WHITEPAPER

Ensuring medical image fidelity:
Evaluation methods and techniques

Dr. Vaibhavi Sonetha
AVP, Citius Healthcare Consulting

Abstract

The accuracy and quality with which a medical imaging system captures, processes, and displays images is referred to as image fidelity. High fidelity in medical imaging is critical for accurate diagnosis, treatment planning, and patient management. There are various metrics available for assessing image quality and this paper explores several of them. The primary focus is on methods known as FR-IQA (Full-Reference Image Quality Assessment), which evaluate the quality of an image by comparing it with the reference image of perfect quality. These comparisons can be categorized as pixel-based or perception-based. The paper discusses various metrics within each category, outlining their advantages, limitations, and applications. Perception-based IQA is often considered superior to pixel-based IQA as the former is sensitive to the human visual system and the latter isn't.

Introduction

The development of a new radiology viewer and rendering pipeline is crucial to address the evolving needs and challenges of medical imaging. This includes incorporating advanced visualization, improving efficiency, enhancing diagnostic accuracy, ensuring seamless integration with clinical workflows, and compliance with evolving regulatory standards. Ultimately, these improvements contribute to improved patient care and diagnostic reliability.

A key aspect of such advancements is image fidelity—the degree of accuracy and faithfulness with which an image reproduces the visual details of the original subject. In the context of radiology or medical imaging, image fidelity would involve how well the viewer or rendering pipeline preserves the fine details, contrast, and overall clarity of the medical images being displayed, compared to the original data acquired from imaging modalities such as X-rays, CT, MRI, Ultrasound, X-ray, SPECT, PET, Mammography, etc. Assessing the quality of medical images is also very useful in evaluating new imaging software, hardware, acquisition methods, and algorithms.

Techniques for image quality assessment

Image quality can be assessed using two main techniques: subjective and objective methods. A widely used subjective assessment technique is the Double-Stimulus Continuous-Quality Scale (DSCQS), applied to evaluate MR, ultrasound, and telemedicine images. Subjective assessments rely on the judgment of qualified personnel, making them time-consuming and potentially biased due to individual factors such as perception, color preference, alertness, and display settings. Hence, though they are conventional, they are not ideal assessment tools.

In contrast, the objective method utilizes mathematical algorithms to assess the quality of medical images. The classification of the method can also be based on the presence of a reference image for comparison. If the reference image is available for comparison, then the method is known as Full Reference Image Quality Assessment (FR-IQA). In case of the absence of a reference image, the method is known as No Reference Image Quality Assessment (NR-IQA) or blind assessment.

Full Reference Image Quality Assessment (FR-IQA)

Various mathematically developed metrics are utilized to compare the quality of a captured image with a reference image in the case of objective assessment. The most widely used metrics are mean squared error (MSE) and peak signal-to-noise ratio (PSNR). MSE is computed by averaging the squared pixel intensity differences of two images, and PSNR is the ratio of the maximum value of pixel intensity to MSE.[1] Some of the other similar metrics are peak mean squared error (PMSE), root mean squared error (RMSE), average difference (AD), maximum difference (MD), mean absolute error (MAE), Shannon's information content, contrast-to-noise ratio (CNR) and normalized cross-correlation (NK).

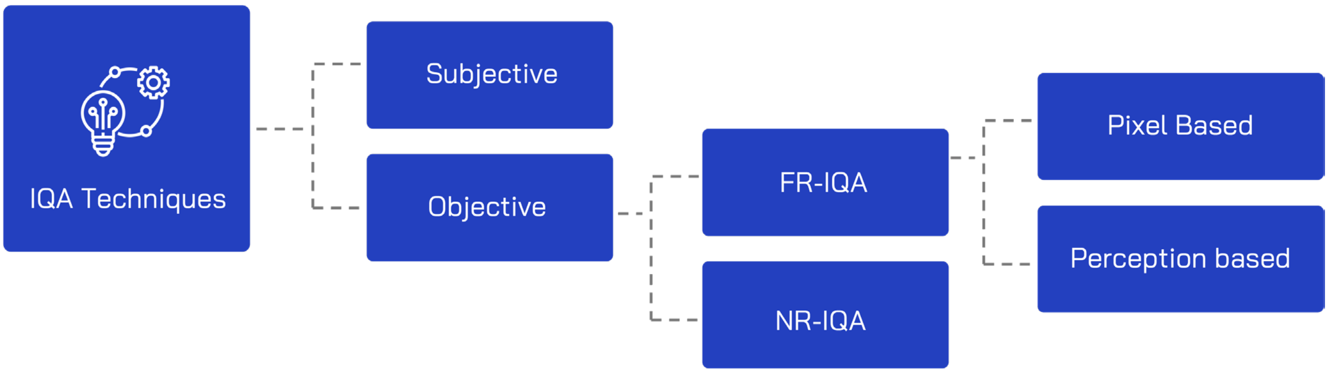

Fig. 1 Image quality assessment (IQA) techniques

Most of these metrics are based on pixel-wise differences but do not consider human perception of image quality in the assessment. A direct pixel-by-pixel comparison of a new image with a reference image may not accurately represent the true diagnostic capabilities of the new system, as it overlooks antialiasing, color perception, structural similarity, image resolution and size. This underscores the need for more advanced and comprehensive methods to ensure accurate and reliable image quality evaluation when assessing a new radiology viewer or a newer rendering pipeline.

Perception based FR-IQA

Metrics that are developed based on human visual perception include the perceptual difference model (PDM), structural similarity index (SSIM) family, learned perceptual image patch similarity (LPIPS), and deep image structure and texture similarity (DISTS), among others. These metrics aim to model human visual perception to evaluate image quality.

The perceptual difference model (PDM) is a computerized human vision model that incorporates optics and sensitivity of the retina, spatial contrast sensitivity function, and channels of spatial frequency of the visual cortex. It determines the visual differences between the two images and generates a spatial map depicting the magnitude of differences, along with a scalar PDM error per pixel.[2]

The structural similarity approach is based on the concept that the measure of structural information change is a good approximation to perceived image distortion. It compares the structure of the captured image with that of the reference image by computing luminance, contrast, and structure similarity as shown in Table 1.[3] The impact of all three of them together is captured as the SSIM index as in equation 1. SSIM is calculated for various local regions of an image. MSSIM is the mean value for structural similarity and is calculated as in equation 2. Multi-scale structural similarity (MS-SSIM) is an extension of the SSIM index which operates by computing SSIM at multiple scales or levels of the image, thereby capturing global as well as local structural information. This approach makes MS-SSIM more robust to variations in image resolution and better at capturing perceptual image quality. Table 2 compares the MSE, PSNR, PDM, and SSIM objective FR-IQA metrics.

Table 1: Parameters of structural similarity (SSIM)

Luminance Similarity	Contrast Similarity	Structure similarity
Mean value of pixel intensity	Standard deviation in pixel intensity	Correlation bet. pixel intensity
$l(\mathbf{x},\mathbf{y}) = \frac{2\,\mu_x\,\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$	$c(\mathbf{x},\mathbf{y}) = \frac{2\,\sigma_x\,\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$	$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\,\sigma_y + C_3}$

$$\mathrm{SSIM}\,(x,y) = \frac{(2\mu_x\mu_y + C_1)\,(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)\,(\sigma_x^2 + \sigma_y^2 + C_2)}$$

$$\mathrm{MSSIM}\,(x,y) = \frac{1}{M}\sum_{j=1}^{M}\mathrm{SSIM}\,(x_j, y_j)$$

Table 2: Comparison of objective FR-IQA metric

Metric	Objective	Advantages	Limitations	Application
MSE Range: 0 to ∞ Ideal value = 0	Measuring degree of similarity or distortion level	Convention, Simple, parameter free, compute inexpensive	Can't capture perceptual structural information / distortion, signs of error, image content dependent variation	Any imaging modality
PSNR Range: ∞ to -∞ Ideal value = ∞	Same as MSE	Useful for images with dynamic range of pixel	Same as MSE	Any imaging modality
PDM Ideal value = 0	Visual differences between images	Captures differences as by human visual system	Applicable only to study with similar input images and similar processing	Perceived image degradation by compression, Evaluation of MRI images & parallelly reconstructed MR images
MSSIM Ideal value = 1	Measuring similarity from luminance, contrast, structure	Spatially varying, perceived image quality, weighted model for human attracted textures/segments	Computationally expensive, need for equation optimization as per application, not appropriate for real-time medical images with no reference image	Quality of compressed & noisy MR images, blurred X-ray images, contrast ultrasound images, compressed CT images [4,5]

The learned perceptual image patch similarity (LPIPS) and deep image structure and texture similarity (DISTS) measure the perceptual similarity between two images leveraging a deep neural network (DNN). They utilize pre-trained convolutional neural networks (CNN) like VGG, AlexNet, or SqueezeNet to extract feature maps from the input images. The feature maps from the corresponding layers of CNN are compared between two images using a weighted distance metric. The weighted distances from all considered layers are then aggregated to produce a final similarity score.[6]

Different samples with the same texture have different features, but visually, they appear nearly the same. Most of the IQA methods, except DISTS, are highly sensitive to the different features of the same textured images and thus disregard the aspects of perceptual similarity. The DISTS metric is sensitive to structural distortions/artifacts and insensitive to the resampling of visual textures. The weighted distance metric of DISTS incorporates texture measurements using the global means and structure measurements using the global correlations. Equation 3 represents the DISTS index.[7]

$$D(x, y; \alpha, \beta) = 1 - \sum_{i=0}^{m} \sum_{j=1}^{n_i} \left(\alpha_{ij} l(\tilde{x}_j^{(i)}, \tilde{y}_j^{(i)}) + \beta_{ij} s(\tilde{x}_j^{(i)}, \tilde{y}_j^{(i)}) \right)$$

Where α, β = learnable weights, "l" = texture measurement parameter, "s" = structure measurement parameter, m = number of convolution layers, ni = number of feature maps in the ith convolution layer.

There are various other pixel-based and perception-based metrics for image quality assessment. Each metric has its strengths and weaknesses, and the selection of the most appropriate one depends on the application, type of distortion, computational complexity, and perceptual accuracy, etc.

Conclusion

Image quality assessment is the process of evaluating the perceived quality of an image, either involving human observers for evaluation or using metrics that can automatically perform the measurement. Assessing the image quality of a new radiology viewer or rendering pipeline is crucial to ensure accurate diagnosis, clinical validation, consistency, reliability, enhanced user experience, effective clinical workflows, and patient safety. If the reference image is available, then comparison can determine quality. There are many mathematical metrics available for this purpose. They are either pixel-based or perception-based. Pixel-based metrics are insensitive to structural changes, sensitive to small changes in pixel values, unable to capture high-level features such as texture, and have limited robustness. To address these limitations, more advanced methods have been developed to capture the complexities of human visual perception.

Pixel-based metrics can be useful for image quality validation in scenarios where exact pixel-level accuracy and fidelity are critical, such as validating the quality of lossless compression algorithm, developing image processing algorithms for denoising, deblurring, etc. When the visual quality and diagnostic utility of the images are paramount, perceptual quality metrics are appropriate. They provide a more human-centric evaluation of image quality ensuring that the applied process does not compromise the diagnostic quality of medical images. They shall be utilized for the evaluation of medical images transmitted over networks for remote diagnosis, developing algorithms for segmentation, detection, or classification of medical images, assessing the impact of compression on the diagnostic quality of medical images, and comparing output from various devices of the same imaging modality.

It's crucial to select appropriate IQA metrics to ensure an effective outcome. The selection parameters include application requirements, type of distortions present in an image, perceptual accuracy, computational efficiency, and robustness.

References:

1. Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures," in IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, 2009.
2. L. S. Chow, R. Paramesran, "Review of medical image quality assessment," in Biomedical Signal Processing and Control, vol. 27, pp. 145-154, 2016.
3. Z. Wang, A C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," in IEEE TRANSACTIONS ON IMAGE PROCESSING, vol. 13, no. 4, pp. 600-612, 2004.
4. B. Kumar, G.R. Sinha, K. Thakur, "Quality assessment of compressed MR medical images using general regression neural network", Int. J. Pure Appl. Sci. Technol., vol. 7, no. 2, pp. 158–169, 2011.
5. K.S. Rangaraju, K. Kumar, G.E. Healthcare, B.C.H. Renumadhavi, "Review paper on quantitative image quality assessment—medical ultrasound images", Int. J. Eng., vol. 1, no. 4, pp. 1–6, 2012.
6. K. Ding, K. Ma, S. Wang, et al, "Comparison of Full-Reference Image Quality Models for Optimization of Image Processing Systems," in International Journal of Computer Vision, vol. 129, pp. 1258–1281, 2021.
7. K. Ding, K. Ma, S. Wang and E. P. Simoncelli, "Image Quality Assessment: Unifying Structure and Texture Similarity," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 5, pp. 2567-2581, 2022.

Citius Healthcare Consulting

Citius Healthcare Consulting is a trusted partner in navigating the complexities of Healthcare and Life Sciences. As a transformative force, we empower organizations to overcome their critical business and technology challenges, driving sustainable growth.

By merging the management consulting expertise of what was formerly FluidEdge Consulting, with the digital healthcare capabilities of CitiusTech, our goal is to empower healthcare organizations with solutions that address their most critical challenges. Leveraging our deep domain knowledge and CitiusTech's cutting-edge HealthTech and Life Sciences innovations, Citius Healthcare Consulting strives to deliver impactful outcomes that enhance patient care and drive operational efficiency.

Building on our legacy of excellence, we continue to support Providers, Payers, Life Sciences companies, and MedTech organizations across the healthcare ecosystem-

- Experienced consultants: An average of 15+ years of experience
- Proven track record: Over 1,200 global projects delivered
- Cost Savings: Exceeding $300M in savings and a consistent return on investment through cost recovery and improved revenue capture

We are committed to shaping the future of healthcare.

Discover more at: www.citiustech.com/citius-healthcare-consulting