DOI:10.32703/2617-9040-2025-46-10

UDC 004.925.4

***Karnatov Serhii*[1*]**

[1]Postgraduate student, Automation and Computer-Integrated Transport Technologies Department, National Transport University, M. Omelyanovicha-Pavlenko St., 1, Kyiv, 01010, Ukraine. ORCID: https://orcid.org/0009-0006-6254-6166.

***Corresponding author:** serafim@ukr.net.

Analysis of PSNR, SSIM, LPIPS metrics in the context of human perception of visual similarity

This paper presents a comprehensive comparative analysis of three well-known image quality assessment (IQA) metrics: PSNR, SSIM, and LPIPS. It explores their basic principles, mathematical foundations, advantages, and limitations, particularly as they relate to human visual perception. The evolution of IQA metrics from simple pixel-by-pixel comparisons (PSNR) to structural approaches (SSIM) and, more recently, to learned perceptual metrics (LPIPS) is discussed. A critical analysis of the effectiveness of each metric in assessing various visual distortions, including noise, blur, and compression artifacts, is presented. Inherent issues in human visual perception, such as the role of semantics, texture, color, and visual artifacts, are explored as fundamental causes of discrepancies between objective metric estimates and subjective human judgments. The paper highlights the "unproven effectiveness" of deep features in LPIPS, and discusses its vulnerabilities, such as adversarial attacks and limitations in global semantic understanding. Finally, it outlines directions for future research aimed at developing more robust, interpretable, and perceptually consistent IQA metrics that can better account for the complexity of the human visual system and the evolving demands of modern image processing and generative artificial intelligence technologies.

***Keywords**: Image quality assessment, PSNR, SSIM, LPIPS, human perception, visual distortions, generative models, objective metrics, subjective assessment.*

Introduction. Image quality and similarity assessment (Image Quality Assessment, IQA) is a fundamental task in modern image processing and computer vision technologies. Objective quantification of visual quality or the degree of similarity between images is critical for a wide range of applications, from data compression and image restoration to the development of artificial intelligence generative models (AIGC), medical imaging, and video surveillance systems. The goal of research in this area is to develop quantitative measures that reliably reflect visual quality and match human perception.

Historically, IQA metrics have undergone significant evolution. Early approaches, such as mean square error (MSE) and peak signal-to-noise ratio (PSNR), were based on simple pixel-by-pixel comparisons. However, it quickly became apparent that such metrics have significant limitations. As noted in [1], "classical pixel-by-pixel metrics, such as Euclidean distance l_2 , are inadequate for evaluating structured outputs such as images because they assume inter-pixel independence." This has led to a fundamental problem: the disconnect between the objective numerical measurements that these metrics provide and the subjective human perception of visual quality and similarity. The human visual system (HVS) processes visual information in a much more complex way, taking into account structural

relationships, semantic context, and other perceptual aspects that are ignored by simple pixel-by-pixel metrics.

This gap has stimulated the development of more sophisticated metrics. The emergence of the structural similarity index (SSIM) was an important step forward, as it takes into account the degradation of structural information, which better correlates with human perception [2]. Later, with the development of deep learning, metrics such as LPIPS (Learned Perceptual Image Patch Similarity), which are trained directly on large datasets of human similarity judgments, attempting to model HVS even more accurately [3, 4]. This evolution reflects a gradual awareness of the complexity of human visual perception and the desire to create tools that more adequately reflect subjective experience.

Accurate IQA metrics are particularly relevant in the context of the rapid development of artificial intelligence generative models (AIGCs) that are capable of creating photorealistic images. Traditional metrics often fail to adequately assess the quality of such images, which may contain subtle but visually significant artifacts or semantic inconsistencies. Metrics that focus on per-pixel or simple structural differences may not capture these nuances. This creates an urgent need for metrics that better align with human perception of fine detail and semantic coherence, which are critical for evaluating AI-generated content [5, 6].

Analysis of recent research and problem statement. Research in the field of image quality assessment is an extremely active area in computer vision and signal processing, and numerous works demonstrate both progress and unsolved challenges.

Early work focused on metrics based on signal error, such as PSNR, derived from MSE. They were dominant due to their simplicity of calculation and clear mathematical interpretation. However, as Wang et al. point out in their pioneering work on SSIM [2], "traditional image quality metrics such as MSE and PSNR do not correlate well with human perception of image quality." Research also indicates that "PSNR does not always correlate with perceived visual quality because human perception of images can be affected by factors not reflected in pixel differences." According to Zhang's work and al. [7], PSNR "even gives the same value for all very different degradations ". This is clear evidence of the inadequacy of PSNR for perceptual assessment.

The answer to these limitations was the introduction of structural metrics. Wang and al. [2] proposed SSIM based on the idea that HVS is highly adapted to extract structural information. SSIM showed a significantly better correlation with human judgments compared to PSNR, which was confirmed in many subsequent studies. However, SSIM also has its limitations. In the work "Image quality assessment: From error visibility this structural similarity" [2, 8] shows that "a small spatial shift of an image can mean that it has a very low SSIM score, although the subjective image quality is the same as the reference". SSIM can also be insensitive to hue changes [9, 10]. Research on medical image quality assessment [11] indicates that SSIM "cannot identify a hole (local information loss)". The development of multi-scale SSIM (MS-SSIM) [12] was an attempt to overcome the limitations of single-scale analysis by recognizing the multi-scale nature of LSI.

With the advent of the deep learning era, a new class of perceptual metrics has emerged. LPIPS, proposed by Zhang and al. [4], is a prime example of this direction. This metric uses features extracted from pre-trained CNNs and is trained on large datasets of human perceptual judgments (e.g., the BAPPS dataset containing hundreds of thousands of comparisons [4]). Original work by Zhang and al. showed that deep features significantly outperform classical metrics in correlation with human perception. This is also confirmed in [13, 14], where it is noted that "CNNs trained on ImageNet learn a hierarchy of features from simple (contours, textures) to complex (parts of objects, objects), which is somewhat analogous to hierarchical processing in HVS". Despite this, LPIPS is not an ideal metric. As noted in [15], LPIPS is vulnerable to adversarial attacks, where small, visually imperceptible changes to a human result in a significant change in the LPIPS estimate. In addition, its patch-oriented analysis may not take into account global semantic coherence [16]. Problems with estimating massive local information loss (e.g., a "hole" in an MRI) have also been noted for LPIPS [11].

The relevance of the problem also increases in the context of assessing the quality of images generated by artificial intelligence. Traditional metrics are often unable to adequately assess the quality

of such images, which may contain subtle but visually significant artifacts or semantic inconsistencies [16]. This creates an urgent need for metrics that better match human perception of fine details and semantic coherence.

Thus, despite significant progress, the development of IQA metrics that fully reflect the complexity of human visual perception remains an open problem. Each of the considered metrics – PSNR, SSIM, LPIPS – has its own strengths and weaknesses, and none of them can be a universal "gold standard" for all types of distortions and applications. There is a need for a deep comparative analysis of these metrics to clearly outline the scenarios of their effective and ineffective use.

Research goals and objectives. The main goal of the research is to conduct a comparative analysis of the PSNR, SSIM, and LPIPS metrics in order to identify their effectiveness in reflecting human visual perception of image similarity and quality, as well as to determine their advantages, disadvantages, and specifics of application in the context of various types of visual distortions.

Materials and methods of research. Classical metrics for assessing image quality PSNR (Peak Signal-to-Noise Ratio) is one of the oldest and simplest metrics used to evaluate image quality, especially in the context of lossy compression.

Principle of operation and formula: PSNR measures the ratio between the maximum possible signal (image) power and the power of distorting noise, which affects the accuracy of its representation. PSNR is based on the mean square error (MSE), which is calculated as the average square of the differences in pixel intensities between the original (I) and distorted (K) images of size $M \times N$

$$MSE = \frac{1}{M \times N} \cdot \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left[I(i,j) - K(i,j) \right]^2 . \quad (1)$$

After calculating the MSE, the PSNR is calculated using the formula:

$$PSNR = 10 \cdot log_{10} \left(\frac{MAX_I^2}{MSE} \right), \quad (2)$$

where MAX_I is the maximum possible pixel value of the image (for example, 255 for an 8-bit image); PSNR is measured in decibels (dB), and higher values generally indicate better quality of the reconstructed image. Typical PSNR values for lossy image and video compression range from 30 to 50 dB for 8-bit data.

Advantages: The main advantages of PSNR are its mathematical simplicity, ease of calculation, and clear physical meaning, especially for estimating additive white Gaussian noise. Due to these characteristics, PSNR is widely used as a baseline performance indicator for lossy compression algorithms and other image processing methods.

Disadvantages: Despite its widespread use, PSNR has a number of significant drawbacks that limit its applicability for adequate assessment of perceptual quality:

- Low correlation with human perception: PSNR often correlates poorly with how humans perceive image quality. The metric treats all per-pixel errors equally, regardless of their visual impact, structural context, or location in the image.
- Insensitivity to structural distortions: PSNR does not take into account structural information in the image. Therefore, it is insensitive to distortions such as blurring, blocking, edge displacement, or other artifacts that destroy the structure but may have a small mean square error.
- Misleading results: Two images with the same PSNR value can have drastically different visual quality to the human eye [7].
- Ignoring masking effects of LSR: PSNR does not take into account such important aspects of human vision as masking effects, when distortions in some areas of the image (for example, textured or high-contrast) are less noticeable than in others.

The mathematical simplicity and ease of calculation of PSNR, which have contributed to its widespread use, are at the same time its fundamental disadvantage for perceptual assessment.

SSIM (Structural Similarity Index Measure). The structural similarity index (SSIM) has been proposed as an alternative to PSNR that better matches human perception of image quality because it estimates the degradation of structural information [2].

Working principle and formula: The basic principle of SSIM is that the human visual system is highly adapted to extract structural information from a scene. Therefore, a metric that measures the preservation of structure should correlate better with subjective quality assessment. SSIM is a perception-based model that considers image degradation as a perceived change in structural information, unlike MSE or PSNR. SSIM is computed locally for image windows and compares three key features: luminance, contrast, and structure.

The general SSIM formula for two windows x and y has the form [2]:

$$SSIM(x,y) = \left[l(x,y)^{\alpha} \cdot c(x,y)^{\beta} \cdot s(x,y)^{\gamma} \right], \tag{3}$$

where $l(x, y)$, $c(x, y)$ and s(x, y) are the brightness, contrast and structure comparison components, respectively;

α, β, γ > 0 are parameters that determine the relative importance of each component (usually set to 1).

The SSIM value ranges from -1 to 1 (or 0 to 1 in some implementations), where 1 means perfect similarity. The overall score for the entire image is usually obtained as the mean value of the local SSIMs (MSSIM).

Advantages:

- Better correlation with HVS: SSIM correlates significantly better with human perception of image quality compared to PSNR because it takes into account structural changes that are important for HVS [2].
- Sensitivity to important aspects: The metric is sensitive to changes in brightness, contrast, and structural details.
- Local Similarity Map: SSIM allows you to generate a local similarity map (SSIM map), which shows how quality varies across an image, providing more information than a single global value.
- Consideration of masking effects: SSIM implicitly takes into account the masking effects of brightness and contrast through the way its components are calculated.

Disadvantages:

- Imperfect LSI modeling: SSIM still does not fully model all aspects of human visual perception. It may not perform well with certain types of distortions, such as strong blurring, significant color shifts, or small spatial shifts [2].
- Sensitivity to geometric transformations: SSIM is sensitive to rotations and scaling, although there are modifications such as CW-SSIM that attempt to address this issue.
- Single-scale limitation: Classical SSIM analyzes images at a single scale, which is a limitation because the LMS perceives information at different levels of detail. This drawback is partially addressed in multi-scale SSIM (MS-SSIM) [12].
- Problems with medical images: SSIM may underestimate distortion near sharp edges or have instability in low-dispersion regions, which is especially relevant for medical images [11].
- Insufficient consideration of semantics: SSIM, while taking structure into account, is still a relatively low-level metric and does not analyze the semantic content of the image.

SSIM is an important step towards creating more perceptually relevant metrics. However, its components (brightness, contrast, structure) are still relatively low-level statistical characteristics of pixel intensities in a local area, not capturing deep semantic aspects or the more complex cognitive processes of human perception.

Modern learning-based metrics: LPIPS. With the advent of deep learning, a new class of image quality and similarity metrics has emerged that are trained directly on human perceptual judgment data. One of the most famous such metrics is LPIPS (Learned Perceptual Image Patch Similarity).

LPIPS (Learned Perceptual Image Patch Similarity). How it works: LPIPS computes the perceptual similarity between two images by measuring the distance between their representations in a deep feature space extracted using convolutional neural networks (CNNs). These CNNs (e.g., AlexNet, VGG, SqueezeNet) are typically pre-trained on large datasets for high-level tasks such as image classification (e.g., ImageNet), or they can be specifically trained or fine -tuned on datasets containing human ratings of perceptual similarity of images, such as BAPPS (Berkeley-Adobe Perceptual Patch Similarity) dataset [4].

The LPIPS calculation process is as follows:

- Two images (reference and distorted) are passed through the selected pre-trained CNN.
- Activation maps (features) are extracted from several layers of the network. Different layers correspond to different levels of abstraction of visual information, from low-level textures to more complex structural elements.
- These activation maps for each image are processed (e.g., normalized across channels).
- The distance (usually the weighted Euclidean distance l_2) between the corresponding activation maps for the two images is calculated.
- These distances are averaged across spatial dimensions and across layers (with specific weights for each layer, which can also be learned) to produce a single LPIPS value. A low LPIPS value indicates high perceptual similarity between images.

Advantages:

- High correlation with human perception: LPIPS shows significantly better correlation with human judgments of image similarity compared to traditional metrics such as PSNR and SSIM [4].
- Ability to capture complex visual distinctions: Through the use of deep features, LPIPS can distinguish subtle textural and structural nuances.
- The "unreasonable efficiency" of deep features: Studies have shown that even features from networks trained on high-level tasks turn out to be surprisingly effective at estimating low-level perceptual similarity.

Limitation:

- Vulnerability to adversarial attacks: One of the most significant shortcomings of LPIPS is its vulnerability to adversarial attacks. Small changes, visually imperceptible to a human, can lead to a significant change in the LPIPS estimate, which does not correspond to human perception [15].
- Patch-oriented analysis and global semantics: LPIPS computes similarity based on the comparison of individual image patches. While this allows for the analysis of local details and textures, this approach may not fully account for global semantic coherence or long-term spatial dependencies across the entire image [16].
- Dependence on the architecture and training of the underlying CNN: The quality of the features used by LPIPS depends on the specific CNN architecture and the data on which it was trained.
- Computational complexity: Compared to PSNR and SSIM, LPIPS can be a more resource-intensive metric.
- Interpretability: Like many deep learning models, LPIPS can be less interpretable ("black box") compared to metrics like SSIM.
- Specific types of distortion: LPIPS may have difficulty estimating images with very low texture or homogeneous regions. Problems with variations between stereo pairs have also been noted.

Despite these limitations, LPIPS has become an important tool in image quality assessment, especially for generative models and image restoration tasks where perceptual quality is a priority [17].

The problem of human perception of image similarity. Understanding how humans perceive visual similarity is key to developing and evaluating the effectiveness of objective image quality metrics. The human visual system (HVS) is an extremely complex mechanism whose work goes far beyond the simple registration of light intensities.

Human visual perception is a multi-step process that begins with the detection of photons of light by the retina and ends with the formation of a conscious image and its interpretation in the brain. It is important to emphasize that what we see is not a simple translation of the retinal stimulus; the brain

actively processes, filters, and supplements the information received. This active interpretation is one of the main reasons for the discrepancies between objective metrics and subjective assessment.

Early theories of perception, such as Hermann von Helmholtz's theory of unconscious inference, argued that the brain makes assumptions and inferences based on incomplete sensory data and prior experience. For example, we unconsciously assume that light falls from above, faces are usually perceived upright, and nearer objects can obscure more distant ones. These built-in assumptions help us quickly interpret a visual scene, but they can also lead to visual illusions.

Gestalt theory has provided a number of important principles for organizing visual elements into coherent images or "gestalts" [18]. These principles include:

- Proximity: Objects located nearby are perceived as a group.
- Similarity: Elements that are similar in shape, color, size, or other characteristics are grouped together.
- Closure: The HVS tends to "complete" the missing parts of a figure in order to perceive it as complete.
- Symmetry: Symmetrical objects are more easily perceived as a single whole.
- Common destiny: Elements moving in the same direction are perceived as connected.
- Continuity (or good continuation): The HVS prefers the perception of smooth, continuous lines and contours.
- Good Gestalt (law of pregnancy): The simplest, most regular, and most stable shapes are perceived.
- Past Experience: Prior knowledge and experience influence how we interpret visual stimuli.

These principles show that the LNS actively structures visual input, rather than passively registering it. David Marr proposed to consider vision as a feature processing process that involves extracting basic components of a scene, such as edges, corners, regions, and textures, forming the so-called "primary sketch" [19]. This processing is hierarchical: from low-level features (color, brightness, orientation of local contours), the system proceeds to the analysis of medium-level characteristics (texture, grouping of elements, spatial arrangement of the scene) and, finally, to high-level interpretation (recognition of objects, understanding their relationships and the semantic content of the scene).

The LMS is also a highly adaptive system. Its sensitivity to different spatial frequencies (detail), contour orientations, contrast levels, and motion dynamics is not constant, but varies depending on the viewing conditions and the characteristics of the stimulus itself. Masking effects play an important role, when the presence of some visual elements (for example, complex texture or high contrast) can reduce the visibility of other elements or distortions in the same image area. In addition, visual attention (saliency) directs processing resources to the most significant or informative parts of the scene, which means that not all parts of the image are perceived with the same detail and importance.

The complexity and multifactorial nature of LPS is the root cause of the inadequacy of simple metrics. Objective metrics often focus on one or a few aspects (e.g., pixel-wise difference for PSNR, local structure for SSIM), while humans integrate a vast amount of information at different levels. PSNR ignores virtually all of the listed aspects of LPS. SSIM partially accounts for structure and some masking effects, but not semantics or global organization. LPIPS, trained on human data, tries to implicitly capture this complex processing, but is still limited by its architecture and training data.

Despite continuous progress in the development of objective image quality metrics, discrepancies between their estimates and human subjective perception remain a significant problem. These discrepancies arise due to the fundamental complexity of HVS and its ability to take into account context, semantics, and previous experience, which are difficult to formalize in the form of mathematical models.

Examples of counterintuitive metric results:

- PSNR can show the same values for images with completely different types of distortion (e.g., blur, noise, compression artifacts), which are visually perceived in completely different ways [7].
- SSIM, although it correlates better with HVS, also has its weaknesses. For example, it may give a low score to images with small spatial shifts that appear almost identical to humans. Conversely, SSIM

may show high similarity for images with significant color changes, if their structure is preserved. SSIM may also have problems with the assessment of strongly blurred images or images with massive local information loss [11].

- LPIPS, despite its perceptual validity, is also not without its drawbacks. Its patch -oriented analysis can lead to situations where patches are locally similar, but globally the image is incoherent or semantically incorrect [16]. The most well-known problem with LPIPS is its vulnerability to adversarial attacks [15].

These discrepancies often arise because metrics "focus on the wrong thing." Human perception is flexible and can emphasize different aspects depending on the context, task, and importance of the information.

Comparing the performance of PSNR, SSIM, and LPIPS metrics is key to understanding their strengths and weaknesses in different practical scenarios. Their behavior varies significantly depending on the type of visual distortion present in the image.

These discrepancies often arise because metrics have a fixed "focus." Human perception, in contrast, is flexible and adapts to the context and task.

To quantitatively assess the correspondence of objective metrics to human perception, correlation coefficients with subjective estimates, such as MOS and DMOS, obtained as a result of psychophysical experiments are used [20].

A general trend observed in numerous studies on various databases (e.g. LIVE, TID2013, CSIQ, KADID-10k):

- LPIPS typically exhibits the highest correlation (measured, for example, by Spearman's rank correlation coefficient (SRCC) or Pearson's linear correlation coefficient (PLCC)) with MOS/DMOS, outperforming SSIM and PSNR [4].
- SSIM usually shows better correlation with human judgment than PSNR [2].
- PSNR often has the lowest correlation with human quality ratings.

Metric quality assessment on compression samples. To compare the evaluation of different metrics, we will take an image of 512x512 pixels and compress it using the fractal method (FIC). The sample for the study is the portrait of Isaac Newton, posted on Wikipedia. The image was compressed using the fractal method with the rank block size: 256, 128, 64, 32, 16, 8, 4. Accordingly, the quality of the compressed image improved with decreasing block size. Figure 1 shows a collage of compressed images and the original. The results of calculating the PSNR, SSIM, LPIPS metrics are given in Table 1.

Table 1. **Image compression quality metrics using the FIC method**

Rank block size (pix)	PSNR (dB)	SSIM	LPIPS
256	16.9385	0.4605	0.9095
128	20.0715	0.4985	0.7824
64	23.2162	0.5300	0.6997
32	26.5780	0.5849	0.6234
16	28.5952	0.6581	0.4810
8	30.5655	0.7714	0.2693
4	35.2494	0.9326	0.0642

Given the significant difference in the concept of metrics, they should be normalized for comparison. The range of PSNR values can be quite large, but we normalize the indicator to a value of 40 as the maximum, often, with such a quality of the resulting image, a person cannot distinguish the processed image from the original. The resulting indicator will take values from 0 to 1 (the higher, the better). SSIM takes values from 0 to 1, the higher, the better. We leave it unchanged. LPIPS takes values from 0 - 1, but the lower the better. Therefore, for comparison, the inverse value to 1 should be used. Comparative values are presented in the form of diagram 1.

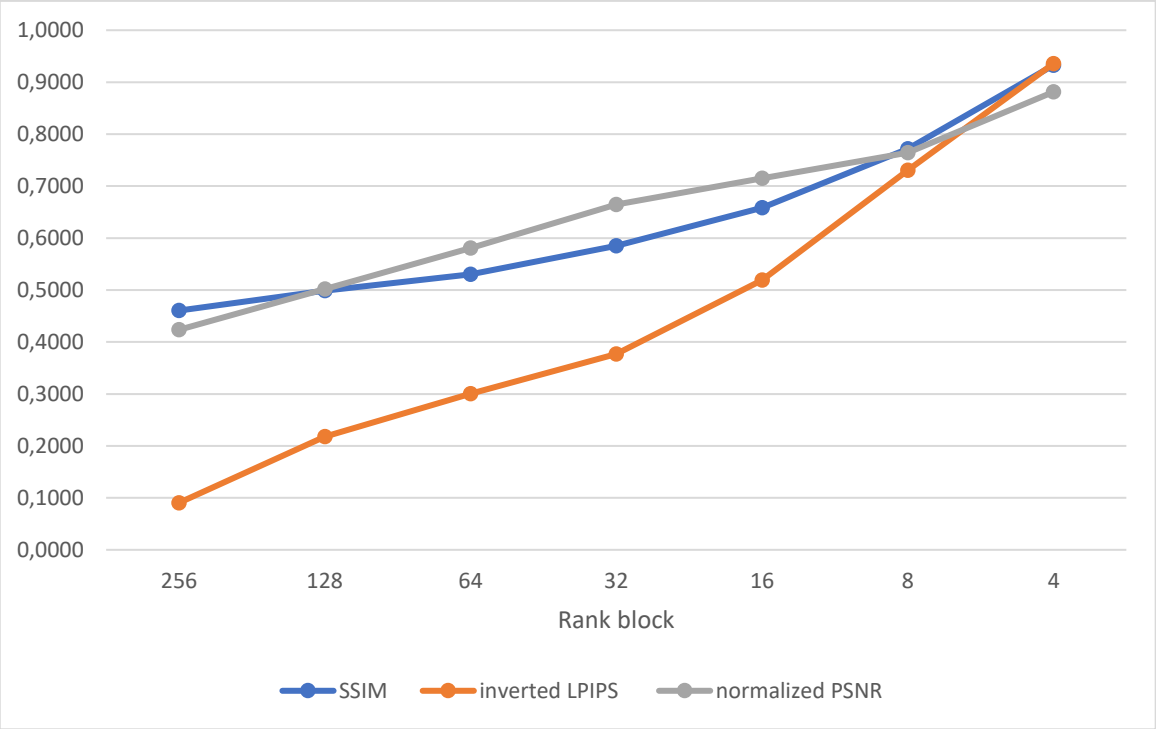

Diagram 1. **Comparison of SSIM, inverted LPIPS, normalized PSNR**

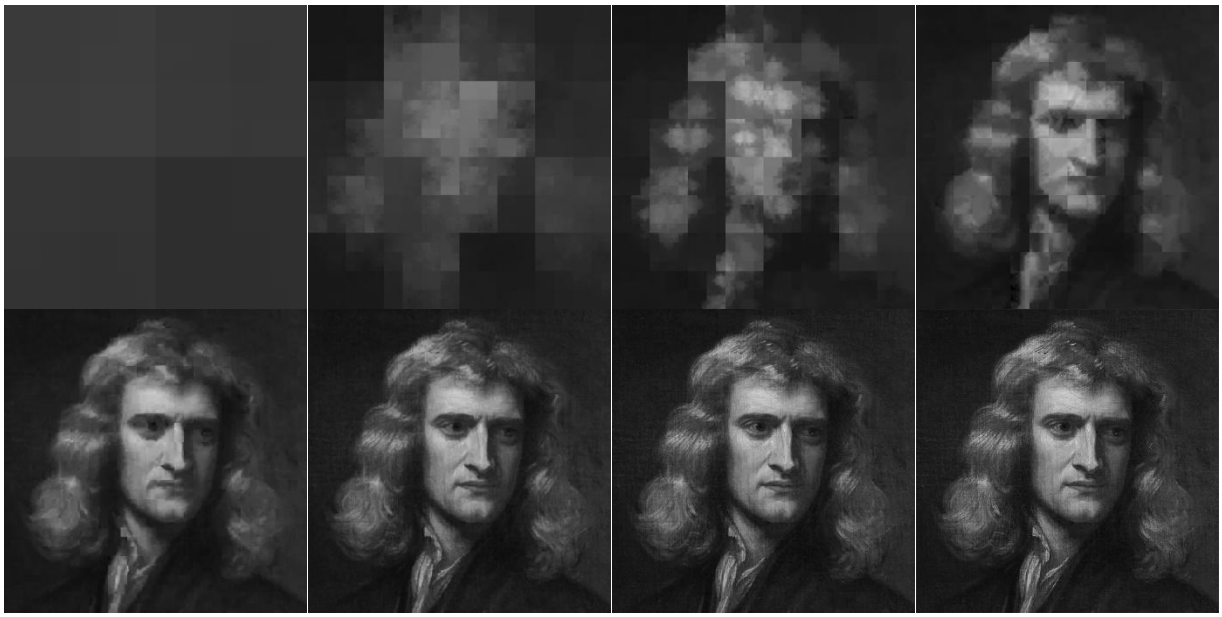

Fig. 1. ***Images*** **compressed by FIC with rank blocks: 256, 128, 64, 32, 16, 8, 4 and original**

As can be seen from Diagram 1 for PSNR metrics, even SSIM gray square is very similar to "Newton". The gray background, which occupies most of the original image, creates conditions for a high estimate of the similarity of PSNR and SSIM. PSNR steadily increases with decreasing block size (from 16.9 dB to 35.2 dB). The graph (gray line "normalized PSNR") shows an almost linear increase. This uniform increase does not reflect the real "jump" in quality that we see between blocks 32 and 16. PSNR equally "evaluates" the transition from illegible squares to a barely recognizable face. SSIM also increases with decreasing block size (from 0.46 to 0.93). Its curve (blue line) is slightly steeper than that of PSNR, which better reflects the improvement in quality in the last stages. Although SSIM is better

than PSNR, it still does not fully correlate with human perception, especially at low-quality stages. The "inverted LPIPS" curve grows slowly in the initial stages (when the image is indistinct) and goes up very sharply in the range from 32 to 4. This corresponds perfectly to visual analysis: the metric captures a slight improvement when the image remains "blocky" and gives a high score precisely at those steps where the image becomes recognizable and clear. LPIPS is noticeably better at assessing quality, which is confirmed by visual comparison of drawings.

Conclusions: Analysis of PSNR, SSIM, and LPIPS metrics revealed significant differences in their operating principles, advantages, disadvantages, and most importantly, in their ability to reflect human perceptions of image similarity and quality.

- PSNR, as the simplest of the considered metrics, is based on the per-pixel root mean square error. Its advantages are ease of calculation and clear interpretation for certain types of additive noise. However, a fundamental disadvantage of PSNR is its low correlation with human visual perception. It ignores structural information, semantic content, and complex LSR mechanisms such as masking effects. This leads to situations where images with the same PSNR can have radically different visual quality, as well as the inability to adequately assess the impact of many common distortions such as blurring or compression artifacts.
- SSIM is a significant step forward because it was developed taking into account that the LMS is highly adapted to extract structural information. By comparing the brightness, contrast, and structure of local image regions, SSIM shows a much better correlation with human judgment than PSNR. It is more sensitive to structural distortions that are visually significant. However, SSIM still has limitations: it can be insensitive to significant color changes if the structure is preserved, it does not cope well with strong blurring or small spatial shifts, and it does not sufficiently take into account global semantic information.
- LPIPS represents a modern approach based on deep learning. Using features extracted from pre-trained convolutional neural networks and training directly on large datasets of human perceptual judgments, LPIPS achieves the highest correlation with human perception among the considered metrics. It is able to capture more complex visual differences, including textural and structural nuances. Unlike the mathematical models PSNR and SSIM, LPIPS is "trained" to understand which changes in the image are important to a person. It ignores minor shifts or noise that the human eye does not pay attention to, but which greatly degrade the PSNR/SSIM indicators. LPIPS is able to better assess the preservation of objects and textures. In this example, it correctly determines that a significant object (face) appears at the "Block 16" stage, and "rewards" this with a significant improvement in the assessment. This metric is particularly useful for evaluating the performance of generative neural networks (GANs) and other deep learning algorithms, where the goal is to produce a realistic, rather than pixel- precise, image. However, LPIPS is not a panacea. Its main drawbacks include vulnerability to adversarial attacks, potential problems with estimating global semantic consistency through patch - oriented analysis, dependence on the architecture and training of the underlying CNN, and higher computational complexity.

The problem of human perception of similarity and its reflection in metrics lies in the fundamental complexity of HVS. Human perception is an active, interpretive process that takes into account a huge number of factors: from low-level features (color, brightness) through medium-level (texture, Gestalt principles) to high-level (semantics, context, previous experience). None of the existing objective metrics is able to fully capture this complexity.

Thus, the experimental results clearly demonstrate that for the evaluation of compression quality oriented to the end user (human), LPIPS is a much more informative and reliable metric than the classical PSNR and SSIM. LPIPS is the most perceptually relevant metric of the three considered for a wide range of distortions, followed by SSIM, and PSNR demonstrates the worst correspondence to human perception. However, even LPIPS can give counterintuitive results in certain scenarios, especially when assessing global semantic coherence or in the presence of atypical artifacts not represented in the training sample.

Future research directions will likely focus on developing even more sophisticated metrics that:

- Better integrate global semantic understanding (perhaps using transformative architectures).

- Are more robust to adversarial attacks and unknown types of distortions.
- They ensure better interpretability of their assessments.
- Take into account the specifics of the task and content (e.g. for medical imaging, AIGC).

In conclusion, while objective metrics are indispensable tools in image processing, it is important to remember their limitations and to complement their analysis with subjective human evaluation whenever possible, especially in critical applications. Understanding the principles of LMS operation and continuously improving metrics will bring us closer to creating tools that truly reflect the human vision of quality.

REFERENCES

1. Johnson, J., Alahi, A., & Fei-Fei, L. (2016, September). Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision* (pp. 694-711). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-46475-6_43.
2. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), 600-612. https://doi.org/10.1109/TIP.2003.819861.
3. Breger, A., Biguri, A., Landman, M. S., Selby, I., Amberg, N., Brunner, E., ... & Schönlieb, C. B. (2025). A study of why we need to reassess full reference image quality assessment with medical images. *Journal of Imaging Informatics in Medicine*, 1-26. https://doi.org/10.1007/s10278-025-01462-1
4. Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 586-595). https://arxiv.org/abs/1801.03924.
5. Arabboev, M., Begmatov, S., Rikhsivoev, M., Nosirov, K., & Saydiakbarov, S. (2024). A comprehensive review of image super-resolution metrics: classical and AI-based approaches. *Acta IMEKO*, *13*(1), 1-8. https://doi.org/10.21014/actaimeko.v13i1.1679.
6. Gertsiy, O. (2024). Research on graphic data formats for compact representation and comparison of images. *Transport systems and technologies*, (43), 173-187. https://doi.org/10.32703/2617-9059-2024-43-14.
7. Zhang, L., Zhang, L., Mou, X., & Zhang, D. (2011). FSIM: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, *20*(8), 2378-2386. https://doi.org/10.1109/TIP.2011.2109730.
8. Shrestha, B. (2005). Evaluation of JPEG2000 for lossless medical image compression. *Mississippi State University Libraries*. https://www.gri.msstate.edu/publications/docs/2005/03/4328BijayShrestha_2005.pdf.
9. Russ, J. C. (2006). *The image processing handbook*. CRC press.
10. Gonzalez, R.C., & Woods, R.E. (2017). *Digital Image Processing*. Pearson, New York, 4 editions.
11. Singh, G.K., Agarwal, A., & Reddy, N.V. (2023). Comparison of PSNR, SSIM, and LPIPS in medical imaging. In *2023 IEEE 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1–6). https:// www.researchgate. net / figure / LPIPS - SSIM - and - PSNR - Metrics - Comparison - This - table - evaluates - the - quality - of - generated _ tbl 2_382080667.
12. Wang, Z., Simoncelli, E. P., & Bovik, A. C. (2003, November). Multiscale structural similarity for image quality assessment. In *The thrity-seventh asilomar conference on signals, systems & computers, 2003* (Vol. 2, pp. 1398-1402). IEEE. https://doi.org/10.1109/ACSSC.2003.1292216.
13. Kuzovkin, I., Vicente, R., Petton, M., Lachaux, J. P., Baciu, M., Kahane, P., ... & Aru, J. (2018). Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. *Communications biology*, *1*(1), 107. https://doi.org/10.1038/s42003-018-0110-y.
14. Zhang, K., Liang, J., Van Gool, L., & Timofte, R. (2021). Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4791-4800). https://doi.org/10.48550/arXiv.2103.14006.
15. Zhai, G., & Min, X. (2020). Perceptual image quality assessment: a survey. *Science China Information Sciences*, *63*(11), 211301. https://doi.org/10.1007/s11432-019-2757-1.
16. Gu, S., Bao, J., Chen, D., & Wen, F. (2020, August). Giqa: Generated image quality assessment. In *European conference on computer vision* (pp. 369-385). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-58621-8_22.
17. Shoshan, A., Gandelsman, Yo., Bagon, S, & Dekel, T. (2024). R-LPIPS: An Adversarially Robust Perceptual Similarity Metric. *Scientific Reports*. https://doi.org/10.48550/arXiv.2307.15157.
18. Koffka, K. (1935). *Principles of Gestalt Psychology*. Harcourt, Brace & Co.
19. Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual informatio*. WH Freeman and Company.
20. International Telecommunication Union. (2019). *Methodology for the subjective assessment of the quality of television pictures* (Recommendation BT.500-14). https://www.itu.int/rec/R-REC-BT.500-14-201910-S/en russ

Карнатов Сергій[1]

[1]Аспірант, Кафедра Автоматизація та комп'ютерно-інтегровані технології транспорту, Національний транспортний університет, вул. Михайла Омеляновича-Павленка, 2, 01010, м. Київ, Україна. ORCID: https://orcid.org/0009-0006-6254-6166.

Аналіз метрик PSNR, SSIM, LPIPS у контексті людського сприйняття візуальної подібності.

Анотація. *Ця стаття пропонує комплексний порівняльний аналіз трьох відомих метрик оцінки якості зображення (IQA): PSNR, SSIM та LPIPS. У ній досліджуються їхні основні принципи, математичні основи, переваги та обмеження, зокрема, що стосуються їхньої кореляції зі зоровим сприйняттям людини. Обговорюється еволюція метрик IQA від простих попіксельних порівнянь (PSNR) до структурних підходів (SSIM) та, нещодавно, до метрик вивченого сприйняття (LPIPS). Представлено критичний аналіз ефективності кожної метрики в оцінці різних візуальних спотворень, включаючи шум, розмиття та артефакти стиснення. Притаманні людському зоровому сприйняттю проблеми, такі як роль семантики, текстури, кольору та візуальних артефактів, досліджуються як фундаментальні причини розбіжностей між об'єктивними метричними оцінками та суб'єктивними людськими судженнями. У статті висвітлюється «необґрунтована ефективність» глибоких ознак у LPIPS, а також розглядаються його вразливості, такі як атаки з боку суперників та обмеження в глобальному семантичному розумінні. Зрештою, у ньому окреслено напрямки майбутніх досліджень, спрямованих на розробку більш надійних, інтерпретованих та перцептивно узгоджених метрик IQA, які можуть краще враховувати складність зорової системи людини та мінливі вимоги сучасних технологій обробки зображень та генеративного штучного інтелекту.*

Ключові слова*: Оцінка якості зображень, PSNR, SSIM, LPIPS, людське сприйняття, візуальні спотворення, генеративні моделі, об'єктивні метрики, суб'єктивна оцінка.*