.72	1.04	0.04	0.05
3rd root GST	7.47	0.24	0.12	0.84	0.85	3.88	1.03	0.03	0.03
CS GST	12.27	0.51	0.26	1.08	2.16	7.42	1.09	0.06	0.09
negative	14.01	0.79	0.35	1.27	5.68	23.21	1.24	0.10	0.12
HE	24.64	1.02	0.45	1.65	10.25	41.40	1.27	0.14	0.23
0.74-rooting	16.61	0.92	0.42	1.67	18.34	87.76	1.37	0.12	0.21
0.84-rooting	5.42	1.47	0.64	3.16	99.78	533.30	2.31	0.30	0.40
2 rooting	24.05	1.41	0.52	2.00	25.05	121.46	1.45	0.19	0.29

The combined metric has shown particular value in medical imaging applications [51–53].

1. *Diagnostic Imaging:* EMEE components detect subtle tissue variations and small anatomical structures; GIA ensures that contrast and brightness levels remain clinically relevant; combined metrics assist in the early detection of pathological changes; and applications include mammography, CT scan analysis, and MRI quality assessment.
2. *Microscopy:* Local metrics evaluate cellular structure details; global assessment maintains consistent imaging conditions, particularly useful in digital pathology and cell analysis; and it is also used in satellite and aerial imaging.
3. *Remote sensing:* EMEE identifies small ground features and texture changes; GIA evaluates large-scale patterns and atmospheric effects. Combining these metrics can improve land-use classification accuracy and support timely applications in environmental monitoring and urban planning.

4. *Weather monitoring*: Local metrics detect cloud formation patterns; global assessment tracks large-scale weather systems and improves weather prediction model accuracy.

7. Adaptive Contrast-Weighted Structural Similarity Measure

Image similarity and quality assessment remain challenging because they require balancing low-level signal fidelity with perceptual relevance. Traditional full-reference metrics often rely on pixel-wise or structural differences that do not always align with human visual perception, particularly when images contain multiple distortion types, local contrast changes, or domain-specific degradation patterns. Similar limitations have been widely discussed in both full-reference and no-reference quality assessment, particularly when objective scores lack reproducibility or fail to generalize across content and distortion conditions [54].

No-reference image quality assessment addresses part of this limitation by estimating perceptual quality directly from the distorted image. Earlier NR-IQA methods used natural scene statistics, sharpness, blur, gradient profiles, and transform domain statistics to capture perceptually meaningful degradation without requiring a pristine reference [55–58]. More recent approaches, including transformer-based and deep-learning-based models, have improved perceptual alignment but often introduce higher computational complexity and reduced interpretability [59–62]. These observations motivate the development of a similarity measure that combines the interpretability of classical full-reference metrics with perceptually motivated no-reference weighting.

In this section, we introduce the adaptive contrast-weighted structural similarity measure (ACSSIM). The method integrates local structural similarity with contrast-based weighting derived from no-reference image characteristics. By combining these components, the ACSSIM aims to improve sensitivity to perceptually relevant distortions while maintaining the simplicity and interpretability of classical full-reference metrics.

Let $f = \{f_{n,m}\}$ and $g = \{g_{n,m}\}$ be the reference and distorted images, respectively. The ACSSIM framework operates by decomposing the images into local regions and computing a weighted similarity score based on both structural and contrast information. The novel similarity metric of these images can be calculated as follows:

1. Splitting the images into blocks (*or windows*) $W_k, k = 1:K$, of $L_1 \times L_2$ pixel each, where $L_1, L_2 = 3, 5, 8$, or 16.
2. Calculating the local contrast terms for the images f and g in each window W_k , denoted by $AC(f_{n,m}; W_k)$ and $AC(g_{n,m}; W_k)$, respectively. Here, the metric $AC = AC_5$ (Equation (25)); the AMEE metric can also be used, C_{AMEE} (Equation (31)).
3. Calculating the local image similarity measures (ISM)

$$ISM(f, g; W_k) = 1 - SSIM(f, g; W_k)$$

$$\text{or } ISM(f, g; W_k) = MSE(f, g; W_k).$$

4. Averaging the local similarity responses using normalized contrast-adaptive weights over all windows to obtain the final ACSSIM value:

$$ACSSIM(f, g) = \frac{\sum_{k=1}^K w_k \cdot ISM(g, f; W_k)}{\sum_{k=1}^K w_k}, w_k = \left(AC(f_{n,m}; W_k) AC(g_{n,m}; W_k) \right)^{\frac{1}{2}}.$$

To evaluate the behavior of the proposed metric, a controlled experiment was conducted using synthetically generated distortions applied to a single reference image (Figure 23). The distortions include intensity changes, contrast variations, additive noise,

blur, and compression artifacts. The distortions were designed so that their mean squared error (MSE) values are approximately equal, enabling a controlled comparison in which pixel-wise fidelity remains constant while perceptual quality varies.

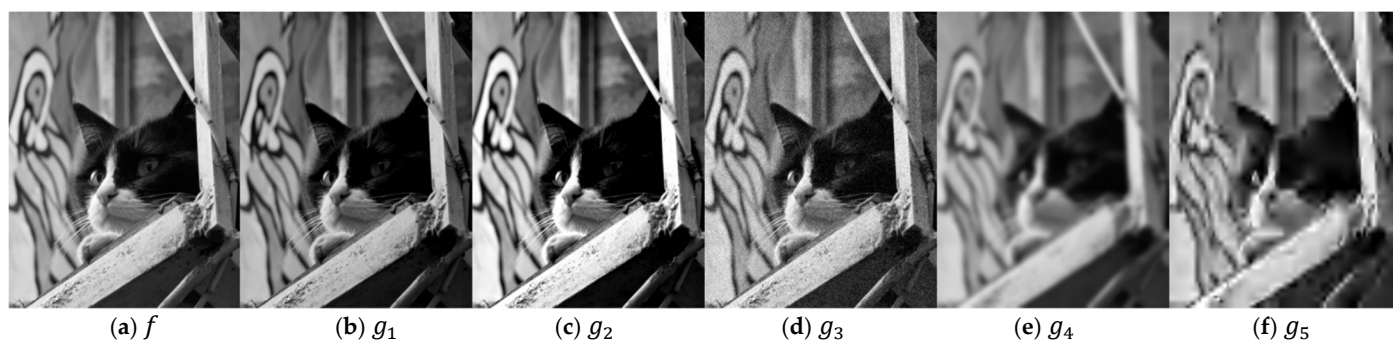


Figure 23. Example of artificially generated test distortions for the cat image: (a) original reference image, (b) mean-shifted image, (c) contrast-modified image, (d) noisy image, (e) blurred image, and (f) JPEG2000-compressed image.

The results of this experiment are summarized in Table 17. A key observation is that classical fidelity-based metrics such as MSE and PSNR produce nearly identical values for several different distortions, as expected from the controlled design of the examples. However, the perceived quality of the images differs noticeably, showing that these pixel-error metrics are not sufficiently sensitive to structural and perceptual changes. SSIM responds more strongly to structural degradation, but its lowest score is obtained for the noisy image, although the visual examples suggest that blur and JPEG compression introduce more severe perceptual degradation. In contrast, the proposed ACSSIM metric provides a more consistent ordering of distortion severity by assigning progressively larger distortion scores from mild luminance and contrast changes to stronger noise, blur, and compression artifacts.

Table 17. Classical FR metric values for the cat example shown in Figure 23.

Image Pair	MSE ↓	PSNR ↑	SSIM ↑	MS-SSIM ↑	FSIM ↑	GMSD ↓	VIF ↑	ACSSIM ↓
(f, f)	0	80	1	1	1	0	1	0
(f, g_1)	373.9	22.4	0.8839	0.9884	0.9887	0.0392	0.9330	0.0281
(f, g_2)	373.9	22.4	0.8174	0.9528	0.9452	0.0773	0.9165	0.0642
(f, g_3)	373.9	22.4	0.4294	0.8839	0.9029	0.1343	0.3262	0.3558
(f, g_4)	373.8	22.4	0.6746	0.8546	0.8134	0.1800	0.1584	0.4041
(f, g_5)	350.7	22.68	0.6482	0.8327	0.8281	0.1771	0.1040	0.4312

A qualitative comparison of distortion severity is presented in Table 18. The table includes human visual analysis and AI-assisted analysis to provide an additional subjective interpretation of the examples. Both analyses show strong agreement with the ACSSIM-based ranking for mild distortions, such as mean shift and contrast change, as well as for the intermediate noise case. Minor differences appear for blur and JPEG compression, where the perceived severity depends on whether loss of fine detail or visible compression artifacts is considered more disturbing. This result highlights the complexity of perceptual quality assessment and supports the need for structure- and contrast-aware similarity measures.

Table 18. Qualitative comparison of distortion severity based on human and AI-assisted analysis.

Distortion	Human Analysis	ChatGPT Analysis(GPT-5.4)	ACSSIM Ranking	GPT Ranking
Mean shift	Barely noticeable global intensity change; no structural degradation	Minor luminance shift with negligible perceptual impact; structural content fully preserved	1	1
Contrast	Slightly more noticeable than mean shift; enhances intensity differences but preserves details	Mild contrast variation; edges and structures remain intact with minimal perceptual degradation	2	2
Noise	Visible noise present, but object structure and details remain clearly recognizable	Grain introduces visual disturbance but preserves edges and object boundaries	3	3
Blur	Significant degradation; edges and fine details are strongly smoothed, reducing clarity	Strong loss of high-frequency information; edges and contours are heavily degraded	4	5
JPEG	Most severe distortion; introduces strong artifacts that distort structures and visual consistency	Compression artifacts create artificial patterns and local distortions, but some structure remains	5	4

To further validate the proposed method, ACSSIM was evaluated on the TID2013 and KADID-10k datasets, two widely used benchmarks for image quality assessment. TID2013 contains 25 reference images and 3000 distorted images across 24 distortion types, with associated subjective scores [62]. KADID-10k contains 81 reference images and 10,125 distorted images across 25 distortion types and five degradation levels, with associated subjective scores [63]. A quantitative comparison was conducted using representative full-reference IQA methods implemented via the comprehensive image quality assessment toolbox pyiqa (see Table 19). As ACSSIM is a full-reference metric, comparison with established FR-IQA methods provides the most appropriate evaluation setting. The goal of ACSSIM is not to compete with complex learned models, but rather to enhance the classical SSIM framework while preserving key advantages such as simplicity, interpretability, and low computational cost. In this sense, ACSSIM can be viewed as a lightweight extension of structural similarity that incorporates local contrast weighting. On TID2013, ACSSIM achieves PLCC = 0.6995 and SRCC = 0.6709, improving upon the SSIM baseline (PLCC = 0.6558, SRCC = 0.6269). On KADID-10k, ACSSIM achieves PLCC = 0.5984 and SRCC = 0.6778, also improving upon SSIM (PLCC = 0.5755, SRCC = 0.6188).

Although ACSSIM does not outperform stronger FR-IQA methods such as MS-SSIM, FSIM, GMSD, PieAPP, and TOPIQ-FR, its value lies in its simplicity and interpretability. From a computational perspective, ACSSIM remains close to SSIM, since it adds only a local contrast weighting step and does not require deep feature extraction, model training, or complex multi-scale processing. Moreover, many learned IQA methods are trained or calibrated on large-scale synthetic distortion datasets, including KADID-10k or datasets with similar distortion types, which can provide an advantage when evaluated on such benchmarks. Therefore, these results support ACSSIM's intended role as an efficient and interpretable enhancement of a classical similarity metric, rather than a replacement for advanced data-driven approaches.

Table 19. Quantitative comparison of FR-IQA algorithms on the TID2013 and KADID-10k datasets.

Method	Dataset	TID2013		KADID-10k		Reference
	Metric Type	PLCC	SRCC	PLCC	SRCC	
MSE	Classical fidelity baseline	0.4321	0.6395	0.3273	0.6408	-
PSNR	Classical fidelity baseline	0.6601	0.6869	0.5557	0.6757	-
SSIM	Classical FR-IQA	0.6558	0.6269	0.5755	0.6188	[9]
MS-SSIM	Classical FR-IQA	0.7819	0.7859	0.6802	0.8256	[64]
VIF	Classical FR-IQA	0.7336	0.6770	0.6845	0.6792	[65]
FSIM	Classical FR-IQA	0.8322	0.8509	0.7094	0.8537	[66]
GMSD	Classical FR-IQA	0.8553	0.8044	0.8048	0.8474	[67]
LPIPS	Learned FR-IQA	0.7530	0.7445	0.7484	0.8225	[31]
PieAPP	Learned FR-IQA	0.7195	0.8438	0.7708	0.8647	[68]
DISTS	Learned FR-IQA	0.7538	0.7077	0.8057	0.8137	[69]
AHIQ	Learned FR-IQA	0.6796	0.6479	0.6734	0.6788	[70]
CKDN	Learned FR-IQA	0.6805	0.5944	0.4976	0.4675	[71]
TOPIQ-FR	Learned FR-IQA	0.9158	0.9165	0.9871	0.9857	[72]
ACSSIM	Proposed FR-IQA	0.6995	0.6709	0.5984	0.6778	

Figure 24 presents the distortion-wise SRCC values obtained on TID2013 for the baseline MSE and SSIM metrics, as well as their corresponding contrast-adaptive weighted ACSSIM variants. The proposed normalized ACSSIM formulation improves rank-order prediction across several distortion categories, showing its strongest gains for contrast change, where ACSSIM (SSIM) improves over SSIM by 0.1044 SRCC, and for quantization noise, where the gain reaches 0.0850 SRCC. Additional positive gains are observed for sparse sampling and reconstruction, JPEG/JPEG2000-related distortions, and multiplicative Gaussian noise, indicating that contrast-adaptive local weighting can better emphasize perceptually relevant regions for several structured or compression-like degradations. However, the improvement is not uniform across all distortion types. In particular, local block wise distortions show a decrease relative to SSIM, and smaller negative or near-neutral gains are also observed for impulse noise, additive noise in color components, and mean shift. These cases suggest that the current grayscale, contrast-weighted formulation is less effective for distortions that are highly localized, color-component dependent, or globally uniform, motivating future extensions with color-aware weighting and distortion-adaptive pooling.

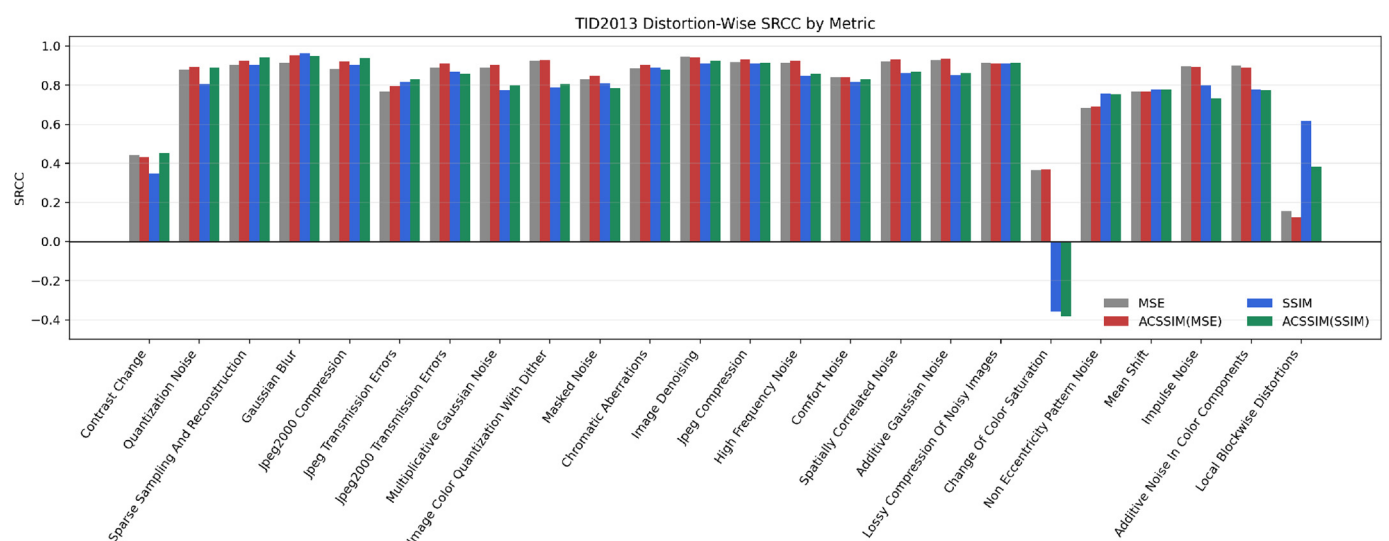


Figure 24. Distortion-wise SRCC comparison of baseline metrics and ACSSIM on TID2013.

Despite its advantages, the ACSSIM framework possesses specific limitations that warrant consideration:

- **Parameter sensitivity:** The framework's performance is partially dependent on the selection of the weighting constants for contrast and entropy. Sub-optimal parameter tuning may lead to "over-masking" in certain high-texture regions, potentially ignoring structurally relevant noise.
- **Block size dependency:** The accuracy of the local NR descriptors (EME/AMEE) is sensitive to block-size selection. If the block is too small, entropy measures become statistically insignificant; if too large, the "local" nature of the HVS masking is lost.
- **Global illumination shifts:** Like many FR-based metrics, the ACSSIM may be overly sensitive to uniform global luminance shifts that do not necessarily degrade the semantic content of the image but significantly alter the local Weber-contrast weights.

To address these limitations and expand the utility of the framework, several avenues for future research are identified:

- **Automated parameter optimization:** Investigating "shallow" machine learning techniques to dynamically tune ACSSIM weights based on the detected distortion category.
- **Integration of saliency maps:** Incorporating visual attention models to weight regions of interest more heavily, further aligning the metric with human foveal perception.
- **Extension to hypercomplex domains:** Applying ACSSIM within quaternion or octonion algebraic frameworks to better handle multi-channel color dependencies and image restoration.
- **Temporal consistency for video:** Adapting the framework for video quality assessment (VQA) by introducing temporal entropy weights to account for motion-based artifacts.
- **Cross-domain validation:** Testing robustness in specialized fields, such as medical imaging (MRI/CT) and remote sensing, where the preservation of high-frequency structural detail is a diagnostic necessity.

8. Conclusions

This study offers a tutorial on assessing image similarity and quality, bridging the gap between Full-Reference (FR) and No-Reference (NR) approaches. Our results show that traditional metrics like PSNR, SSIM, RMSE, and geometric distances often misalign with human visual perception, especially under compound distortions or non-linear contrast changes. Using psychovisual principles and information-theoretic descriptors, we developed ACSSIM, a lightweight framework that improves classical metrics with adaptive, contrast-sensitive local weighting. Tests on the TID2013 and KADID-10k benchmarks reveal that ACSSIM correlates better with Mean Opinion Scores (MOS) than the standard SSIM. Although not meant to replace complex deep-learning models, ACSSIM offers a better performance-efficiency trade-off suitable for real-time, resource-limited scenarios. This research reinforces the strategy that perceptually motivated enhancements of classical methods remain a powerful, interpretable, and practical approach to bridging the gap between objective measurement and subjective human experience.

Future work will focus on three main directions: automatic tuning of ACSSIM weighting parameters for different distortion types, integration of saliency maps to better model human visual attention, and extension of the framework to video and specialized domains such as medical imaging and remote sensing, where structural detail is especially important.

Author Contributions: Conceptualization, S.S.A.; methodology, S.S.A., and A.M.G.; software, A.M.G., and H.A.; validation, S.S.A., and A.M.G.; formal analysis, S.S.A., and A.M.G.; investigation, S.S.A., A.M.G., and H.A.; resources, S.S.A., and A.M.G.; data curation, S.S.A., and A.M.G.; writing—original draft preparation, S.S.A., and A.M.G.; writing—review and editing, S.S.A., A.M.G., and H.A.; visualization, S.S.A., A.M.G., and H.A.; supervision, S.S.A.; project administration, S.S.A.; funding acquisition, S.S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The authors used publicly available datasets in the experiments. Our codes is available on the web page <https://ceid.utsa.edu/agrigoryan/codes/> (accessed on 24 June 2026).

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

Abbreviations

The following abbreviations are used in this manuscript:

PSNR	Peak signal- to- noise ratio
SSIM	Structural similarity index measure
ACSSIM	Adaptive contrast-weighted structural similarity measure
FR	Full-reference
NR	No-reference
MAE	Mean absolute error
MSE	Mean square error
RMSLE	Root mean square logarithmic error
EME	Enhancement measure estimate
EWEE	Extended Weber entropy enhancement measure
AMEE	Absolute measure of enhancement by entropy
MC	Michelson contrast
IQA	Image Quality Assessment
PLCC	Pearson Linear Correlation Coefficient
SRCC	Spearman Rank Correlation Coefficient
GPT	Generative pre-trained transformer

References

1. Streijl, R.C.; Winkler, S.; Hands, D.C. Mean opinion score (MOS) revisited: Methods and applications, limitations and alternatives. *Multimed. Syst.* **2016**, *22*, 213–227.
2. Gao, Y.; Min, X.; Zhu, Y.; Li, J.; Zhang, X.-P.; Zhai, G. Image quality assessment: From mean opinion score to opinion score distribution. In *Proceedings of the 30th ACM International Conference on Multimedia*; Association for Computing Machinery: New York, NY, USA, 2022.
3. Kumar, B.; Singh, S.P.; Mohan, A.; Anand, A. Performance of quality metrics for compressed medical images through mean opinion score prediction. *J. Med. Imaging Health Inform.* **2012**, *2*, 188–194.
4. Silva, E.A.; Panetta, K.; Agaian, S.S. Quantifying image similarity using measure of enhancement by entropy. In *Proceedings of the SPIE 6579, Mobile Multimedia/Image Processing for Military and Security Applications 2007*, Orlando, FL, USA, 11–12 April 2007; p. 65790U.
5. Nercessian, S.; Agaian, S.S.; Panetta, K.A. An image similarity measure using enhanced human visual system characteristics. In *Mobile Multimedia/Image Processing, Security, and Applications 2011*; SPIE: Bellingham, WA, USA, 2011; Volume 8063.
6. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84.
7. Wang, Z.; Simoncelli, E.P. Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities. *J. Vis.* **2008**, *8*, 8.
8. Wang, Z.; Wu, G.; Sheikh, H.R.; Simoncelli, E.P.; Yang, E.-H.; Bovik, A.C. Quality-aware images. *IEEE Trans. Image Process.* **2006**, *15*, 1680–1689.

9. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612.
10. Wang, Z. Applications of objective image quality assessment methods. *IEEE Signal Process. Mag.* **2011**, *28*, 137–142.
11. Wang, Z.; Bovik, A.C. Reduced- and no-reference image quality assessment: The natural scene statistic model approach. *IEEE Signal Process. Mag.* **2011**, *28*, 29–40.
12. Wang, Z.; Li, Q. Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 1185–1198.
13. Hu, Z.; Yang, G.; Du, Z.; Huang, X.; Zhang, P.; Liu, D. No-reference image quality assessment based on global awareness. *PLoS ONE* **2024**, *19*, e0310206.
14. Zhang, H.; Yuan, B.; Dong, B.; Jiang, Z. No-reference blurred image quality assessment by structural similarity index. *Appl. Sci.* **2018**, *8*, 2003.
15. Röhling, T.; Jennifer, M.; Dirk, S. A fast histogram-based similarity measure for detecting loop closures in 3-d lidar data. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: Piscataway, NJ, USA, 2015.
16. Chaparala, P.; Mallarapu, P.; Pichuka, S.; Maram, G. New histogram-based user and item profiles for recommendation systems. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*; IEEE: Piscataway, NJ, USA, 2023.
17. Ma, Y.; Gu, X.; Wang, Y. Histogram similarity measure using variable bin size distance. *Comput. Vis. Image Underst.* **2010**, *114*, 981–989.
18. Lowe, D.G. Distinctive image features from scale-invariant key points. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.
19. Bay, H.; Ess, A.; Tuytelaars, T.; van Gool, L. Speeded-up robust features (surf). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359.
20. Oliva, A.; Torralba, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vision.* **2001**, *42*, 145–175.
21. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 20–25 June 2005; pp. 886–893.
22. Ling, H.; Okada, K. Diffusion distance for histogram comparison. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 17–22 June 2006; pp. 246–253.
23. Singh, S.; Siramdas, S.; Tanmayi, K.; Koli, L. Similarity measure of images using sift and orb feature matching. *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)* **2022**, *10*, 2354–2356.
24. Ghildyal, A.; Feng, L. Shift-tolerant perceptual similarity metric. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2022.
25. Taghanaki, S.A.; Choi, K.; Khasahmadi, A.H.; Goyal, A. Robust representation learning via perceptual similarity metrics. In *International Conference on Machine Learning*; PMLR: 2021. <https://proceedings.mlr.press/v139/taghanaki21a.html> (accessed on 24 June 2026)
26. Ghazanfari, S.; Garg, S.; Krishnamurthy, P.; Khorrami, F.; Araujo, A. R-LPIPS: An adversarially robust perceptual similarity metric. *arXiv* **2023**, arXiv:2307.15157.
27. Ghildyal, A. Alignment of perceptual similarity metrics with human perception. Ph.D. Thesis, Portland State University, Portland, OR, USA, 2025.
28. Spyridonos, P.; Gaitanis, G.; Likas, A.; Seretis, K.; Moschovos, V.; Feldmeyer, L.; Heidemeyer, K.; Zampeta, A.; Bassukas, I.D. Image perceptual similarity metrics for the assessment of basal cell carcinoma. *Cancers* **2023**, *15*, 3539. <https://doi.org/10.3390/cancers15143539>.
29. Al-Shurbaji, T.A.; AlKaabneh, K.A.; Alhadid, I.; Masa'dEh, R. An optimized scale-invariant feature transform using chamfer distance in image matching. *Intell. Autom. Soft Comput.* **2022**, *71*, 971–985.
30. Snell, J.; Ridgeway, K.; Liao, R.; Roads, B.D.; Mozer, M.C.; Zemel, R.S. Learning to generate images with perceptual similarity metrics. In *2017 IEEE International Conference on Image Processing (ICIP)*; IEEE: Piscataway, NJ, USA, 2017.
31. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE 9 Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2018.
32. Yang, P.; Wang, H.; Yang, J.; Qian, Z.; Zhang, Y.; Lin, X. Deep learning approaches for similarity computation: A survey. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 7893–7912.
33. Xu, J.; Zhou, W.; Chen, Z. Blind omnidirectional image quality assessment with viewport oriented graph convolutional networks. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 1724–1737.

34. Sheikh, H.R.; Bovik, A.C. A visual information fidelity approach to video quality assessment, January 2005. https://www.researchgate.net/publication/228368222_A_visual_information_fidelity_approach_to_video_quality_assessment (accessed on 24 June 2026)
35. Grigoryan, A.M.; Agaian, S.S. Transform-based image enhancement algorithms with performance measure. In *Advances in Imaging and Electron Physics*; Academic Press: Cambridge, MA, USA, 2004; Volume 130, Chapter 9, pp. 165–242.
36. Agaian, S.S. Visual morphology. In Proceedings of the SPIE 3646, Nonlinear Image Processing X, 139, 5 March 1999. <https://doi.org/10.1117/12.341081>.
37. Fomin, S.V.; Kolmogorov, A.N. *Elements of the Theory of Functions and Functional Analysis*; Dover Publications: Gaden, NY, USA, 1999.
38. Agaian, S.S.; Silver, B.; Panetta, K.A. Transform coefficient histogram-based image enhancement algorithms using contrast entropy. *IEEE Trans. Image Process.* **2007**, *16*, 741–758.
39. Grigoryan, A.M.; Agaian, S.S. Image processing contrast enhancement. In *Wiley Encyclopedia of Electrical and Electronics Engineering*; 2017; 22p. <https://onlinelibrary.wiley.com/doi/abs/10.1002/047134608X.W5525.pub2> (accessed on 24 June 2026)
40. Panetta, K.; Chen G.; Agaian, S. No reference color image contrast and quality measures. *IEEE Trans. Consum. Electron. Trans. Consum. Electron.* **2013**, *59*, 643–651.
41. Panetta, K.; Samani, A.; Agaian, S. Choosing the optimal spatial domain measure of enhancement for mammogram images. *Int. J. Biomed. Imaging* **2014**, *2014*, 937849.
42. Frigg, R.; Werndl, C. Entropy—A Guide for the perplexed. Archived 13 August 2011 at the Wayback Machine. In *Probabilities in Physics*; Beisbart, C., Hartmann, S., Eds.; Oxford University Press: Oxford, UK, 2010.
43. Shannon, C. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
44. Panetta, K.A.; Wharton, E.J.; Agaian, S.S. Human visual system-based image enhancement and logarithmic contrast measure. *IEEE Trans. Syst. Man. Cybern. Part B (Cybern.)* **2008**, *38*, 174–188.
45. Agaian, S.; Grigoryan, A.; Panetta, K. A new measure of image enhancement. In Proceedings of the IASTED International Conference on Signal Processing & Communication, Marbella, Spain, 19–22 September 2000; pp. 19–22.
46. Agaian, S.S.; Panetta, K.; Grigoryan, A.M. Transform-based image enhancement algorithms with performance measure. *IEEE Trans. Image Process.* **2001**, *10*, 367–382.
47. Wang, Z.; Sheikh, H.; Bovik, A. No-reference perceptual quality assessment of JPEG compressed images. In *International Conference on Image Processing*; IEEE: Piscataway, NJ, USA, 2002.
48. Marziliano, P.; Dufaux, F.; Winkler, S.; Ebrahimi, T. Perceptual blur and ringing metrics: Application to JPEG2000. *Signal Proc. Image Commun.* **2004**, *19*, 163–172.
49. Caviedes, J.; Oberti, F. A new sharpness metric based on local kurtosis, edge and energy information. *Signal Process. Image Commun.* **2004**, *19*, 147–161. <https://doi.org/10.1016/j.image.2003.08.002>.
50. Panetta, K.; Agaian, S.; Zhou, Y.; Wharton, E.J. Parameterized logarithmic framework for image enhancement. *IEEE Trans. Syst. Man. Cybern. Part B (Cybern.)* **2011**, *41*, 460–473.
51. Beghdadi, A.; Qureshi, M.A.; Amirshahi, S.A.; Chetouani, A.; Pedersen, M. A critical analysis on perceptual contrast and its use in visual information analysis and processing. *IEEE Access* **2020**, *8*, 156929–156953.
52. Zhuang, Y.; Chen, S.; Jiang, N.; Hu, H. An effective WSENet-based similarity retrieval method of large lung CT image databases. *KSII Trans. Internet Inf. Syst. (TIIS)* **2022**, *16*, 2359–2376.
53. Agaian, S.; Ayunts, H.; Trongtirakul, T.; Hovhannisyan, S. A new method for judging thermal image quality with applications. *Signal Process.* **2025**, *229*, 109769.
54. Pinson, M.H. Why no-reference metrics for image and video quality lack accuracy and reproducibility. *IEEE Trans. Broadcast.* **2023**, *69*, 97–117. <https://doi.org/10.1109/TBC.2022.3191059>.
55. Sheikh, H.R.; Bovik, A.C.; Cormack, L. No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Trans. Image Process.* **2005**, *14*, 1918–1927. <https://doi.org/10.1109/TIP.2005.854492>.
56. Liang, L.; Wang, S.; Chen, J.; Ma, S.; Zhao, D.; Gao, W. No-reference perceptual image quality metric using gradient profiles for JPEG2000. *Signal Process. Image Commun.* **2010**, *25*, 502–516. <https://doi.org/10.1016/j.image.2010.01.007>.
57. Brandão, T.; Queluz, M.P. No-reference image quality assessment based on DCT domain statistics. *Signal Process.* **2008**, *88*, 822–833. <https://doi.org/10.1016/j.sigpro.2007.09.017>.
58. Ferzli, R.; Karam, L.J. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *IEEE Trans. Image Process.* **2009**, *18*, 717–728. <https://doi.org/10.1109/TIP.2008.2011760>.

59. 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), Waikoloa, HI, USA, 4–8 January 2022; pp. 1220–1230.
60. Lu, W.; Sun, W.; Min, X.; Zhu, W.; Zhou, Q.; He, J.; Wang, Q.; Zhang, Z.; Wang, T.; Zhai, G. Deep neural network for blind visual quality assessment of 4K content. *IEEE Trans. Broadcast.* **2023**, *69*, 406–421.
61. Zhang, Z.; Zhou, Y.; Li, C.; Zhao, B.; Liu, X.; Zhai, G. Quality assessment in the era of large models: A survey. *ACM Trans. Multimed. Comput. Commun. Appl.* **2025**, *21*, 189. <https://doi.org/10.1145/3722559>.
62. Ponomarenko, N.; Jin, L.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Astola, J.; Vozel, B.; Chehdi, K.; Carli, M.; Battisti, F.; et al. Image database TID2013: Peculiarities, results and perspectives. *Signal Process. Image Commun.* **2015**, *30*, 57–77.
63. Lin, H.; Hosu, V.; Saupe, D. KADID-10k: A Large-Scale Artificially Distorted IQA Database. In Proceedings of the 2019 Tenth International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany, 5–7 June 2019; pp. 1–3.
64. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 9–12 November 2003; 2, pp. 1398–1402.
65. Sheikh, H.R.; Bovik, A.C.; de Veciana, G. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* **2005**, *14*, 2117–2128.
66. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386.
67. Xue, W.; Zhang, L.; Mou, X.; Bovik, A.C. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Trans. Image Process.* **2014**, *23*, 684–695.
68. Prashnani, E.; Cai, H.; Mostofi, Y.; Sen, P. PieAPP: Perceptual image-error assessment through pairwise preference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1808–1817.
69. Ding, K.; Ma, K.; Wang, S.; Simoncelli, E.P. Image quality assessment: Unifying structure and texture similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 2567–2581.
70. Lao, S.; Gong, Y.; Shi, S.; Yang, S.; Wu, T.; Wang, J.; Xia, W.; Yang, Y. Attentions help CNNs see better: Attention-based hybrid image quality assessment network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, New Orleans, LA, USA, 19–20 June 2022; pp. 1140–1149.
71. Yin, G.; Wang, W.; Yuan, Z.; Han, C.; Ji, W.; Sun, S.; Wang, C. Content-variant reference image quality assessment via knowledge distillation. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 3134–3142.
72. Chen, C.; Mo, J.; Hou, J.; Wu, H.; Liao, L.; Sun, W.; Yan, Q.; Lin, W. TOPIQ: A top-down approach from semantics to distortions for image quality assessment. *IEEE Trans. Image Process.* **2024**, *33*, 2404–2418.

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.