ACTA IMEKO
ISSN: 2221-870X
March 2024, Volume 13, Number 1, 1 - 9

A comprehensive review of image super-resolution metrics: classical and AI-based approaches

Mukhriddin Arabboev[1], Shohruh Begmatov[1], Mokhirjon Rikhsivoev[1], Khabibullo Nosirov[1], Saidakmal Saydiakbarov[2]

[1] *Department of Television and Radio Broadcasting Systems, TUIT named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan*
[2] *Faculty of Telecommunication Technologies, TUIT named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan*

ABSTRACT
Image super-resolution is a process that aims to enhance the quality and resolution of images using various techniques and algorithms. The process aims to reconstruct a high-resolution image from a given low-resolution input. To determine the effectiveness of these algorithms, it's crucial to evaluate those using specific metrics. In this paper, we take a closer look at the most commonly used image super-resolution metrics, including classical approaches like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Peak Signal to Noise Ratio (PSNR), and Structural Similarity Index (SSIM). We also discuss advanced metrics like Learned Perceptual Image Patch Similarity (LPIPS), Fréchet Inception Distance (FID), Inception Score (IS), and Multi-Scale Structural Similarity Index (MS-SSIM). Furthermore, we provide an overview of classical and AI-based super-resolution techniques and methods. Finally, we discuss potential challenges and future research directions in the field and present our experimental results by applying image super-resolution metrics. In the result and discussion section, we have practiced some given metrics and proposed our image super-resolution results.

Section: RESEARCH PAPER

Keywords: Image super-resolution; metrics; LPIPS; FID; IS; MSE; RMSE; SSIM; MS-SSIM; PSNR; AI-based methods

Citation: M. Arabboev, Sh. Begmatov, M. Rikhsivoev, Kh. Nosirov, S. Saydiakbarov, A comprehensive review of image super-resolution metrics: classical and AI-based approaches, Acta IMEKO, vol. 13 (2024) no. 1, pp. 1-8. DOI: 10.21014/acta_imeko.v13i1.1679

Section Editor: Laura Fabbiano, Politecnico di Bari, Italy

Received September 22, 2023; **In final form** February 29, 2024; **Published** March 2024

Corresponding author: Mukhriddin Arabboev, e-mail: mukhriddin.9207@gmail.com

1. INTRODUCTION

Image super-resolution (SR) is a process that enhances the resolution of low-resolution (LR) images to create high-resolution (HR) images [1]-[2]. This field rapidly evolves and has many applications in surveillance, remote sensing, medical imaging, and multimedia content production. However, acquiring high-resolution images directly can be difficult due to hardware limitations, imaging conditions, and other factors. SR overcomes this challenge by reconstructing HR images from LR inputs using sophisticated algorithms to infer and restore the missing details [3].

The demand for high-resolution images has surged in recent years due to the growing popularity of social media platforms, streaming services, and 4K/8K displays. In surveillance and remote sensing applications, SR can be used to improve the quality of LR images captured by security cameras and satellites, respectively, enabling more accurate object detection and tracking. In medical imaging, SR can enhance the resolution of MRIs and CT scans, leading to better diagnosis and treatment planning. In multimedia content production, SR can upscale LR images and videos to HR without sacrificing quality, providing consumers with a more immersive viewing experience.

Despite its many advantages, SR remains challenging, especially for real-time applications. SR algorithms must accurately and efficiently reconstruct HR images even when given noisy or incomplete LR inputs. Additionally, SR algorithms must be able to adapt to a wide range of image types and domains.

This paper provides a comprehensive overview of the most frequently used metrics for evaluating SR algorithms. We discuss the advantages and limitations of each metric in detail and guide in selecting the appropriate metrics for specific applications. We also review various SR methods and techniques, including classical and AI-based approaches. Our review meticulously outlines each method's unique characteristics and potential applications, offering valuable insights into the evolving landscape of SR technology.

2. CLASSICAL METRICS

Image super-resolution (SR) is a popular technique for reconstructing high-resolution images from low-resolution inputs [4]. Evaluating the performance of SR algorithms is essential for identifying their effectiveness and potential improvements. Classical methods for image SR fall under three categories: interpolation-based, reconstruction-based, and learning-based. Interpolation-based methods like nearest-neighbor, bilinear, and bicubic interpolation estimate the missing pixel values using nearby pixels. Reconstruction-based methods like projection onto convex sets (POCS) and iterative back-projection (IBP) formulate the SR problem as an inverse problem and try to minimize a cost function for finding an optimal solution. Learning-based methods like example and sparse coding-based SR learn a mapping between low-resolution and high-resolution images from training examples. These methods are widely used in image SR tasks but often have limitations such as blurring, ringing artifacts, and lack of fine details.

2.1. Mean Squared Error (MSE)

Mean Squared Error (MSE) [5] is a straightforward metric that calculates the average squared difference between the pixel values of two images. A pixel value is the numerical representation of the color or intensity stored for each pixel location in a digital image grid, as explained by [6]. Each pixel contains a value that represents its color or intensity, and in an 8-bit RGB image, each component of a pixel (red, green, and blue) can have 256 possible values (0-255). The collective pixel values encode the visual information that makes up the digital image. Modifying individual pixel values can alter colors, brightness, and other image properties, as [6] explains. It is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(I_1 - I_2)^2\,, \tag{1}$$

where I_1 and I_2 are the pixel values of the two images, and n is the total number of pixels. Lower MSE values indicate better image quality. The term "better image quality" refers to how clear and detailed a digital image is, showing how accurately it represents the original subject or scene that was captured. Several factors determine the quality of an image, such as resolution (higher resolution means more pixels and finer details), noise levels (less noise and fewer compression artifacts make the image appear better), color accuracy (images with colors that closely match the original scene are considered higher quality), sharpness (images with well-defined edges and no blurring), and artifacts (fewer distortions, aliasing effects, or other anomalies indicate better image quality) [7].

In summary, an image with higher resolution, lower noise, accurate colors, strong sharpness, and fewer artifacts is generally considered to have better quality compared to an image with more defects.

MSE is computationally efficient and easy to implement, making it a popular choice for evaluating SR algorithms. However, it can be sensitive to small changes in pixel values that may not be perceptually relevant, leading to poor correlation with human perception.

The mean squared error (MSE) is a commonly used and simple metric for measuring the quality of an image. It is calculated by finding the squared intensity differences between the distorted and reference images' pixels, then averaging them. The peak signal-to-noise ratio (PSNR) is also used to determine the related quantity [8].

2.2. Root Mean Squared Error (RMSE)

In research studies, the Root Mean Square Error (RMSE) is a widely accepted statistical metric used to measure model performance. Another useful measure commonly used in model evaluations is the mean absolute error (MAE). Despite being in use for many years, there is still no consensus on the most appropriate metric for model errors. However, in geosciences, the RMSE is often presented as the standard metric for model errors[9].

Also, RMSE is a variation of MSE that measures the square root of the average squared difference between corresponding pixels in two images [10]. The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(I_1 - I_2)^2}\,, \tag{2}$$

where n is the number of pixels, and I_1 and I_2 are the corresponding pixel values of the two images [10].

2.3. Peak Signal-to-Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio (PSNR) is a measure of fidelity that is used to evaluate the quality of an image or video. It is independent of the dynamic range of the media being evaluated. In the literature on learned image and video restoration and compression, the performance of methods is usually reported by the arithmetic mean of the PSNR of each image or frame in the test set. For example, the average PSNR of all images in the test set was reported for the well-known EDSR image super-resolution model [11].

PSNR is a widely used metric to measure the quality of reconstructed images by comparing the ground truth image's peak signal value with the super-resolution algorithm's error [12]. Mathematically, PSNR is defined as:

$$PSNR = 20 \cdot \log_{10}(MAX_I) - 10 \cdot \log_{10}(MSE), \tag{3}$$

where MAX_I refers to the maximum luminance value among the red, green, and blue components for an RGB color image. Specifically, in an 8-bit RGB image MAX_I would be 255. MSE is the mean squared error between the estimated HR image and ground truth HR image pixel luminance values. PSNR is measured in decibels (dB), with higher values indicating better image quality [13].

PSNR is commonly used to assess the performance of SR algorithms due to its computational efficiency and ease of implementation. However, it has been found to have limited relevance to human perception, as it fails to consider the structural and perceptual dissimilarities between images.

2.4. Structural Similarity Index (SSIM)

The Structural Similarity Index Measure (SSIM) [10] is a metric that evaluates the perceived quality of images by considering their structural information, luminance, and contrast. Typically, it compares two images - an estimated HR image and a ground truth HR image - by analyzing the mean, standard deviation, and cross-covariance of pixel intensities within local windows.

The SSIM method is based on human perception and considers image degradation a change in structural information. It also considers other important factors affecting perception, such as luminance and contrast masking. 'Structural information' refers to strongly interdependent or spatially close pixels, which provide important information about visual objects in the image domain. Luminance masking refers to the phenomenon where

distortion is less visible at the edges of an image. In contrast, contrast masking refers to the phenomenon where distortions are less visible in an image's texture.

SSIM estimates the similarity between two images or videos - the original and the recovered [5]. SSIM is defined as:

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma}, \quad (4)$$

where x and y are the two images and higher values indicate better image quality [12]. The equation uses different variables to compare two images. l stands for luminance, which is used to compare the brightness of the images. c is contrast, which helps to differentiate between the brightest and darkest regions of the images. s is structure and it compares the local luminance patterns of the images to determine their similarity or dissimilarity. Additionally, the equation includes three positive constants: α, β, and γ [14].

SSIM is a more appropriate metric than PSNR for evaluating the perceived quality of SR algorithms, as it correlates better with human perception. However, SSIM is computationally more demanding than PSNR, as it requires local statistics computation.

3. ADVANCED METRICS OF IMAGE SUPER-RESOLUTION

Image super-resolution algorithms aim to improve the resolution of low-quality images. Advanced metrics evaluate these algorithms' quality and efficiency and compare different techniques. These metrics are essential in determining the effectiveness and ability of super-resolution techniques to produce high-quality, visually pleasing results.

3.1. Learned Perceptual Image Patch Similarity (LPIPS)

Perceptual similarity metrics, or LPIPS, measure image similarity [11]. They use deep learning to extract features from images and calculate the distance between these features. This approach is based on the understanding that human perception of image quality goes beyond mere pixel-wise differences. However, LPIPS is vulnerable to adversarial attacks that produce results that do not align with human visual similarity judgment [15]. The features used in LPIPS are extracted using a deep convolutional neural network (CNN) pre-trained on an image classification task. The LPIPS metric is defined as follows:

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot \left(\hat{y}^l_{hw} - \hat{y}^l_{0hw} \right) \right\|_2^2 . \quad (5)$$

In Equation (5), we can see how the distance between reference and distorted patches x, x_0 are determined using network $\mathcal{F}$. We extract feature stacks from L layers and unit-normalize them in the channel dimension. This unit-normalized feature stack is designated as $\hat{y}^l, \hat{y}^l_0 \in \mathbb{R}^{H_l \times W_l \times C_l}$ for layer l. We then scale the activations channelwise using vector $w^l \in \mathbb{R}^{C_l}$, and calculate the ℓ_2 distance. Finally, we average spatially and sum channel-wise. It's worth noting that if we use $w_l = 1 \forall l$ for all layers, it's equivalent to computing cosine distance [11].

3.2. Fréchet Inception Distance (FID)

Fréchet Inception Distance (FID) is [16] a metric that quantifies the similarity between two sets of images based on their features. It is commonly used to evaluate the quality of images generated by Generative Adversarial Networks (GANs).

A pre-trained Inception network is used to extract image features, which are then modelled as multivariate Gaussian distributions to calculate FID. The feature distributions of two sets of images, such as generated and ground truth HR images, are compared to measure the distance between them. The FID is then calculated as the Fréchet distance between these two distributions:

$$FID = \|\mu_r - \mu_g\|^2 + T_r\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r \Sigma_g\right)^{1/2}\right). \quad (6)$$

The Fréchet distance is a measure of similarity between two multivariate normal distributions. It is calculated based on the means μ_r, μ_g and covariance matrices Σ_r, Σ_g [17].

Where μ_r and Σ_r are the mean and covariance of features from the real dataset and μ_g and Σ_g are the mean and covariance of features from the generated dataset [16].

The first term $\|\mu_r - \mu_g\|^2$ calculates the distance between the means of the two distributions. c

The second term $T_r\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r \Sigma_g\right)^{1/2}\right)$ calculates the distance between the covariances. The covariances describe the shape and spread of the distributions. Taking the trace measures the total variance between the two distributions. A lower FID score means that the generated images more closely match the statistics of real images, as the means and covariance's of the distributions are closer. Lower FID values indicate higher image quality. Unlike simpler metrics such as MSE and PSNR, FID has been proven to correlate strongly with human perception. However, it is more computationally expensive due to its dependence on a pre-trained deep neural network and the calculation of multivariate Gaussian distributions [18].

3.3. Inception Score (IS)

GANs commonly use the Inception Score (IS) to evaluate image quality and diversity [11]. The IS measures the conditional label distribution predicted by the Inception network. The IS is defined as:

$$IS(G) = \exp\left(E_{x \sim p_g} D_{KL}\left(p(yvertx) \parallel p(y)\right)\right), \quad (7)$$

where signifies that x is an image drawn from the distribution, represents the D_{KL}-divergence between the distributions, denotes the conditional class distribution, and represents the marginal class distribution. The term exp in the calculation is included to simplify value comparison. However, for the sake of convenience, we can disregard it without impacting the overall interpretation, thus utilizing [19].

3.4. Multi-Scale Structural Similarity Index (MS-SSIM)

MS-SSIM, or Multi-Scale Structural Similarity Index, compares two images to determine their similarity. It is an extension of the original Structural Similarity (SSIM) index developed as a classical method. However, MS-SSIM is a hybrid method combining classical and AI-based techniques to achieve its results.

MS-SSIM employs classical image processing techniques to determine the structural similarity between two images at multiple scales. Additionally, it uses machine learning techniques to optimize the different scales' weighting and improve the index's accuracy.

In essence, MS-SSIM is an image quality metric that uses classical and AI-based techniques to accurately measure the similarity between two images [13]. The MS-SSIM is defined as:

$$MS - \text{SSIM}\ (X, Y) = \frac{1}{M} \sum_{j=1}^{M} \text{SSIM}\left(x_j, y_j\right), \quad (8)$$

where x and y are the compared images, M is the number of scales [20].

4. CLASSICAL AND AI-BASED SUPER-RESOLUTION TECHNIQUES AND METHODS

4.1. Classical super-resolution techniques

Super-resolution techniques enhance the resolution of an image or a sequence of images, reconstructing a high-resolution (HR) image from one or more low-resolution (LR) images. These techniques can be divided into single-image super-resolution (SISR) and multi-image super-resolution (MISR). SISR techniques improve the resolution of a single low-resolution image by estimating the missing high-frequency details. MISR techniques use multiple low-resolution images of the same scene to reconstruct a high-resolution image. Classical super-resolution techniques can be interpolation-based (e.g., bicubic interpolation, resampling) or model-based (e.g., sparse coding, dictionary learning) [3].

4.2. AI-based super-resolution techniques

Super-resolution techniques that use artificial intelligence (AI) rely on deep learning, especially convolutional neural networks (CNNs) and generative adversarial networks (GANs) [21]. Some examples of these techniques are SRCNN [22], EDSR [23], and SRGAN [24].

Deep learning-based AI methods have gained significant attention in recent years due to their superior performance in various image-processing tasks, including super-resolution (SR2). CNNs have been the most popular choice for image SR tasks, and architectures like SRCNN, VDSR, and EDSR have achieved state-of-the-art performance. These methods use many convolutional layers to learn a mapping between the low-resolution (LR) and high-resolution (HR) images.

GANs have also been used for image SR tasks, and methods like SRGAN and ESRGAN have achieved impressive results. GANs consist of a generator and a discriminator, where the generator aims to generate realistic HR images from the LR input, and the discriminator tries to distinguish between real and generated images. The generator and discriminator are trained simultaneously, improving the generated images' perceptual quality.

In addition to CNNs and GANs, other deep learning-based methods, such as recurrent neural networks (RNNs) and attention mechanisms, have also been employed for image SR tasks. These methods have shown promising results, further demonstrating the potential of AI-based techniques for image SR.

5. COMPARISON OF METRICS

Our review the image quality metrics discussed earlier and evaluate their strengths, weaknesses, and suitability for various super-resolution scenarios. To make it easy to compare, we have summarized the comparison in Table 1.

5.1. Classical metrics vs. AI-based metrics

Classical metrics such as MSE, RMSE, SSIM, and PSNR are commonly utilized as they are simple and easy to compute. However, they often do not align well with human perception, especially for deep learning-based SR techniques [8].

AI-based metrics such as LPIPS, FID, and IS have been developed to address the limitations of classical metrics. Although these metrics are more complicated and computationally expensive, they have been shown to better correlate with human perception for a wide range of image processing tasks including super-resolution.

Table 1. A comparison of image quality metrics for super-resolution.

Metric	Strengths	Weaknesses	Suitable for
MSE/ RMSE	Easy to compute, widely used	Poor correlation with human perception	Simple comparisons, when human perception is not critical
SSIM/ MS-SSIM	Incorporates structural information, a better correlation with human perception than MSE	Computationally expensive, not always suitable for deep learning-based methods	Comparing methods with different structural properties
PSNR	Easy to compute, widely used	Poor correlation with human perception	Simple comparisons, when human perception is not critical
LPIPS	Good correlation with human perception, robust to noise	Requires pre-trained deep network, computationally more expensive than classical metrics	Evaluating deep learning-based methods, when human perception is important
FID	Compares feature distributions, correlates well with human perception	Requires pre-trained network, sensitive to noise	Evaluating generated image quality, when human perception is critical
IS	Measures quality and diversity, easy to compute	Sensitive to noise, not always correlated with human perception	Evaluating generated image diversity, when human perception is not critical

When assessing SR techniques, it is crucial to consider the specific application's requirements. AI-based metrics like LPIPS and FID may be more appropriate if human perception is critical. However, if computational efficiency or simplicity is the main concern, classical metrics such as MSE and PSNR may be more suitable.

In Table 2, we reviewed authors' papers that related super-resolution metrics and researched the average results of authors who used image super-resolution metrics in the dataset. The table reports the average results of state-of-the-art image super-resolution methods on common benchmark datasets like Set5 and Set 14. It shows a comparison of techniques based on different commonly used evaluation metrics - PSNR, SSIM, MSE, LPIPS, and MS-SSIM. Methods include Bicubic interpolation, A+, JOR, RFL, SelfEx, and others that established foundations using optimizations or sparse representations. More recent deep models like RED, DnCNN, TNRD, FDSR, SRCN, ESRGAN, EDSR, and SRGAN are also included. Not all methods report results on the full suite of metrics. PSNR and SSIM remain the most widely reported to analyze reconstruction quality, while LPIPS and MS-SSIM capture perceptual aspects. The table provides an overview of average performance for different supervised approaches based on various evaluation criteria on standard benchmarks. It summarizes progress in image super-resolution and helps identify the most suitable method based on the evaluation metric of interest. FID and IS datasets were not compared with other papers.

Table 2. The average results of the authors used image super-resolution metrics in the dataset (Set 5 and Set 14).

Model	PSNR	SSIM	MSE	LPIPS	MS-SSIM
Bicubic [22]	✓	✓	-	-	✓
A+ [25]	✓	✓	-	-	✓
JOR [26]	✓	✓	-	-	-
RFL [27]	✓	✓	-	-	-
SelfEx [28]	✓	✓	-	-	-
CSCN [29]	✓	✓	-	-	-
RED [30]	✓	✓	-	-	-
DnCNN [31]	✓	✓	-	-	-
TNRD [32]	✓	✓	-	-	-
FDSR [33]	✓	✓	-	-	-
SRCN [34]	✓	✓	✓	-	✓
ESRGAN [35]	✓	✓	-	✓	-
EDSR [23]	✓	✓	-	✓	-
SRGAN [24]	✓	✓	-	✓	-
SC [36]	✓	✓	-	-	✓
ANR [33]	✓	✓	-	-	✓

Table 3 provides a comprehensive overview of the training results reported by various authors for single-image super-resolution methods on the Set 14 dataset. The table compares the performance of state-of-the-art approaches for different upsampling scale factors, including x2, x3, and x4, using widely adopted full reference metrics, PSNR (higher is better) and SSIM (higher is better). The comparison includes a range of seminal deep learning-based methods, such as CNN [22], SRGAN [23], ANR [26], and SRCNN [37], representing early works in the field.

More recent approaches like VDSR [30], DnCNN-3 [30], and TNRD [30] have improved upon earlier methods by achieving better quantitative results. By reporting results on standard datasets, including Set14, this comparison table allows evaluation of progress made over time in advancing the field through deep learning models. It summarizes improvements based on reference metrics, allowing for a quantitative analysis of gains for different upsampling ratios. In Figure 1, we illustrated the authors' and our comparison results of PSNR and SSIM with other methods via graphs/pictures.

We conducted a study shown in Figure 2, where we trained using Set 14 and compared our results with the Bicubic and CRCNN methods used by other authors. Our approach involved using Low-Resolution Images to produce High-resolution pictures. Our results, as depicted in Figure 1 and Figure 2, showed better PSNR values of 30.06 and SSIM of 0.91. We suggest future research in this field should focus on exploring novel AI-based techniques, incorporating domain-specific knowledge, and investigating unsupervised and semi-supervised learning approaches [31].

6. CONCLUSION

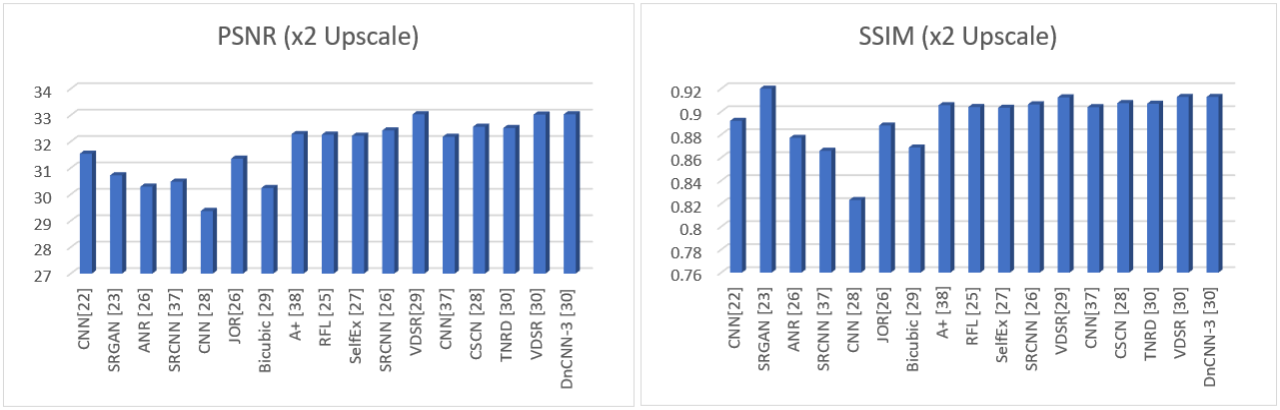
This research presents a comprehensive review of popular image super-resolution techniques and metrics. The study covers both classical and modern deep learning-based approaches, which were examined in detail and evaluated experimentally. A variety of metrics assessing perceptual and computational aspects were compared.

Although no single metric emerged as definitively superior, specific trends were identified. Classical metrics such as MSE, PSNR, and SSIM remain widely adopted due to their simplicity and efficiency. However, perceptual metrics like LPIPS and FID may better gauge human perception where it is paramount. Hybrid techniques combining favorable properties also showed promise.

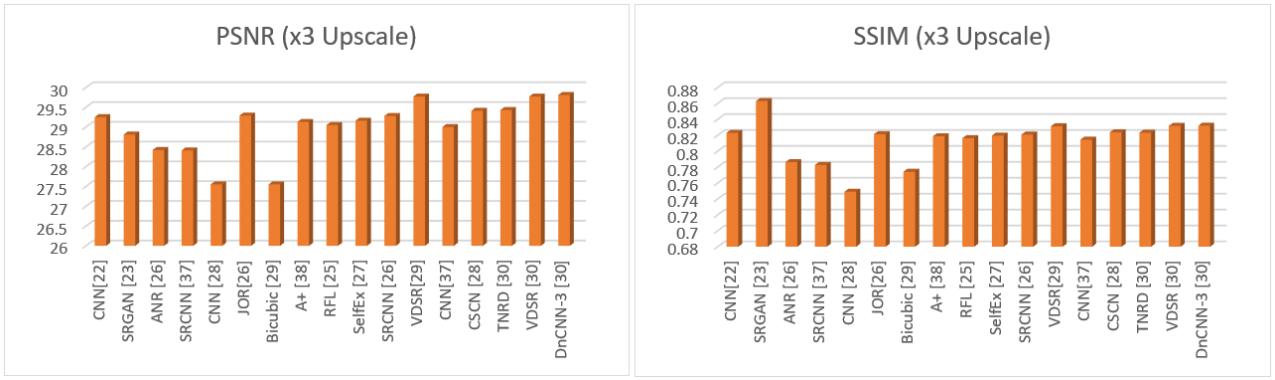
The appropriate choice of metric depends on specific application needs and goals, ranging from computational efficiency to perceptual accuracy demands. As the field progresses with novel approaches, continued examination and advancement of objective evaluation measures hold

Table 3. The training results of the authors used image super-resolution metrics in the dataset (Set 14).

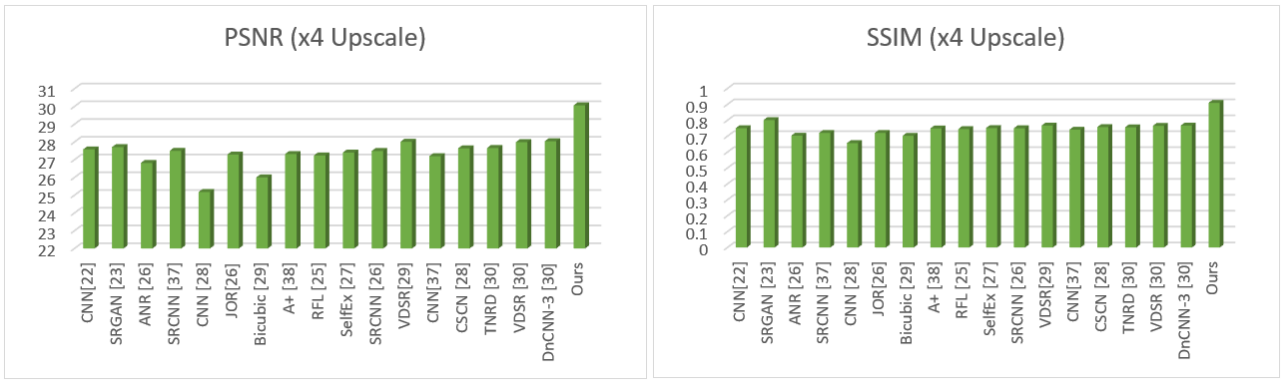
Methods	Dataset	PSNR (x2 Upscale)	SSIM (x2 Upscale)	PSNR (x3 Upscale)	SSIM (x3 Upscale)	PSNR (x4 Upscale)	SSIM (x4 Upscale)
CNN [22]	Set 14	31.54	0.8920	29.25	0.8230	27.58	0.7510
SRGAN [23]	Set 14	30.72	0.9200	28.81	0.8630	27.71	0.8020
ANR [26]	Set 14	30.29	0.8772	28.42	0.7863	26.82	0.7040
SRCNN [37]	Set 14	30.48	0.8660	28.41	0.7827	27.50	0.7212
CNN [28]	Set 14	29.37	0.8231	27.55	0.7491	25.18	0.6570
JOR [26]	Set 14	31.35	0.8879	29.29	0.8214	27.29	0.7208
Bicubic [29]	Set 14	30.24	0.8688	27.55	0.7742	26.00	0.7027
A+ [38]	Set 14	32.28	0.9056	29.13	0.8188	27.32	0.7491
RFL [25]	Set 14	32.26	0.9040	29.05	0.8164	27.24	0.7451
SelfEx [27]	Set 14	32.22	0.9034	29.16	0.8196	27.40	0.7518
SRCNN [26]	Set 14	32.42	0.9063	29.28	0.8209	27.49	0.7503
VDSR [29]	Set 14	33.03	0.9124	29.77	0.8314	28.01	0.7674
CNN [37]	Set 14	32.18	0.9039	29.00	0.8145	27.20	0.7413
CSCN [28]	Set 14	32.56	0.9074	29.41	0.8238	27.64	0.7578
TNRD [30]	Set 14	32.51	0.9069	29.43	0.8232	27.66	0.7563
VDSR [30]	Set 14	33.02	0.9128	29.77	0.8318	27.99	0.7659
DnCNN-3 [30]	Set 14	33.03	0.9128	29.81	0.8321	28.04	0.7672
Ours	Set 14	-	-	-	-	30.06	0.91



a)



b)



c)

Figure 1. Authors' comparison results on PSNR and SSIM using Set 14. a) the results of PSNR and SSIM in x2 upscaling factor; b) PSNR and SSIM results in x3 upscaling factor; c) the authors' and our comparison results of PSNR and SSIM in x4 upscaling factor.

importance-metrics reflecting human perception while maintaining efficiency warrant ongoing research attention.

Several open challenges, such as developing robust super-resolution models handling diverse conditions and noise, also require addressing. By synthesizing key findings, this work aimed to guide metric selection and highlight important directions for future work in advancing objective assessment and algorithm design.

ACKNOWLEDGMENT

We acknowledge the contribution of the ITMade innovation laboratory at the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi.

REFERENCES

[1] V. K. Ha, J.C. Ren, X.Y. Xu, S. Zhao, G. Xie, V. Masero, A. Hussain, Deep Learning Based Single Image Super-resolution: A Survey, Int. J. Autom. Comput., vol. 16, no. 4, 2019, pp. 413–426. DOI: 10.1007/s11633-019-1183-x

[2] X. Liu, S. Chen, L. Song, M. Woźniak, S. Liu, Self-attention negative feedback network for real-time image super-resolution, J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 8, 2022, pp. 6179–6186.
DOI: 10.1016/j.jksuci.2021.07.014

[3] X. Wang, J. Yi, J. Guo, Y. Song, J. Lyu, J. Xu, W. Yan, J. Zhao, Q. Cai, H. Min, A Review of Image Super-Resolution Approaches Based on Deep Learning and Applications in Remote Sensing, Remote Sens., vol. 14, no. 21, 2022, pp. 1–34.
DOI: 10.3390/rs14215423

[4] M. Rikhsivoev, A. Yusupov, S. Begmatov, M. Arabboev, K. Nosirov, S. Saydiakbarov, S. Vakhkhobov, Z. Khamidjonov, Working principles of multi-frame image super-resolution, Science

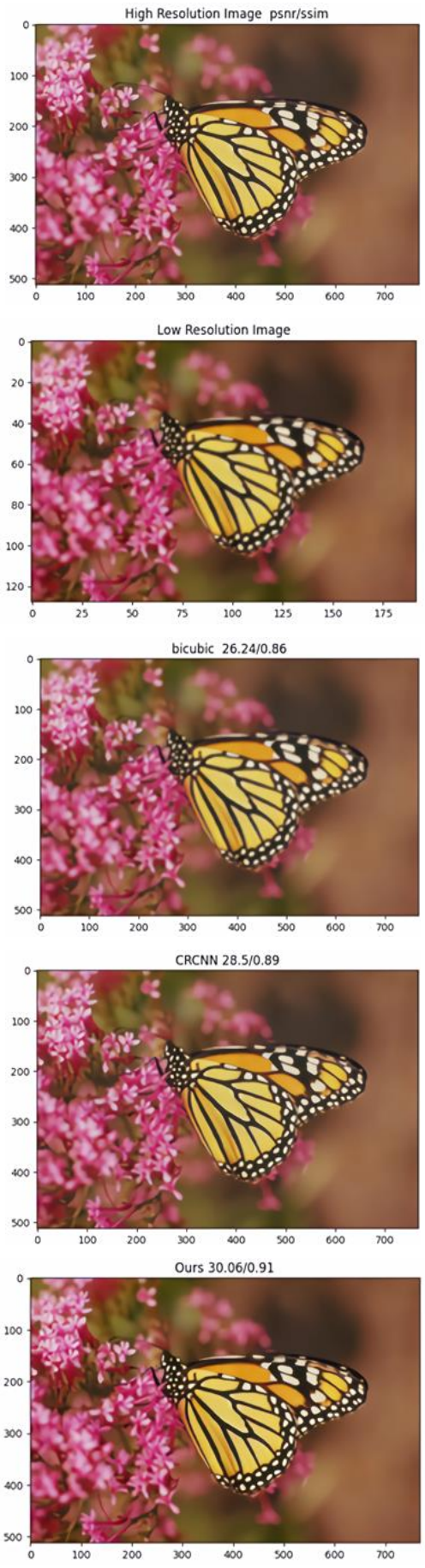


Figure 2. Results of our method with comparison to other image super-resolution outputs.

and innovation, vol. 2, no. Special Issue 3, 2023, pp. 244-249. DOI: 10.5281/zenodo.7856110

[5] S. S. Channappayya, A. C. Bovik, C. Caramanis, R. W. Heath, Design of linear equalizers optimized for the structural similarity index, IEEE Trans. Image Process., vol. 17, no. 6, 2008, pp. 857–872. DOI: 10.1109/TIP.2008.921328

[6] R. Fisher, S. Perkins, A. Walker, E. Wolfart, Hypermedia Image Processing Reference (HIPR), Artificial Intelligence, 01, 1996, pp. 1–317.

[7] J. M. Todd, Digital image processing (second edition), 1988. DOI: 10.1016/0143-8166(88)90012-7

[8] U. Sara, M. Akter, M. S. Uddin, Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study, J. Comput. Commun., vol. 07, no. 03, 2019, pp. 8–18. DOI: 10.4236/jcc.2019.73002

[9] T. Chai, R. R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)?, Geosci. Model Dev., vol. 7, no. 3, 2014, pp. 1247–1250. DOI: 10.5194/gmd-7-1247-2014

[10] M. Arabboev, S. Begmatov, K. Nosirov, J. C. Chedjou, K. Kyamakya, Development of a novel method of adaptive image interpolation for image resizing using artificial intelligence, Proc. of the IVUS 2022: 27th Int. Conf. on Information Technology, 12 May, 2022, Kaunas, Lithuania, pp. 32–38.

[11] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, Proc. of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-23 June 2018, pp. 586-595. DOI: 10.1109/CVPR.2018.00068

[12] Y. Wang, X. Li, H. Ma, Q. Ma, Q. Ding, M. Pirouz, Image super-resolution reconstruction based on improved generative adversarial network, Journal of Network Intelligence, vol. 6, no. 2, 2021, pp. 154-163.

[13] M. Arabboev, S. Begmatov, K. Nosirov, S. Tashmetov, S. Saydiakbarov, J. C. Chedjou, K. Kyamakya, Development of a Novel Method for Image Resizing Using Artificial Neural Network, Lecture Notes in Computer Science, vol 13741, 2023, pp. 527–539. DOI: 10.1007/978-3-031-27199-1_53

[14] R. Kumar, V. Moyal, Visual Image Quality Assessment Technique using FSIM, Int. J. Comput. Appl. Technol. Res., vol. 2, no. 3, 2013, pp. 250–254. DOI: 10.7753/ijcatr0203.1008

[15] M. Kettunen, E. Härkönen, J. Lehtinen, E-LPIPS: Robust Perceptual Image Similarity via Random Transformation Ensembles, arXiv, 2019. DOI: 10.48550/arXiv.1906.03973

[16] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, GANs trained by a two time-scale update rule converge to a local Nash equilibrium, Proc. of the 31st Int. Conf. on Neural Information Processing Systems, Long Beach, CA, USA, 4-9 December, 2017, pp. 6629–6640. DOI: 10.18034/ajase.v8i1.9

[17] D. C. Dowson, B. V. Landau, The Fréchet distance between multivariate normal distributions, J. Multivar. Anal., vol. 12, no. 3, 1982, pp. 450–455. DOI: 10.1016/0047-259X(82)90077-X

[18] S. Ghazanfari, S. Garg, P. Krishnamurthy, F. Khorrami, A. Araujo, R-LPIPS: An Adversarially Robust Perceptual Similarity Metric, arXiv, 2023. DOI: 10.48550/arXiv.2307.15157

[19] S. Barratt, R. Sharma, A Note on the Inception Score, arXiv, 2018. DOI: 10.48550/arXiv.1801.01973

[20] Z. Wang, E. Simoncelli, A. C. Bovik, Multiscale structural similarity for image quality assessment, Proc. of the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9-12 November 2003, pp. 1398-1402. DOI: 10.1109/ACSSC.2003.1292216

[21] S. Begmatov, M. Arabboev, K. Nosirov, M. Rikhsivoev, Review Types of GAN Method for Image Super-Resolution, Science and innovation, vol. 2, no. Special Issue 3, 2023, pp. 175-179. DOI: 10.5281/zenodo.7854589

[22] Y. Li, D. Liu, H. Li, L. Li, Z. Li, F. Wu, Learning a Convolutional Neural Network for Image Compact-Resolution, IEEE Trans. Image Process., vol. 28, no. 3, 2019, pp. 1092–1107. DOI: 10.1109/TIP.2018.2872876

[23] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21-26 July 2017, pp. 105-114. DOI: 10.1109/CVPR.2017.19

[24] J. Yang, J. Wright, T. S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process., vol. 19, no. 11, 2010, pp. 2861–2873. DOI: 10.1109/TIP.2010.2050625

[25] S. Schulter, C. Leistner, H. Bischof, Fast and accurate image upscaling with super-resolution forests, Proc. of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7-12 June 2015, 3791–3799. DOI: 10.1109/CVPR.2015.7299003

[26] D. Dai, R. Timofte, L. Van Gool, Jointly Optimized Regressors for Image Super-resolution, Comput. Graph. Forum, vol. 34, no. 2, 2015, pp. 95–104. DOI: 10.1111/cgf.12544

[27] J. Bin Huang, A. Singh, N. Ahuja, Single image super-resolution from transformed self-exemplars, Proc. of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7-12 June 2015, pp. 5197-5206. DOI: 10.1109/CVPR.2015.7299156

[28] Z. Wang, D. Liu, J. Yang, W. Han, T. Huang, Deep networks for image super-resolution with sparse prior, Proc. of the 2015 IEEE Int. Conf. on Computer Vision (ICCV), Santiago, Chile, 7-13 December 2015, pp. 370-378. DOI: 10.1109/ICCV.2015.50

[29] J. Kim, J. K. Lee, K. M. Lee, Accurate image super-resolution using very deep convolutional networks, Proc. of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27-30 June 2016, pp. 1646-1654. DOI: 10.1109/CVPR.2016.182

[30] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising, IEEE Trans. Image Process., vol. 26, no. 7, 2017, pp. 3142–3155. DOI: 10.1109/TIP.2017.2662206

[31] Y. Chen, T. Pock, Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration, IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, 2017, pp. 1256–1272. DOI: 10.1109/TPAMI.2016.2596743

[32] Z. Lu, Z. Yu, P. Yali, L. Shigang, W. Xiaojun, L. Gang, R. Yuan, Fast Single Image Super-Resolution Via Dilated Residual Networks, IEEE Access, vol. 7, 2019, pp. 109729–109738. DOI: 10.1109/ACCESS.2018.2865613

[33] C. Dong, C. C. Loy, K. He, X. Tang, Image Super-Resolution Using Deep Convolutional Networks, IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, 2016, pp. 295–307. DOI: 10.1109/TPAMI.2015.2439281

[34] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C. C. Loy, ESRGAN: Enhanced super-resolution generative adversarial networks, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11133 LNCS, 2019, pp. 63–79. DOI: 10.1007/978-3-030-11021-5_5

[35] B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, Enhanced Deep Residual Networks for Single Image Super-Resolution, Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21-26 July 2017, pp. 1132-1140. DOI: 10.1109/CVPRW.2017.151

[36] R. Timofte, V. De, L. Van Gool, Anchored neighborhood regression for fast example-based super-resolution, Proc. of the 2013 IEEE Int. Conf. on Computer Vision, Sydney, NSW, Australia, 1-8 December 2013, pp. 1920-1927. DOI: 10.1109/ICCV.2013.241

[37] D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars, Learning a deep convolutional network for image super-resolution, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8692, 2014, pp. 184–199. DOI: 10.1007/978-3-319-10593-2_13

[38] R. Timofte, V. De Smet, L. Van Gool, A+: Adjusted anchored neighborhood regression for fast super-resolution, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9006, 2015, pp. 111–126. DOI: 10.1007/978-3-319-16817-3_8