



ANALYSIS OF THE IMPACT OF RGB-TO-ACHROMATIC COLOR SPACE TRANSFORMATIONS ON SINGLE-IMAGE SUPER-RESOLUTION PERFORMANCE

Hürkal HÜSEM^{1*}, Zeynep GÜRKAŞ AYDIN², Önder DEMİR³

¹Istanbul University-Cerrahpaşa, Institute of Graduate Studies, Department of Computer Engineering, 34320, Istanbul, Türkiye

²Istanbul University-Cerrahpaşa, Faculty of Engineering, Department of Computer Engineering, 34320, Istanbul, Türkiye


³Marmara University, Faculty of Technology, Department of Computer Engineering, 34840, Istanbul, Türkiye


Abstract: Super-resolution techniques are employed to enhance the quality of digital images. Color spaces are developed to model colors in various digital environments. In the literature, several studies suggest that applying color space transformations and subsequently employing super-resolution techniques on the transformed images improve image quality. This study analyzes the impact of color space transformations on super-resolution applications. The analysis is conducted by performing the super-resolution process entirely in the RGB color space, followed by converting the obtained result into a different color space and comparing the quality metrics. The findings reveal that it is possible to achieve higher scores by converting RGB images into YCbCr or CIELab color spaces, despite no actual improvement in perceived image quality. Our experiments involve applying image enhancement techniques solely within the RGB color space, converting the results into alternative color spaces, and comparing them with ground truth images in Set5, Set14, BSDS100, Urban100, and DIV2K. Working in color spaces other than RGB does not lead to significant visual quality improvement. Our experiments demonstrate that solely through color space conversion, traditional metrics such as PSNR and SSIM, as well as deep learning-based metrics like DISTS and A-DISTS, can yield higher scores. Therefore, the observed improvements in quality metrics resulting from color space transformations may be misleading and may not reflect actual enhancements in image fidelity. With the A-DISTS metric that evaluates human perception, our study examines not only the impact of transformations from RGB to alternative color spaces on metrics but also evaluates the alignment of these metrics with human perception, an area that has received limited attention in the literature.


Keywords: Single-image super-resolution, Achromatic axis, Color space conversion, Image enhancement

*Corresponding author: Istanbul University-Cerrahpaşa, Institute of Graduate Studies, Department of Computer Engineering, 34320, Istanbul, Türkiye

E mail: contact@hurkal.com (H. HÜSEM)

Hürkal HÜSEM  <https://orcid.org/0000-0002-5414-6481>

Zeynep GÜRKAŞ AYDIN  <https://orcid.org/0000-0002-4125-0589>

Önder DEMİR  <https://orcid.org/0000-0003-4540-663X>

Received: December 06, 2024

Accepted: January 15, 2025

Published: March 15, 2025

Cite as: Hüsem H, Gürkaş-Aydın Z, Demir Ö. 2025. Analysis of the impact of RGB-to-achromatic color space transformations on single-image superresolution performance. *BSJ Eng Sci*, 8(2): 330-340.

1. Introduction

Image data captured by imaging systems, such as digital cameras, can be enhanced using various methods. Super-resolution aims to improve the data obtained by image detection systems with various techniques (Candès and Fernandez-Granda, 2014). With resolution enhancement techniques, it is possible to use more economical solutions instead of more expensive optical systems that can provide shorter exposure or high-quality images.

Color spaces are mathematically defined models developed to describe colors to determine which components can be used to obtain color. It is important to choose the appropriate color space according to the characteristics of the environment in which the image will be processed or printed to obtain true-to-life colors (Yilmaz et al., 2002). Although most of the resolution improvement studies work with RGB color space, which is modeled according to the intensity of red, green, and blue color components, it has been observed that color

spaces including achromatic data along with color axes give successful results in resolution improvement studies (Dong et al., 2016; John et al., 2016; Z. Wang et al., 2021). According to Z. Wang et al. (2021), RGB color space is frequently used compared to other color spaces, and improvements made on the Y component of YCbCr (or YCC) color space in previous models are mentioned. It has been revealed that models trained using different color spaces with luminance axis other than RGB can make a significant difference in success (Z. Wang et al., 2021). John et al. (2016) show that YCbCr and CMYK models produce better PSNR results and stated that CIELab color space could also be used in resolution enhancement (John et al., 2016). Dong et al. (2016) highlighted the superior performance of the Y channel within the YCbCr color space compared to training across all axes, yet demonstrated that the best results were achieved using the RGB color space (Dong et al., 2016). Gong et al. (2017) similarly propose the use of the classified dictionary learning method in super-resolution



processes, recommending the L^* coordinate in the CIE Lab color space and the Y coordinate in the YIQ system (Gong et al., 2017).

Single-image super-resolution (SISR) techniques aim to enhance the resolution of a single low-resolution image, a critical task in various image processing applications. A common practice in SISR is to convert RGB images into alternative color spaces, such as YCbCr and CIE Lab, assuming that this conversion improves image quality. However, the efficacy of this approach remains unclear, prompting our investigation.

Unlike previous works, which primarily focus on quantitative assessments using metrics like PSNR and SSIM, we scrutinize the perceptual quality of images resulting from color space transformations. Our methodology involves applying image enhancement techniques exclusively in the RGB color space, followed by conversion into YCbCr and CIE Lab for comparison with the original images. By doing so, we aim to elucidate whether the observed improvements in scores translate to tangible enhancements in visual quality.

Instead of applying resolution enhancement techniques on completely different color spaces, the color spaces of the input and output images are processed in RGB format as usual, then converted after the enhancement process. It has been demonstrated that color space transformation, when excluded from the resolution enhancement process, may independently appear to improve performance.

This study uses the Set5 (Bevilacqua et al., 2012), Set14 (Zeyde et al., 2012), BSDS100 (Martin et al., 2001), and Urban100 (Huang et al., 2015), DIV2K with unknown degradation (Agustsson and Timofte, 2017) datasets, which are frequently used in resolution enhancement studies. All images within the datasets consist of real-world high-resolution visuals paired with their corresponding synthetically generated low-resolution counterparts (Su et al., 2024).

Section II discusses the interpolation methods, neural network architectures, and transformer-based models and image quality assessment methods utilized in this study; Section III presents the experiments conducted; and Section IV provides the results and suggestions.

2. Materials and Methods

This section provides a detailed examination of the techniques and models employed in this study, alongside the image quality metrics utilized for evaluation.

2.1. Resolution Enhancement Methods

Generalizing the effect of color space conversion, interpolation methods and deep learning methods that have successful results in resolution enhancement are examined and have been studied.

2.1.1. Interpolation methods

Figure 1 shows the most widely used and known interpolation methods, which are nearest neighbor, bilinear, bicubic, and Lanczos interpolation, applied to 3D planes.

2.1.2. Nearest neighbor interpolation

The pixel value desired to be obtained in nearest neighbor interpolation takes the value of the known nearest point. Although it is the easiest interpolation method to implement and understand, created images become low quality. This is because the rectangular function is equivalent to the sinc function in the Fourier transform and there is a rapid loss in the transition gain (Fadnavis, 2014). Image distortion, called "pixelating" in nearest neighbor interpolation, is a more noticeable problem, especially as the upscaling factor is increased. In our nearest neighbor interpolation experiments, Euclidean distance is used for distance metric.

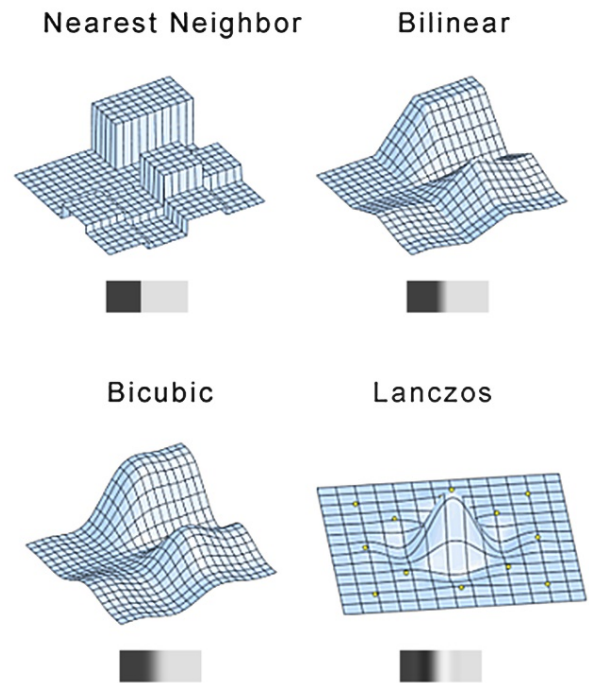


Figure 1. Demonstration of nearest neighbor, bilinear, bicubic, and Lanczos interpolation methods on examples (Getreuer, 2011).

2.1.3. Bilinear interpolation

According to the table shown in Figure 2, the data to be estimated is obtained from pairs of points. E value is obtained from A-B pairs and F value is obtained from C-D pairs. Then, P is obtained from the predicted E-F point pairs using weighted average. More realistic results are achieved than nearest neighbor interpolation (Han, 2013).

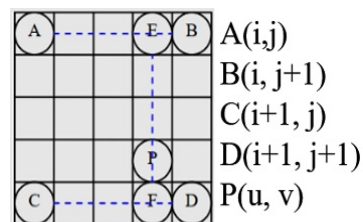


Figure 2. Application of the bilinear interpolation method on the matrix (Han, 2013).

2.1.4. Bicubic interpolation

It works similarly to the bilinear interpolation algorithm. The calculation is made here considering that it is a square matrix with 16 elements, where the points A-B-C-D are placed at the corners as shown in Figure 3. In this algorithm, the influence is expanded more than in the double linear interpolation algorithm. While calculating the point e between points A and B, a linear and non-broken curve is used by using the points A-1 and B+1 (Han, 2013).

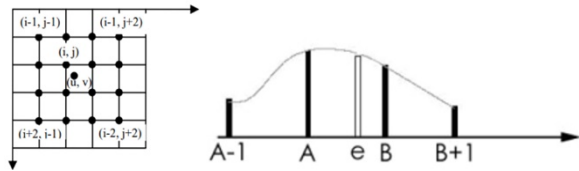


Figure 3. Diagram of bicubic interpolation algorithm (Han, 2013).

2.1.5. Lanczos interpolation

Lanczos interpolation is a low-band filtering method close to the ideal sinc function. It is performed using the sinc filter. The computational cost is more costly than other methods because it contains trigonometric functions. The sinc function uses a windowed pruned sinc function, as shown in Figure 4 (Burger and Burge, 2008). It is more successful than the other interpolation methods mentioned in removing noises in the image.

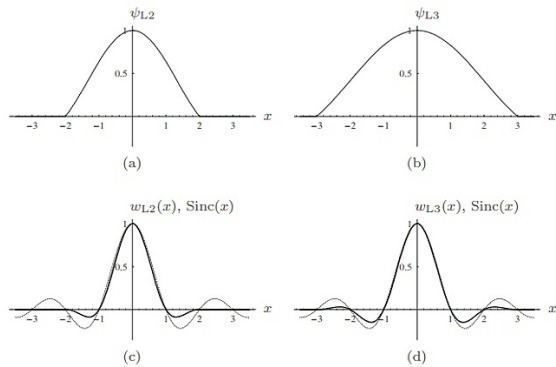


Figure 4. Lanczos window functions (a, b) and their corresponding interpolation kernels (c, d) (The original sinc function is shown as dashed.) (Burger and Burge, 2008).

2.1.6. Deep learning methods

Deep learning has introduced transformative advancements in computer vision, with architectures like Swin Transformer and GANs leading the way in feature extraction and high-resolution image synthesis. Swin Transformer excels at multi-scale feature learning through hierarchical attention mechanisms, while GANs leverage adversarial training for generating highly realistic images, exemplifying state-of-the-art deep learning applications in image processing (Ma et al., 2022; Wang et al., 2023).

2.1.7. SRGAN (Super-Resolution generative adversarial networks)

SRGAN is a generative adversarial network-based neural network that focuses on a single-image super-resolution called perceptual loss, which is the weighted sum of adversarial and content loss, with a new approach, as seen in equation 1 (Ledig et al., 2017).

$$I^{SR} = I_X^{SR} + 10^{-3} I_{Gen}^{SR} \tag{1}$$

SRGAN's discriminator network works as expected as standard GAN architecture. In a generative network, residual blocks are used to make the training process easier and to keep the previous layers connected and active. Another crucial point is the input data. Instead of using random noise, the input image is used directly. SRGAN generates realistic textures on images but they have some artifacts (Wang et al., 2018).

In our SRGAN experiments, all parameters and network types are defined as mentioned in the original study (Ledig et al., 2017).

2.1.8. Real-ESRGAN (Enhanced super-resolution generative adversarial networks)

ESRGAN (Wang et al., 2018) tries to enhance image quality according to SRGAN results. ESRGAN uses SRGAN network architecture, adversarial loss, and perceptual loss. Besides this, residual-in-residual dense block is introduced, and batch normalization is abandoned. The idea of predicting relative realness comes from relativistic average GAN (RaGAN) (Jolicœur-Martineau, 2018). RaGAN compares images to determine which is more realistic instead of resolving, "Is this real or fake?" (Wang et al., 2018).

Most image restoration and enhancement approaches are not good at real-world data because real-world data is more complicated than the ones generated by classical degradation models consisting of down sampling, blurring, noising, and JPEG compression. Real-ESRGAN, mainly based on ESRGAN, is focused on real-world data to overcome its complexity by using completely synthetic data while training the network. U-Net discriminator (Ronneberger et al., 2015) is used instead of a VGG-style discriminator to make local details clearer (X. Wang et al., 2021).

In our Real-ESRGAN experiments, all parameters and network types are defined as mentioned in the original study (X. Wang et al., 2021).

2.1.9. Swin2SR (SwinV2 transformer for compressed image super-resolution and restoration)

Transformer is a network architecture built for natural language processing and becomes an important advancement in this topic by using the attention mechanism (Vaswani et al., 2017). In image processing, CNNs are very useful, but Vision Transformers (ViT) achieve better results than CNNs with fewer computational resources at the training phase (Dosovitskiy et al., 2020). The idea behind ViT is the representation of the image in the transformer network.

While pixel arrays are used in CNNs, ViT uses visual tokens that represent an image split into non-overlapped patches.

CNNs use the same convolution kernel for various image sections, but this may not be a good idea for all situations because they are content-independent. The self-attention mechanism in visual transformers solves this problem, but this time, fixed-sized patches may cause border artifacts and sacrifice information over border pixels. With shallow feature extraction, deep feature extraction, and high-quality image reconstruction modules, SwinIR has fewer parameters but more successful results (Liang et al., 2021).

Swin2SR is an improved form of SwinIR (Liang et al., 2021) using SwinV2 Transformer (Liu et al., 2022) network specialized for super-resolution. With pre-normalization in the SwinV2 Transformer, more training parameters without instabilities become possible. Scaled cosine attention between queries and keys minimizes resolution gaps. With these improvements, despite using %33 fewer iterations in the training process, Swin2SR achieves similar results to SwinIR (Conde et al., 2023).

In our Swin2SR experiments, all parameters and network types are defined as mentioned in the original study (Conde et al., 2023).

2.2. Image Quality Assessments

PSNR and SSIM (Wang et al., 2004), visual comparison metrics that have been used for a long time and have become standardized in the literature, give results contrary to human perception in some techniques that show successful results. Therefore, artificial neural network-based approaches are emerging to perform better optimization in artificial neural network-based visual enhancement methods and to determine how close the proposed images are to the original forms. With LPIPS (Learned Perceptual Image Patch Similarity) (Zhang et al., 2018), DISTs (Deep Image Structure and Texture Similarity) (Ding et al., 2020) and A-DISTs (Adaptive-DISTs) (Ding et al., 2021), which are a metric developed with the idea of combining structural and textural similarity, results become very close to human interpretation.

Quality assessment metrics are divided into two categories: information-oriented and data-oriented. Knowledge-based methods such as PSNR, SSIM, FSIM (Zhang et al., 2011), VIF (Sheikh and Bovik, 2006), and GMSD (Xue et al., 2013) generally use classical distance measurements between two images. For this reason, the mentioned methods do not achieve as accurate results as DeepIQA (Bosse et al., 2017), PieAPP (Popat and Picard, 1997), LPIPS (Zhang et al., 2018), DISTs, and A-DISTs, which are data-driven methods for capturing perceptual similarities. To compare these methods, which produce values in completely different value ranges and different sizes, an approach called "Two-alternative forced choice" (2AFC) is used, in which the observer has to choose one of two options in a way that would prevent from being undecided (Ding et al., 2021).

In Figure 5, the original (a), degraded (b), and resampled original image (c) are shown. Although PSNR, SSIM, FSIM, VIF, GMSD, DeepIQA, PieAPP, and LPIPS metrics all give the result that image (b) is better than the image (c) contrary to human perception, DISTs metric is concluded that image (c) is better. However, in some cases, DISTs can reach a general conclusion over the entire image by ignoring the textures in the image. To overcome this ignorance problem, A-DISTs is developed to analyze local tissue details better with an adaptive approach.

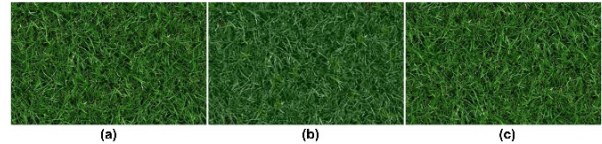


Figure 5. Original grass image (a), degradation after JPEG compression (b), resampled image (c)

The VGG neural network in DISTs does not offer sufficient sensitivity in terms of size. The convolution filters of the VGG neural network are renormalized so that the L2 norm of each filter is equal to 1. Thus, statistics are made easier to compare by ensuring that all convolution filters respond in similar intervals. A statistical feature called the dispersion index is used with A-DISTs as formulated in equation 2. In (3), $\tilde{p}_k^{(i)}$ is used to calculate the difference between X and Y images, and the pattern probability of the i^{th} dimension of the k^{th} patch is expressed. This measurement produces a value in the range of [0,1] (Ding et al., 2020).

As the PSNR and SSIM values increase, it is concluded that there are more compatible images, while results close to zero with DISTs and A-DISTs reveal that the images are more compatible. Resolution enhancement studies still include PSNR and SSIM metrics to compare with previous studies. DISTs and A-DISTs results are also included in our study, as they give the closest results to evaluating local textures and human perception according to their formulation as seen in equations 2 and 3.

$$A - DISTs(X, Y) = 1 - \frac{1}{N} \sum_{i=0}^M \sum_{j=0}^{N_i} S(\tilde{X}_j^{(i)}, \tilde{Y}_j^{(i)}) \quad (2)$$

$$S(\tilde{X}_j^{(i)}, \tilde{Y}_j^{(i)}) = \frac{1}{K_i} \sum_{k=1}^{K_i} (\tilde{p}_k^{(i)} l(\tilde{x}_{j,k}^{(i)}, \tilde{y}_{j,k}^{(i)}) + \tilde{q}_k^{(i)} s(\tilde{x}_{j,k}^{(i)}, \tilde{y}_{j,k}^{(i)})) \quad (3)$$

Another important perceptual metric is Learned Perceptual Image Patch Similarity (LPIPS). Zhang et al. introduces LPIPS designed to align closely with human visual similarity judgments. This metric leverages deep features from intermediate layers of pretrained convolutional neural networks to quantify perceptual similarity between images. These features capture both low-level visual attributes (e.g., edges, textures) and high-level semantic structures. For two images, feature maps $f_i(x_1)$ and $f_i(x_2)$ are extracted from layer l , normalized, and their L2 distances are computed in

equation 4, where w_l represents learned weights optimized to align with human perceptual judgments. The overall LPIPS distance is obtained by aggregating contributions across layers as seen in equation 5. The metric has been validated across various tasks, including super-resolution and image compression, demonstrating higher alignment with human evaluations compared to traditional metrics like PSNR and SSIM. By utilizing different architectures (e.g., AlexNet, VGG, SqueezeNet), LPIPS balances computational efficiency and perceptual fidelity, making it a robust tool for perceptual similarity assessment in computer vision (Zhang et al., 2018). In our experiments, predefined VGG network model is used to evaluate method outputs.

$$d_l(x_1, x_2) = \|w_l \odot (f_l(x_1) - f_l(x_2))\|_2^2 \quad (4)$$

$$LPIPS(x_1, x_2) = \sum_l d_l(x_1, x_2) \quad (5)$$

According to the PSNR calculation in equation 6, MAX_i value depends on the maximum available value in the signal, which is 255 for RGB images for each axes (Dosselmann and Yang, 2005). However, according to John et. al., in all PSNR experiments among various color spaces, the maximum value is assumed to be 255 (John et al., 2016). Similarly, Wang et al. (2021) accepts 255 as the maximum value because of the 8-bit representation. In our study, MAX_i is used as 255, too. In the SSIM original paper, it's unclear what color space is used for the input (Nilsson and Akenine Möller, 2020; Wang et al., 2004). For both SSIM and PSNR calculations, 255 are assumed as the maximum value for all color space axes for a fair comparison within this study.

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_i^2}{MSE} \right) \quad (6)$$

Although the PSNR calculation in resolution enhancement studies is mostly obtained by converting images in various color spaces to YCbCr color space and calculating on the Y channel, it is suggested that a clear approach is needed in this regard (Keles et al., 2021). However, in this study, in which the effect of color space change on the results is examined, the quality assessments of each color space are calculated and evaluated separately in terms of color space.

2.3. Methodology

In our experiments, all images are enhanced exclusively within the RGB color space with 4x upscaling factor. Following the super-resolution process, the upscaled and ground truth images are subsequently converted to the YCbCr and CIELab color spaces. Image Quality Assessment (IQA) metrics are then applied to both the original and enhanced images within their respective color spaces for comprehensive evaluation.

Figure 6 demonstrates the method overview.

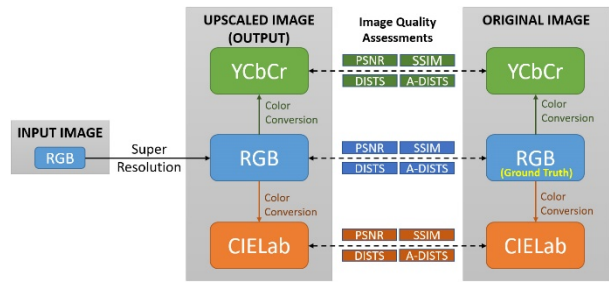


Figure 6. Method overview.

The PSNR and SSIM metrics are calculated for all axes to enable a clearer comparison of the differences between color spaces. However, many studies in the literature focus on the Y axis in the YCbCr color space, and the L^* axis in the CIELab color space that include luminance information, often disregarding the chromatic components. Due to differences in how these metrics are calculated and the perspectives presented in various studies, expressing the changes in metric results as percentages is crucial for understanding the extent of improvement. In our study, the most important consideration when analyzing results should be to focus on the percentage increase in metric scores rather than their absolute values. This is because our study performs enhancements on images exclusively in the RGB color space but evaluates performance metrics on the achromatic axes of the YCbCr and CIELab color spaces.

2.3.1. Interpolation method details

In our bilinear interpolation experiments, the weighted average is used to obtain the missing value; in bicubic interpolation experiments, 4x4 pixel neighborhoods around the target pixel is used based on a cubic polynomial expansion; in Lanczos interpolation experiments, windowed sinc function utilizing 4x4 neighborhood is used.

2.3.2. SRGAN hyper-parameters and model details

The SRGAN generator network consists of 16 residual blocks, each constructed with 3x3 convolutional layers featuring 64 feature maps per layer, and employs Parametric ReLU (PReLU) as the activation function for enhancing the learning capacity of neural networks by allowing the slope of negative inputs to be learned adaptively, improving model flexibility and performance compared to standard ReLU. To upscale image resolution, two sub-pixel convolutional layers are incorporated. The discriminator network, on the other hand, is composed of 8 convolutional layers with a kernel size of 3x3, where the number of feature maps doubles at each layer, starting from 64 and increasing up to 512. Leaky ReLU ($\alpha=0.2$), addresses the "dying ReLU" problem by allowing a small, non-zero gradient for negative inputs, enabling the network to learn even when activations are not strictly positive, is utilized as the activation function in the discriminator (Erdemir et al., 2020). Optimization is performed using the Adam algorithm ($\beta_1=0.9$) with a learning rate set to 10^{-4} for the first 100,000 iterations and reduced to 10^{-5} for the subsequent 100,000

iterations. The loss function combines perceptual loss, which leverages content loss based on high-level feature maps of a pre-trained VGG network, with adversarial loss driven by the discriminator's ability to distinguish super-resolved images from real high-resolution images. During training, each mini-batch contains 16 images with a resolution of 96×96, derived from a dataset of 350,000 images from ImageNet. To enhance stability and avoid suboptimal convergence, the generator is initialized using the SRResNet model pre-trained with MSE loss. This carefully structured design allows SRGAN to produce photo-realistic SR images, effectively bridging the gap between low-resolution inputs and high-resolution outputs (Ledig et al., 2017).

2.3.3. Real-ESRGAN hyper-parameters and model details

The Real-ESRGAN model is designed to address complex real-world image degradations using a refined architecture and carefully tuned hyperparameters. Its generator adopts Residual-in-Residual Dense Blocks (RRDB) architecture, while the discriminator utilizes a U-Net design with spectral normalization to enhance local detail discrimination and stabilize training. The model is trained in two stages: an initial PSNR-oriented phase using L_1 loss and a combined training phase incorporating L_1 , perceptual, and GAN losses with respective weights of {1, 1, 0.1}. Optimized with the Adam optimizer ($\beta_1=0.9$, $\beta_2=0.999$), the learning rate is set to 2×10^{-4} during the first stage and 10^{-4} in the second, with a total of 1,000,000 and 400,000 iterations respectively. The training employs high-resolution patches (256×256) from datasets such as DIV2K and Flickr2K and uses a second-order degradation model incorporating Gaussian and Poisson noise, JPEG compression, and sinc filters. Additional techniques such as Exponential Moving Average (EMA) and sharpening of ground-truth images further enhance training stability and visual sharpness, ensuring robust performance in real-world scenarios (X. Wang et al., 2021).

2.3.4. Swin2SR hyper-parameters and model details

The Swin2SR model leverages an advanced architecture optimized for compressed image super-resolution and restoration, incorporating 6 Residual Swin Transformer Blocks, each containing 6 Swin Transformer Layers with a window size of 8, 180 channels, and 6 attention heads. The training process employs an L_1 loss as the primary objective, supplemented by auxiliary and high-frequency loss functions to enhance the accuracy of low-resolution restoration and preserve fine details. Training is conducted on DIV2K and Flickr2K datasets using 192-pixel high-resolution image patches, with standard augmentations such as rotation and flipping. Optimization is performed with the Adam optimizer ($\beta_1=0.9$, $\beta_2=0.999$) and an initial learning rate of 2×10^{-4} . The model achieves efficient resolution enhancement through a pixel shuffle mechanism, maintaining a relatively compact size of 12 million parameters, enabling it to deliver competitive performance in restoring high-quality images from heavily compressed inputs (Conde et al., 2023).

3. Results

When the quality of the improved images is evaluated according to PSNR and SSIM measurements, as Ledig et al. mentioned, some inconsistencies stand out in various methods because these metrics are based on pixel-wise differences and cannot capture perceptual details well (Ledig et al., 2017). However, for transformer-based approaches, PSNR and SSIM metrics are still working as expected. It is observed qualitatively that better results are achieved with Swin2SR in almost all enhanced images, as seen in

Table 1.

Enhanced images on the Set5 dataset with nearest neighbor, linear, bicubic, Lanczos interpolations, SRGAN, Real-ESRGAN, and Swin2SR results are shown in Figure 7, Figure 8, and Figure 9.

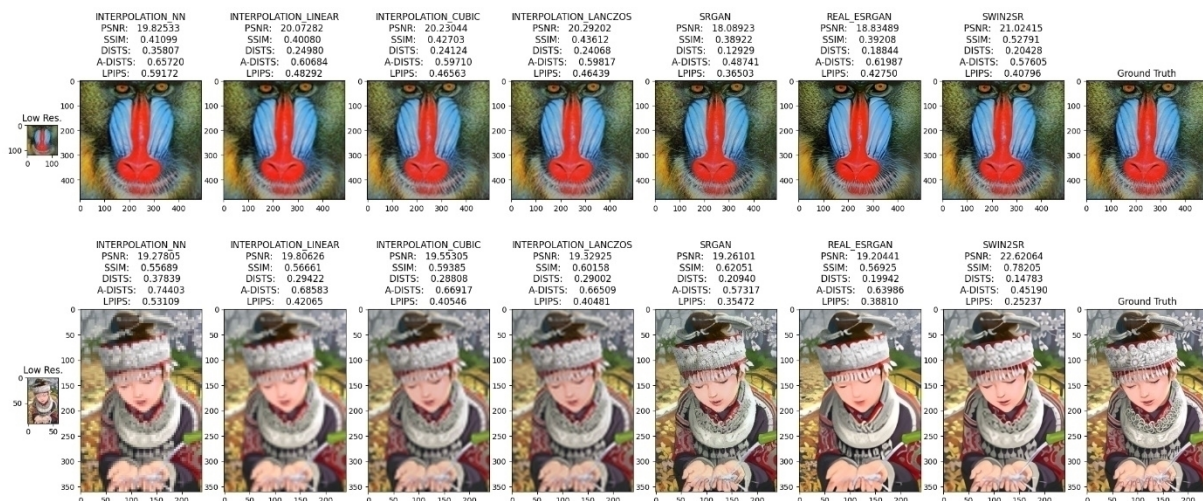


Figure 7. Interpolations, SRGAN, Real-ESRGAN, and Swin2SR results on Set14 - baboon.png and comic.png images



Figure 8. Interpolations, SRGAN, Real-ESRGAN, and Swin2SR results on Set5 dataset

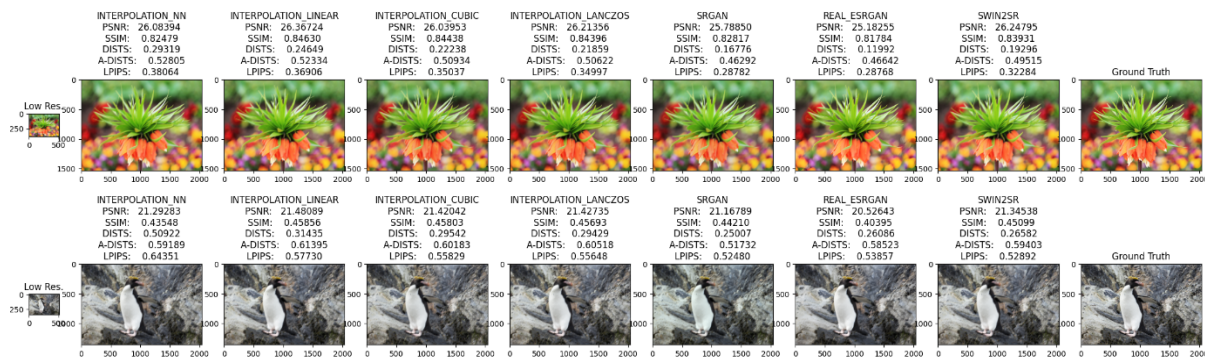


Figure 9. Interpolations, SRGAN, Real-ESRGAN, and Swin2SR results on DIV2K dataset.

The model used in Real-ESRGAN is trained with fully synthetic and noise-free data. Therefore, while it clearly shows the outlines of the objects in the images, in some cases, it is observed that "beauty filters" do smooth the image but lose the details. Besides this, there's a conflict in the baboon image between the DISTS and A-DISTS results. This is because the A-DISTS metric, which focuses much better on local patterns, is better optimized than

DISTS. In such images where texture details are important, A-DISTS is a more proper metric that can be used to determine the successful method.

When the conversion to YCbCr color space is performed on SRGAN and Real-ESRGAN outputs saved by working with original images in RGB color space, it is seen that only the effect of color space change on PSNR and SSIM results is positive.

Table 1 shows the averages of PSNR, SSIM, DISTS, A-DISTS, and LPIPS values calculated before and after conversion to YCbCr and CIELab color spaces according

to the methods applied to the data sets. Note that lower DISTS, A-DIST, and LPIPS values mean better results, unlike PSNR and SSIM.

Table 1. Image quality assessment scores of before and after conversion to YCbCr and CIELab color spaces (Best scores within the dataset and quality metric are marked as “*”.)

| Dataset | Method | Before Conversion (RGB ¹) | | | | | Success of After Conversion (%) | | | | | | | | | |
|-------------------------|---------------------------------|---------------------------------------|-------------------|--------------------|----------------------|--------------------|---------------------------------|-------|-------|---------|-------|---------------------|-------|-------|---------|-------|
| | | PSNR ⁴ | SSIM ⁵ | DISTS ⁶ | A-DISTS ⁷ | LPIPS ⁸ | YCbCr ² | | | | | CIELab ³ | | | | |
| | | | | | | | PSNR | SSIM | DISTS | A-DISTS | LPIPS | PSNR | SSIM | DISTS | A-DISTS | LPIPS |
| Set5 | Interpolation - NN ⁹ | 29.03 | 0.86 | 0.26 | 0.48 | 0.34 | 18.7% | 18.2% | 28.3% | 29.1% | 26.7% | 17.8% | 18.1% | 28.1% | 27.8% | 28.3% |
| | Interpolation - Linear | 30.44 | 0.89 | 0.21 | 0.42 | 0.28 | 17.9% | 16.3% | 5.2% | 20.6% | 8.4% | 17.0% | 16.1% | 8.0% | 21.2% | 13.0% |
| | Interpolation - Cubic | 30.04 | 0.89 | 0.20 | 0.39 | 0.26 | 18.6% | 19.1% | 6.4% | 23.6% | 10.5% | 18.5% | 18.7% | 8.9% | 24.1% | 15.0% |
| | Interpolation - Lanczos | 30.34 | 0.90 | 0.19 | 0.38 | 0.25 | 18.4% | 18.8% | 8.7% | 25.2% | 12.8% | 18.3% | 18.4% | 11.0% | 25.7% | 16.9% |
| | SRGAN ¹⁰ | 31.44 | 0.90 | 0.09 | 0.23 | *0.14 | 17.3% | 12.0% | 36.8% | 43.0% | 36.8% | 16.9% | 11.3% | 39.6% | 44.7% | 40.9% |
| | Real-ESRGAN ¹¹ | 28.81 | 0.85 | 0.14 | 0.37 | 0.21 | 18.5% | 16.5% | 13.5% | 26.7% | 21.0% | 17.8% | 16.3% | 19.5% | 29.3% | 25.1% |
| | Swin2SR ¹² | *35.57 | *0.94 | *0.11 | *0.24 | *0.14 | 15.2% | 6.4% | 31.1% | 38.6% | 32.0% | 14.5% | 6.1% | 34.6% | 41.0% | 36.2% |
| Set14 | Interpolation - NN | 27.07 | 0.82 | 0.26 | 0.50 | 0.37 | 20.1% | 26.9% | 25.2% | 24.4% | 25.6% | 19.5% | 26.9% | 25.6% | 24.0% | 25.5% |
| | Interpolation - Linear | 27.75 | 0.83 | 0.24 | 0.49 | 0.35 | 19.6% | 25.8% | -5.2% | 15.7% | 10.2% | 19.1% | 25.8% | -4.5% | 15.3% | 10.7% |
| | Interpolation - Cubic | 27.31 | 0.84 | 0.24 | 0.47 | 0.33 | 20.1% | 25.7% | -4.3% | 17.8% | 11.3% | 20.1% | 25.8% | -3.5% | 17.4% | 11.5% |
| | Interpolation - Lanczos | 27.11 | 0.84 | 0.23 | 0.46 | 0.33 | 20.3% | 25.4% | -3.3% | 18.8% | 12.0% | 20.3% | 25.4% | -2.9% | 18.2% | 12.1% |
| | SRGAN | 28.36 | 0.83 | *0.12 | 0.34 | *0.22 | 19.4% | 23.5% | 20.5% | 28.3% | 26.8% | 19.1% | 22.8% | 21.2% | 28.3% | 27.6% |
| | Real-ESRGAN | 27.80 | 0.81 | 0.15 | 0.44 | 0.27 | 19.5% | 25.3% | 3.8% | 18.5% | 17.0% | 19.1% | 25.3% | 3.8% | 19.1% | 17.6% |
| | Swin2SR | *31.50 | *0.89 | 0.14 | 0.34 | *0.22 | 16.8% | 14.0% | 11.5% | 26.3% | 22.8% | 16.5% | 14.0% | 9.9% | 26.2% | 22.5% |
| BSDS100 | Interpolation - NN | 25.85 | 0.83 | 0.28 | 0.51 | 0.41 | 22.0% | 43.5% | 25.2% | 21.6% | 24.7% | 21.1% | 43.3% | 24.8% | 20.2% | 24.2% |
| | Interpolation - Linear | 26.75 | 0.86 | 0.27 | 0.49 | 0.37 | 21.3% | 38.7% | -4.6% | 16.9% | 11.6% | 20.5% | 38.3% | -3.5% | 16.2% | 11.1% |
| | Interpolation - Cubic | 26.43 | 0.87 | 0.26 | 0.47 | 0.36 | 21.5% | 36.8% | -4.9% | 18.0% | 9.8% | 20.8% | 36.4% | -4.1% | 17.2% | 9.0% |
| | Interpolation - Lanczos | 26.50 | 0.87 | 0.26 | 0.47 | 0.35 | 21.5% | 36.1% | -4.0% | 18.8% | 10.7% | 20.8% | 35.7% | -3.5% | 18.0% | 10.1% |
| | SRGAN | 28.09 | 0.84 | *0.12 | *0.32 | *0.22 | 20.2% | 34.4% | 27.4% | 31.0% | 27.6% | 19.3% | 32.6% | 25.4% | 30.2% | 26.9% |
| | Real-ESRGAN | 28.25 | 0.85 | 0.16 | 0.41 | 0.27 | 19.8% | 34.7% | 13.6% | 24.0% | 22.6% | 19.1% | 34.1% | 13.4% | 23.4% | 22.1% |
| | Swin2SR | *31.33 | *0.91 | 0.17 | 0.36 | 0.25 | 17.8% | 21.7% | 14.8% | 27.5% | 25.2% | 17.0% | 21.3% | 15.4% | 26.9% | 25.2% |
| Urban100 | Interpolation - NN | 25.33 | 0.81 | 0.26 | 0.47 | 0.36 | 22.5% | 31.6% | 21.4% | 21.2% | 22.1% | 21.6% | 32.1% | 20.6% | 20.5% | 21.9% |
| | Interpolation - Linear | 25.84 | 0.82 | 0.26 | 0.46 | 0.35 | 22.0% | 31.4% | 1.1% | 17.8% | 12.9% | 21.1% | 31.7% | 1.1% | 17.5% | 13.6% |
| | Interpolation - Cubic | 25.13 | 0.82 | 0.25 | 0.44 | 0.33 | 22.7% | 35.2% | 1.7% | 18.9% | 14.6% | 22.1% | 35.4% | 1.6% | 18.5% | 14.9% |
| | Interpolation - Lanczos | 25.12 | 0.82 | 0.25 | 0.44 | 0.33 | 22.7% | 34.7% | 2.4% | 19.4% | 15.0% | 22.1% | 34.8% | 2.2% | 19.0% | 15.1% |
| | SRGAN | 26.93 | 0.84 | *0.11 | 0.33 | 0.21 | 21.1% | 19.7% | 29.2% | 29.8% | 30.9% | 20.2% | 18.5% | 28.3% | 30.1% | 31.1% |
| | Real-ESRGAN | 25.69 | 0.82 | 0.13 | 0.38 | 0.25 | 22.0% | 24.5% | 12.7% | 23.4% | 21.2% | 21.1% | 24.9% | 12.6% | 24.0% | 22.1% |
| | Swin2SR | *30.55 | *0.91 | *0.11 | *0.29 | *0.18 | 18.0% | 11.1% | 26.0% | 30.9% | 29.2% | 17.2% | 11.0% | 26.1% | 31.6% | 30.6% |
| DIV2K (unknown deg.) | Interpolation - NN | 22.93 | 0.57 | 0.41 | 0.59 | 0.54 | 18.1% | 40.3% | 28.9% | 17.0% | 20.6% | 18.8% | 40.3% | 26.3% | 16.5% | 17.7% |
| | Interpolation - Linear | 23.20 | 0.58 | 0.30 | 0.61 | 0.51 | 17.8% | 37.5% | -2.5% | 17.7% | 14.0% | 18.5% | 37.1% | -3.5% | 14.8% | 10.8% |
| | Interpolation - Cubic | 22.80 | 0.57 | 0.28 | 0.59 | 0.49 | 18.4% | 41.9% | -1.6% | 18.3% | 13.5% | 19.1% | 41.5% | -2.0% | 16.2% | 9.9% |
| | Interpolation - Lanczos | 22.95 | 0.58 | 0.28 | 0.59 | 0.49 | 19.0% | 41.0% | 0.3% | 19.4% | 14.2% | 18.8% | 43.7% | -0.9% | 16.5% | 10.9% |
| | SRGAN | 23.00 | 0.57 | 0.23 | *0.53 | *0.44 | 18.3% | 41.7% | 4.4% | 18.0% | 16.2% | 18.8% | 41.3% | 2.5% | 27.7% | 13.1% |
| | Real-ESRGAN | 22.33 | 0.54 | *0.21 | 0.58 | *0.44 | 16.9% | 47.0% | 10.8% | 10.9% | 14.4% | 19.5% | 46.6% | 4.9% | 18.8% | 10.3% |
| | Swin2SR | *23.29 | *0.59 | 0.25 | 0.58 | 0.46 | 36.7% | 38.7% | 2.9% | 22.2% | 14.9% | 18.5% | 37.5% | 1.1% | 19.4% | 10.8% |
| Average (%) | | | | | | | 20.0% | 28.5% | 10.9% | 22.8% | 18.9% | 19.2% | 28.3% | 11.1% | 23.0% | 18.9% |

¹ RGB: Red, Green, and Blue, ² YCbCr: Luminance, Chrominance Blue, and Chrominance Red, ³ CIELab: Commission Internationale de l’Eclairage’s standard; L* is for luminance, a* is for green to red, b* is for blue to yellow, ⁴ Peak Signal-to-Noise Ratio, ⁵ Structural Similarity Index Measure, ⁶ Deep Image Structure and Texture Similarity, ⁷ Adaptive-Deep Image Structure and Texture Similarity, ⁸ Learned Perceptual Image Patch Similarity, ⁹ Nearest Neighbor, ¹⁰ Super Resolution Generative Adversarial Networks, ¹¹ Enhanced Super-Resolution Generative Adversarial Networks, ¹² SwinV2 Transformer for Compressed Image Super-Resolution and Restoration.

4. Discussion

Our study delves into the dual impact of transforming RGB data into alternative color spaces, emphasizing not only the quantitative changes in evaluation metrics but also their correspondence to human perceptual experiences. The findings suggest that while certain transformations improve specific metrics, their perceptual relevance varies, underscoring the importance of incorporating human-centric assessments in future research. These insights highlight a critical gap in the literature, advocating for the development of

methods that balance computational accuracy with perceptual validity.

Wang et al. (2021) and John et al. (2016) have highlighted the effectiveness of the YCbCr color space for improving PSNR and SSIM metrics in super-resolution tasks. Our findings corroborate these observations, demonstrating that transformations to the YCbCr space lead to higher scores in these metrics. However, similar to prior studies, we found that these improvements do not correspond to significant enhancements in perceptual image quality. Furthermore, the results obtained through transformations to the CIELab color

space align with the findings of John et al. (2016), who suggested its potential utility in resolution enhancement. Nevertheless, our experiments reveal that the contributions of these transformations are predominantly metric-specific and do not reflect tangible improvements in visual fidelity (Ding et al., 2021; John et al., 2016).

According to Dong et al. (2016), the YCbCr space allows the method to enhance only the luminance channel, which is crucial for SR tasks aiming at perceptual improvements. This focus reduces the computational

load and avoids distortions in color channels (Dong et al., 2016).

Xu et al. (2020) propose a super-resolution reconstruction method that integrates $L_{2/3}$ sparse regularization with color channel constraints. By converting low-resolution images to the YCbCr color space, the luminance channel is enhanced using the sparse model, while the chrominance channels are refined to reduce color artifacts (Xu et al., 2020). Note that this study applies x2 upscaling factor to only five images that listed in Table 2, PSNR and SSIM metrics are applied on Y channel solely.

Table 2. Bicubic interpolation and proposed method of Xu et al. (2020), x2 upscaling factor with YCbCr color space

| Image | PSNR ¹ | | | SSIM ² | | |
|-----------|-----------------------|--------------------|---------|-----------------------|--------------------|---------|
| | Bicubic Interpolation | Xu et al.'s Method | Success | Bicubic Interpolation | Xu et al.'s Method | Success |
| Flowers | 30.38 | 33.40 | 9.9% | 0.898 | 0.938 | 4.5% |
| Comic | 26.06 | 27.92 | 7.1% | 0.851 | 0.908 | 6.7% |
| Butterfly | 27.42 | 31.19 | 13.7% | 0.915 | 0.958 | 4.7% |
| Skiing | 32.00 | 34.33 | 7.3% | 0.931 | 0.953 | 2.4% |
| Bike | 25.66 | 27.92 | 8.8% | 0.850 | 0.913 | 7.4% |

¹ Peak Signal-to-Noise Ratio, ² Structural Similarity Index Measure

In all the experiments of our study, the resolution improvement process is carried out in RGB color space, and measurements are calculated before and after the conversion to the color spaces specified in the tables. Therefore, the effects of different color spaces on the calculations of success values are examined without any effect on the improvement phase. Although converting an image in RGB color space to YCbCr and CIELab color spaces does not affect the image quality, it positively affects the quality assessment results. The metrics used to assess the success of SISR methods are not sufficient

for evaluating the effects on different color spaces. There is a need for alternative metrics that can effectively assess the success across various color spaces. Without developing these metrics, examining the impact of color space transformation will not be robust.

The success seems to be high in interpolating and deep learning architectures because the value ranges between the color spaces do not match exactly. The extreme values and the number of possible different values are shown in Table 3.

Table 3. Value ranges of RGB, YCbCr, and CIELab color space axes (Yang et al., 2007)

| | RGB ¹ | | | YCbCr ² | | | CIELab ³ | | |
|---------------------------------|------------------|-----|-----|--------------------|-----|-----|---------------------|------|------|
| | R | G | B | Y | Cb | Cr | L* | a* | b* |
| Maximum Value | 255 | 255 | 255 | 235 | 240 | 240 | 100 | 110 | 110 |
| Minimum Value | 0 | 0 | 0 | 16 | 16 | 16 | 0 | -110 | -110 |
| Number of values | 256 | 256 | 256 | 220 | 225 | 225 | 101 | 221 | 221 |
| Possible values per color space | 16,777,216 | | | 11,137,500 | | | 4,932,941 | | |

¹ RGB: Red, Green, and Blue, ² YCbCr: Luminance, Chrominance Blue, and Chrominance Red, ³ CIELab: Commission Internationale de l'Eclairage's standard; L* is for luminance, a* is for green to red, b* is for blue to yellow.

Converting an image from an RGB color space that scales to a wider range of values, such as a YCbCr color space that scales to a narrower range of values, will reduce the variety of color values from 16,581,375 different possible values to 11,137,500. As a result of this decrease, although there is no improvement other than the technique applied in the image quality, the quality seems to increase according to the calculation metrics. The same situation is observed in the CIELab color space, too. This time, while converting from RGB, all possible values in CIELab decrease to 4,932,941.

The impact of color space transformation on performance metrics could similarly be a subject of investigation in other areas of image processing.

Despite these considerations, if a comparison between YCbCr and CIELab color spaces is to be made beyond RGB, YCbCr emerges as computationally efficient due to its linear transformation from RGB and the independent processing of the luminance channel. This makes it ideal for tasks where structural accuracy is prioritized over color fidelity. On the other hand, CIELab's alignment with human perception renders it suitable for applications

requiring high color accuracy, albeit at the cost of increased computational complexity. Consequently, YCbCr is often preferred in real-time applications, whereas CIELab is more suitable for offline processes where perceptual quality is of greater importance.

5. Conclusion

This study explores the intricate relationship between color space transformations and super-resolution techniques, highlighting the balance between quantitative metrics and perceptual quality. While RGB continues to dominate as the standard color space for resolution enhancement, our findings challenge its singular dominance by examining the impact of post-enhancement transformations to any color space with achromatic axis like YCbCr and CIELab. Although these transformations do not inherently enhance visual quality, they significantly influence image quality metrics both pixel-wise and perceptual-based. This discrepancy raises important questions about whether working in different color spaces meaningfully impacts perceived image quality or simply skews numerical evaluations. Future efforts should aim to refine quality assessment frameworks and explore innovative hybrid approaches that maximize both perceptual and computational benefits in super-resolution tasks.

Author Contributions

The percentages of the authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

| | H.H. | Z.G.A. | Ö.D. |
|-----|------|--------|------|
| C | 70 | 15 | 15 |
| D | 80 | 10 | 10 |
| S | 30 | 35 | 35 |
| DCP | 80 | 10 | 10 |
| DAI | 70 | 15 | 15 |
| L | 80 | 10 | 10 |
| W | 80 | 10 | 10 |
| CR | 80 | 10 | 10 |
| SR | 70 | 15 | 15 |

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision.

Conflict of Interest

The authors declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because there was no study on animals or humans.

Acknowledgements

This study is a part of the PhD thesis of Hürkal HÜSEM at the Institute of Graduate Studies, Istanbul University-Cerrahpaşa, Istanbul, Türkiye. The source code for this project is available on GitHub at <https://github.com/hurkal/sisr-color-space>

References

- Agustsson E, Timofte R. 2017. Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, Honolulu, HI, USA, July 21-26, pp: 126-135.
- Bevilacqua M, Roumy A, Guillemot C, Alberi-Morel ML. 2012. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: Proceedings of the 23rd British Machine Vision Conference (BMVC), Surrey, UK, September 3-7, pp: 1-10.
- Bosse S, Maniry D, Müller KR, Wiegand T, Samek W. 2017. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Trans Image Process*, 27(1): 206-219.
- Burger W, Burge MJ. 2008. Introduction to spectral techniques. In: *Principles of Digital Image Processing: Core Algorithms*. Springer, London, UK, 1st ed, pp: 313-342.
- Candès EJ, Fernandez-Granda C. 2014. Towards a mathematical theory of super-resolution. *Commun Pure Appl Math*, 67(6): 906-956.
- Conde MV, Choi UJ, Burchi M, Timofte R. 2023. Swin2SR: SwinV2 transformer for compressed image super-resolution and restoration. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Tel Aviv, Israel, October 23-27, pp: 669-687.
- Ding K, Liu Y, Zou X, Wang S, Ma K. 2021. Locally Adaptive Structure and Texture Similarity for Image Quality Assessment. In: Proceedings of the 29th ACM International Conference on Multimedia, Virtual Event, China, October 20-24, pp: 2483-2491.
- Ding K, Ma K, Wang S, Simoncelli EP. 2020. Image quality assessment: Unifying structure and texture similarity. *IEEE Trans Pattern Anal Mach Intell*, 4(5): 2567-2581.
- Dong C, Loy CC, He K, Tang X. 2016. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans Pattern Anal Mach Intell*, 38(2): 295-307.
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv Prepr ArXiv201011929*.
- Dosselmann R, Yang XD. 2005. Existing and emerging image quality metrics. In: Canadian Conference on Electrical and Computer Engineering, Saskatoon, SK, Canada, May 1-4, pp: 1906-1913.
- Erdemir E, Dragotti P.L, Gündüz D. 2020. Privacy-aware time-series data sharing with deep reinforcement learning. *IEEE Trans Inf Forensics Secur*, 16: 389-401.
- Fadnavis S. 2014. Image interpolation techniques in digital image processing: an overview. *Int J Eng Res Appl*, 4(10): 70-73.
- Getreuer P. 2011. Linear methods for image interpolation. *Image Process Line*, 1: 238-259.
- Gong R, Wang Y, Cai Y, Shao X. 2017. How to deal with color in super resolution reconstruction of images. *Opt Express*, 25(10): 11144-11156.

- Han D. 2013. Comparison of commonly used image interpolation methods. In: Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering, Hangzhou, China, March 22-23, pp: 1556-1559.
- Huang JB, Singh A, Ahuja N. 2015. Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, pp: 5197-5206.
- John N, Viswanath A, Sowmya V, Soman KP. 2016. Analysis of various color space models on effective single image super resolution. In: Berretti S, Thampi S.M, Srivastava P.R, editors. Intelligent Systems Technologies and Applications. Springer, Cham, Switzerland, 1st ed, pp: 529-540.
- Jolicoeur-Martineau A. 2018. The relativistic discriminator: a key element missing from standard GAN. ArXiv Prepr ArXiv180700734.
- Keles O, Yilmaz, MA, Tekalp AM, Korkmaz C, Dogan Z. 2021. On the Computation of PSNR for a Set of Images or Video. In: Picture Coding Symposium, Bristol, UK, June 29 - July 2, pp: 1-5.
- Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp: 4681-4690.
- Liang J, Cao J, Sun G, Zhang K, Van Gool L, Timofte R. 2021. SwinIR: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, October 11-17, pp: 1833-1844.
- Liu Z, Hu H, Lin Y, Yao Z, Xie Z, Wei Y. 2022. Swin transformer v2: Scaling up capacity and resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, New Orleans, LA, USA, June 18-24, pp: 12009-12019.
- Ma J, Tang L, Fan F, Huang J, Mei X, Ma Y. 2022. SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer. IEEECAA J Autom Sin, 9(7): 1200-1217.
- Martin D, Fowlkes C, Tal D, Malik J. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, pp: 416-423.
- Nilsson J, Akenine Möller T. 2020. Understanding ssim. ArXiv Prepr ArXiv200613846.
- Popat K, Picard RW. 1997. Cluster-based probability model and its application to image and texture processing. IEEE Trans Image Process, 6(2): 268-284.
- Ronneberger O, Fischer P, Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, October 5-9, pp: 234-241.
- Sheikh H.R, Bovik A.C. 2006. Image information and visual quality. IEEE Trans Image Process, 15(2): 430-444.
- Su H, Li Y, Xu Y, Fu X, Liu S. 2024. A review of deep-learning-based super-resolution: From methods to applications. Pattern Recognit, 157: 110935.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN. 2017. Attention is all you need. Adv Neural Inf Process Syst, 30.
- Wang X, Xie L, Dong C, Shan Y. 2021. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, October 11-17, pp: 1905-1914.
- Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C. 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European conference on computer vision (ECCV) workshops, Munich, Germany, September 8-14, pp: 0-0.
- Wang Y, Liu W, Sun W, Meng X, Yang G, Ren K. 2023. A progressive feature enhancement deep network for large-scale remote sensing image super-resolution. IEEE Trans Geosci Remote Sens, 61(1): 1-13.
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process, 13(4): 600-612.
- Wang Z, Chen J, Hoi SCH. 2021. Deep Learning for Image Super-Resolution: A Survey. IEEE Trans Pattern Anal Mach Intell, 43(10): 3365-3387.
- Xu Z, Ma Q, Yuan F. 2020. Single color image super-resolution using sparse representation and color constraint. J Syst Eng Electron, 31(2): 266-271.
- Xue W, Zhang L, Mou X, Bovik AC. 2013. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE Trans Image Process, 23(2): 684-695.
- Yang Y, Yuhua P, Zhaoguang L. 2007. A fast algorithm for YCbCr to RGB conversion. IEEE Trans Consum Electron, 53(4): 1490-1493.
- Yilmaz İ, Güllü M, Baybura T, Erdoğan AO. 2002. Renk Uzayları ve Renk Dönüşüm Programı (RDP). Afyon Kocatepe Üniversitesi Fen Ve Mühendis Bilim Derg, 2(2): 19-35.
- Zeyde R, Elad M, Protter M. 2012. On single image scale-up using sparse-representations. In: Curves and Surfaces: 7th International Conference, Avignon, France, pp: 711-730.
- Zhang Lin, Zhang Lei, Mou X, Zhang D. 2011. FSIM: A feature similarity index for image quality assessment. IEEE Trans Image Process, 20(8): 2378-2386.
- Zhang R, Isola P, Efros AA, 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, Salt Lake City, UT, USA, June 18-23, pp: 586-595.