International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-07, Issue-02, MAY 2021

Comparison of Blurring Techniques for Generative Adversarial Network-based Super-Resolution models: an Empirical Study

Divya Mishra, M. Tech, Jamia Milia Islamia, New Delhi, India, dvmshr946@gmail.com

Ankit Mishra, Senior Analyst, EXL Services, Delhi, India, ankitmishra723@gmail.com

Abstract Super-resolution of images is used in many applications such as medical imaging, video surveillance, satellite imagery and astronomy, fraud detection, and more. For image super-resolution applications via Generative Adversarial Networks (GAN) based methods, low-resolution images of corresponding high-resolution datasets are required to generate super-resolution images. Multiple image low-resolution techniques exist in the state-of-the-art literature. It is, however, difficult to identify the best technique to apply, due to differences in the way they are evaluated in current literature.

This paper presents an empirical comparative study of four popular blurring techniques such as average, gaussian, median, and bilateral using 10 random colored images. Then the generated output images with our experiments are examined by popular image-quality metrics named Peak-Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Feature Similarity Index Measure (FSIM). Next, we compare the performance of blurring techniques using the image quality metrics and conclude that the bilateral blurring method performs better on PSNR, SSIM, and FSIM. The least performing method is the average blurring among all. This work is unique and to the best of our knowledge, no such experimental analysis for GAN-based super-resolution models was performed before

Keywords —Image Quality Metrics, Blurring Techniques, Generative Adversarial Networks, Image Super-resolution

I. INTRODUCTION

A. Overview

The central aim of Super-Resolution (SR) is to get a High-Resolution (HR) image from a Low-Resolution (LR) image. An HR image offers a high component of image pixel density and thereby a lot of intelligence concerned with image features provides great help in image analysis. The scope of relevant applications is very wide including military security applications, medical diagnosis lab centers for medical research, image and video resizing, graphics and animation industry, biometric applications, satellite, remote sensing applications to find the area of interest from aerial images, live monitoring, and underwater marine systems to help marine biological researchers. Very old and conventional methods require heavy and expensive hardware devices that are facing lots of problems like device portability, the effect of environmental conditions, parameters affecting device performance, insufficient bandwidth, lack of storage space, and limited computational power resources.

Detailed overview of traditional methods and recent deep learning-based methods are as follows:

B. Traditional Methods

Initially, the traditional methods of recovering HR image from an LR image are done based on nearest neighbors [1], bilinear interpolation [2], bicubic interpolation [3], etc. Nearest neighbors are one of the easy and native algorithms in past literature. Irrespective of the location of any other pixel in an image, the algorithm directly correlates the values of the nearest pixels to the target pixel whose value has to be interpolated. This technique is simple and fast because of less complexity and generalizes well on other training samples which leave out with less requirement of training examples. But the downside of the method is, it produces blurry low image quality effects that don't meet perceptual quality according to the human visual system. Moreover, they are unable to trace high-frequency details like edges and contours of objects within an image.

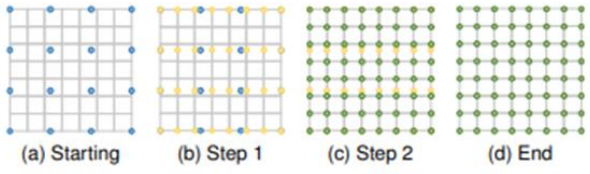

Fig. 1: Interpolation based up-sampling methods

Bilinear interpolation [5] is a sequential linear interpolation performed axis by axis as shown in Fig. 1. The figure shows step by step interpolation of single, quadratic, and then cubic for extracting respective 1 x 1, 2 x 2, and 4 x 4-pixel receptive fields for feature extraction. The method worked as quadratic

 DOI : 10.35291/2454-9150.2021.0278 © 2021, IJREAM All Rights Reserved.

interpolation with a 2 x 2 receptive field, which is not only fast but also results in better performance and improved image quality than nearest neighbors. Similarly, bi-cubic interpolation calculates three times interpolation on the two axes as shown in Fig. 1. This takes 4 x 4 pixels receptive field into account as compared to bilinear interpolation and results in more natural artifacts with compromising on speed. It is noticed that these traditional methods often include errors due to noise amplification, blurring or over-smoothing, computational complexity, produces some unwanted artifacts which are not according to the human visual system, low perceptual quality, and many more other side effects.

C. Deep Learning based Methods

To overcome the shortcomings of traditional methods for image SR, learning-based methods which are based on transposed convolution layer, meta-upscale module, and sub-pixel layer are introduced in the SR research area. Transposed convolution is also known as de-convolution that works on up-scaling of the image dimension and extract features from produced feature map to get the original shape of an image. In other words, it increases the image size by inserting zeros and performing convolution operations.

For gaining the understanding of high-level features without involving additional unnecessary parameters, a recursive learning approach is used which means using the same particular module iteratively multiple times. The method is used by Super-Resolution Convolutional Neural Network (SRCNN) [6] and Deep Residual-CNN (DRCNN) [7] with 16 recursions as an increase in the number of recursions improve the model without adding extra parameters. Later, Deep Recursive Residual Network (DRRN) [8] comes with 25 recursions and performs even better. Now, an increase in multiple recursive blocks improves the image quality but increases the computational complexity, cost, and training time. To overcome this without compromising image quality, Generative Adversarial Networks (GAN) based SR models are introduced.

There exist multiple blurring techniques in the state-of-the-art literature for image LR to feed into the generator network of GAN architecture. It is, however, difficult to identify the best technique to apply, due to differences in the way they are evaluated in current literature. Therefore, in this paper, we present an empirical comparative study of four popular blurring techniques using 10 random colored images and compared them for different blurring techniques using quantitative assessment of image quality metrics. To the best of our knowledge, such empirical comparative analysis has never been performed before to identify the better existing technique to apply. The work is important for the community working on deep learning-based image enhancement techniques particularly on GAN-based SR models. We particularly use blurring techniques to LR image to fed as an input to generator network. Furthermore, our work compares the different blurring techniques available through potential quantitative image quality metrics such as Peak-Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Feature Similarity Index Measure (FSIM), respectively. In this way, one can easily and accurately identify the better existing blurring technique. Hence, the contributions of this paper are as follows:

- We surveyed and gathered four popular blurring techniques namely Average blur, Gaussian blur, Median blur, and Bilateral blur for image LR to feed into generator network of GAN architecture from state-of-the-art literature.
- We experimented on 10 random samples of the coloured image dataset. The experiment is done with the aforementioned four different blurring techniques to figure out which method performs well while retaining maximum information too.
- We evaluated the quality of our experiments using popular image quality metrics from state-of-the-art literature such as PSNR, SSIM, and FSIM.
- We compared the results of each technique and identified bilateral blur as the better technique and average blur as the least performing technique.
- Conclusions drawn from this empirical study are helpful for computer vision researchers especially working in the area of image SR via GAN-based architectures.

The remainder of this paper is organized as follows. In Section 2, we briefly describe the basic concept of generative modeling and GANs. Section 3 explains image LR techniques like the most commonly used image down-scaling and other blurring techniques. Section 4 demonstrates a short introduction to different existing blurring techniques. Section 5 explicate a brief introduction on potential quantitative image quality metrics for evaluation purposes. Section 6 shows experimental results and compares different blurring techniques for each of the image quality metric PSNR, SSIM, and FSIM. Section 7 discuss the results and findings from the experiment along with its potential applications into the SR-based GAN models.

II. GENERATIVE MODELLING & GANS

GAN has a better capability of learning and therefore widely used in vision tasks [9]. To understand GAN models, we first need to briefly understand generative modeling.

A. Generative Modelling

Generative modeling is broadly known as a probabilistic model that describes how new samples are generated from existing ones. The model aims to generate new samples and follow the same rule of data distribution and the same set of feature combinations.

The newly generated samples should follow two rules:

- Data distribution of the new sample should be the same as the original sample.

 DOI : 10.35291/2454-9150.2021.0278 © 2021, IJREAM All Rights Reserved.

- New generated samples should be unique and different from the existing samples.

A generative model must be probabilistic and essentially include randomness that helps in the generation of individual samples of a model. It should not be deterministic, in the sense that the model completely depends on particular calculations like taking maximum or average of all pixel value of an image. By having this fixed deterministic approach, the model is no more generative and produces the ame output image every time. Then the aim of getting new samples from the existing samples is lost. In other words, we can say that the aim is to train a model in such a way that it mimics the data distribution of a new sample from the existing one.

Generative modeling is incomplete without discussing its counterpart, discriminative modeling. For discriminative modeling, each sample in the training dataset must have a label. For this reason, discriminator resembled similar to supervised learning and a generator to the unsupervised learning approach. The discriminator's job is to act as a classifier to differentiate between "real" or already existing samples and "fake" or newly generated samples through the generator network. The difference between both types of modeling is shown in Table 1 through mathematical probabilistic representation

Let us have a dataset with observations X. Suppose the observation in the dataset follows some rule and generate samples to unknown distributions, P_{data}. A generative model is one that mimics the distribution of original dataset into $P_{model.}$ Also,that the P_{model} distribution should have unique samples in space and follows similar distribution as $P_{data.}$

The purpose of generative modeling is successful if we get the following conditions satisfied:

- It can generate samples that seem to appear from $P_{data.}$
- It must generate observations that are unique from the ground truth observations in X.

In a more summarized way, we can say that a model should not reproduce the samples it has aready seen.

B. Generative Adversarial Networks

The model becomes popular in 2016 at the NIPS conference by google researcher [10]. The generative architecture consists of two adversaries' network that is in a regular competition of fooling each other, one is called generator neural network and another one is discriminator neural network. Fig. 2 represents the basic network architecture of GAN.

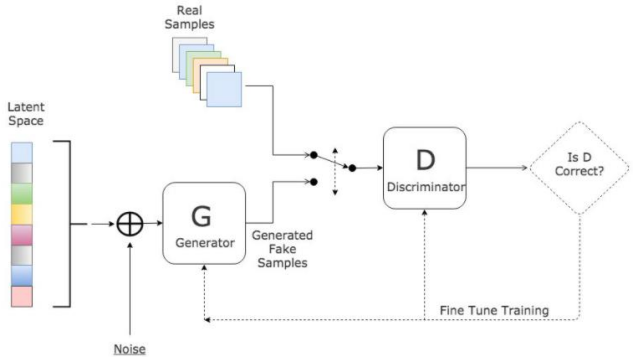

Fig 2: Architecture of GAN

Modelling	Probability Estimation	Explanation
Generative	$p(x)$	The probability of observing the observation
Discriminative	$\rho(y \mid x)$	The probability of new sample 'y' given observation 'x'

The generator aims to generate samples that have similar data distribution as 'real' data points by adding random noise to the network. The network tries to convert this random noise distribution to the original dataset distribution and prepares a 'fake' dataset which should look like as it samples from a real dataset. These generated samples are fed as input to the discriminator network which to judge the "fake" samples from the generator or "real" samples from original ground truth samples. It takes the feature vector of the standard normal distribution of the real dataset and outputs the image of the same shape as the original image dimension. This vector helps the generator network to choose a random point in latent space such that they are not unique and differentiable from the rest data points already present in the real dataset but also follow a similar distribution. This mapping of the high-dimensional vector with an image based on the previous dataset is a very common task in these models for an image processing task.

The discriminator acts as a classifier between two labels to predict real Vs fake. If it is real, then we save the sample else the loss due to the fake sample is fed back to the generator through a feedback loop. The process of fooling the discriminator goes on till it accepts all successive samples generated by the generator network. The rejection of the generator's sample depends on how strictly the discriminator is trained. The GAN training is done by alternate freezing of each network and the key lies here in what manner the two models have to train alternatively so that the generator network is more capable towards the real samples and can easily fool the discriminator.

We can easily train the discriminator as a binary classifier where two classes of images are categorized as 'real' and 'fake'. The predicted output is either '1' for real images or '0' for fake generated images from the generator. This can be

 DOI : 10.35291/2454-9150.2021.0278 © 2021, IJREAM All Rights Reserved.

treated as a purely supervised learning problem. Therefore, training discriminators is not a big hurdle.

Training the generator is more challenging since it doesn't have any training in real data-point mapping in latent space. Instead, we have only noise and input multi-variate feature vector of the standard normal distribution that has to fool the discriminator that is we are expecting output probability close to 1. More importantly, while training we have to freeze the weights of the generator network so that discriminator's weights should get updated and vice-versa. The loss function used in the discriminator is binary cross-entropy because of its binary classification nature. Our target is to train the generator more efficiently not because the discriminator is weak but to strongly train the generator.

C. GAN based Super – Resolution Models.

The LR image is a basic requirement of the generator adversary as an input. The generator architecture then tries to up-sample the image from LR to SR. The generated SR image is passed into the discriminator that acts as a basic binary classifier to differentiate between a 'real' ground truth HR image and a 'fake' generated SR image. During this competent process, the error between 'real' and 'fake' images generates the adversarial loss which is then back-propagated into the generator architecture. Again, the same process of generating SR image and comparison is done, which modifies the generating capability of a model by the feedback of adversarial loss.

For image enhancement purposes, the first GAN-based model Super-Resolution-GAN (SRGAN) [11] is most popular in 2017 and produces significant changes in image improvement. The architectures use the training of an HR image and a down-sampled version of that HR image referred to as an LR image.

Even for other models like Enhanced SRGAN, ESRGAN [12] and Edge-ESRGAN (EESRGAN) [13], the training dataset consists of patches of LR images for the generator network as an input. However, datasets of LR images are not available, and researchers depend on generating LR images by using common techniques [14][15] like down-scaling, blurring, interpolation methods, or by adding noises.

SRGAN is based on PSNR based image quality evaluation method which directly correlates the pixel-wise difference. This difference is not satisfying the visual perception of the human system and misses the high-frequency details like edges of objects in an image. The max-pooling is avoided in every CNN layer and Batch-Normalization (BN) layers are used to avoid over-fitting. To improve the features, a gaussian filter followed by an image downscaling operation is used to produce the respective LR version of the ground truth image. The model performs well on images but not much suitable for video enhancement in real-time. Also, the introduction of BN layers helps in regularization but introduces some irrelevant artifacts in natural images like a patch of grass on the skin of zebra.

It is further modified by using residual blocks and dense connections in ESRGAN [12] to ease the training process and improving the perceptual quality of an image that is lost in SRGAN [11]. Dense connections remove the problem of vanishing gradients, re-usability of features in preceding layers with a reduction in model size too. It helps in the ease of training the model. This feature is utilized in the SR field for SRDenseNet [16] that adopts dense block for 69-layers construction. Also, ESRGAN employed with realistic discriminator rather than an original discriminator that classifies the image as "more realistic" or "less realistic" in the form of a probability score and thus improved the textural details leads to extraction from the feedback network.

III. IMAGE LOW-RESOLUTION METHODS

An image resolution refers to the number of pixels that are displayed per inch of an image. An image LR task is a very common pre-processing method during an SR of an image [17] via GANs. The model becomes popular in 2014 at the NIPS conference by google researcher [10]. The generative architecture consists of two adversaries' network that is in a regular competition of fooling each other, one is called generator and another one is called discriminator. For GAN-based SR models, the LR image counterpart of the respective HR ground truth dataset is required to be fed as an input to the generator. For this reason, researchers need to use and judge the best image LR technique so that it results in best-generated SR images that can easily fool the discriminator. Here, we are using different blurring techniques for image LR to re-implement and compare them to find which available blurring method is best to feed as an input to the generator. The comparison is done based on our short experiment on 10 colored images and results are evaluated through available image quality metrics.

A. Down-scaling

Image down-scaling [18] is one of the oldest techniques of converting an HR image to an LR image. It is typically used in image data transmission, image compression, bandwidth utilization, and fast processing time.

Each pixel in an image signifies one byte that again consists of 8 bits. For an image with M rows and N columns, measures M*N pixels, total MN number of bytes are required to accumulate the disk space per image. But at the same time, it fails to capture minute details and features of an image, hence, loss of data from the original image. Here, a key point that arises is how an image can be down-scaled while keeping most of the details saved. This is defined by the compression ratio that tells us the effectiveness of the down-scaled image. It is defined as the ratio of original image size to down-scaled image size. There are two broad classes of image compression algorithms, lossless compression, and lossy compression.

In lossless compression, the compressed image is numerically identical to the original image. Common file

 DOI : 10.35291/2454-9150.2021.0278 © 2021, IJREAM All Rights Reserved.

formats that use this lossless compression algorithm are PNG and GIF. The techniques used in lossless image compression classified as Run-length coding and Huffman coding. Run-length coding is based on the replacement of long array pixel values to a sequence of a shorter pixel array. The tuple here represents two sequence of values as (A_i, B_i) where A_i represents intensity of pixel and B_i is the frequency of occurrence. The run-length coding supports BMP and TIFF image file formats. On the other hand, the Huffman coding technique uses the bottom-up approach rather than the top-down approach, using a frequency-sorted binary tree. This compression technique is generally used for "discrete" data, like spreadsheets, database records, word files, and even some kinds of raw image and video information. Text compression is a popularly used area by this technique.

In lossy compression, extra redundant bits having similar details and information are permanently removed which is unnoticeable to the user and maintains the uniformity in an image. The most commonly used image file format is a JPEG format. The techniques used in lossy image compression classified as Discrete Cosine Transform (DCT) and fractal compression. In DCT, an image is separated into sub-bands of 8*8 or 16*16 blocks on which DCT is calculated for each block. In fractal compression, the image is circulated in a loop such that the next reconstructed down-sampled image is the true replica of the original image.

B. Blurring

An effective way to reduce the pixel density of an image is by blurring. Blurring smoothens the colors and edges of an image by averaging out sudden changes in pixel intensities. Blurring techniques mainly work on the principle of passing a “low pass filter” through the image, which helps to remove high-frequency (change of pixel value) content from the image (noise) while keeping the majority of the image information intact. It allows the low frequency to enter and stop high frequency. Here, a change in frequency resembles to change in pixel intensity value.

IV. BLURRING TECHNIQUES

Blurring is used to avoid the unnecessary noise and smoothen the image. In this empirical study, we considered the four main blurring methods such as averaging or box blur, gaussian blur, median blur, and bilateral blur for re-implementation on colored images. A digital image can be represented as a matrix of pixels, where each pixel has a value representing the intensity of color. For this study, we only considered Red-Green-Blue (RGB) colored images.

In a gray-scale image, a color depth is added to a binary image, which ranges from darkest black to brightest white. Each pixel would have an intensity between 0 and 255, with 0 being darkest black and 255 being brightest white. F (x, y) would then give the intensity of the image at pixel position (x, y), assuming it is defined over a rectangle, with a finite range [0, 255].

In RGB images, f (x, y) is now a vector of three values instead of one. Where the three channels, R, G, and B are represented. Each pixel of the image has three channels and is represented as a 1x3 vector. Each channel has integer values from 0 to 255.

A. Averaging or Box Blur

In averaging or box blur [19] technique, a box blur which is a “low pass” spatial domain linear filter is convolved on an image to normalize it. In this process, the moving average filter replaces each pixel, with the average value of all the pixels present in the filter window (kernel area). This filter is the simplest of all. As each pixel after applying the filter is the mean of its kernel neighbors, therefore, all of them contribute with equal weights, for forming the final image. A 3x3 normalized box filter would look like as in Eq. 1.

$$K = \frac{1}{9} * \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (1)$$

B. Median Blur

Median blur is a non-linear blurring technique, which is used to lower the resolution and helpful in removing mainly salt-pepper noise from the image. In this median blur process, the moving filter replaces each pixel by the median of all pixels present in the filter window (kernel area) [20]. By applying the median blur, as the median is being calculated for each kernel window, a new value is not being created, the central pixel in a kernel is always replaced by a pixel value present in the image. Thus, an under-represented pixel value in the filter window will not affect the result. As a result, the edges are preserved and it is primarily used to lower the resolution and reduce noise.

C. Gaussian Blur

In a Gaussian pixel, near to the center of a kernel are given more importance than those away from the center [21]. The process is repeated channel by channel for all three channels (RGB) in the case of colored images and a single channel in the case of gray-scale images. Larger kernels average out more pixels near to their center which implies larger kernels produce more blurring effect than smaller ones. Fig. 3 represents 2-dimensional plot of gaussian function.

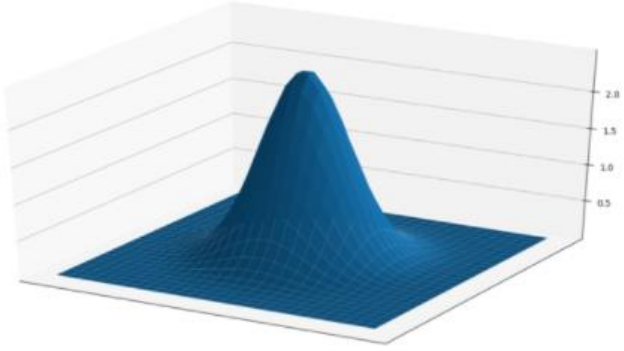

Fig 3: 2D Represention of gaussian function

Let us suppose the plot is overlaid over the kernel having a

 DOI : 10.35291/2454-9150.2021.0278 © 2021, IJREAM All Rights Reserved.

gaussian filter. The height of the plot represents the weight of the kernel i.e. pixels near to center are more important than pixels away from the center to the filtered pixel color. Standard deviation is responsible for flattening the shape of the curve in the gaussian function. The higher the standard deviation or sigma value, the flatter the curve, and the smaller sigma value results in a more pronounced peak. This implies higher sigma value removes more noise in an image and smooth more but at the same time remove finer details from an image.

D. Bilateral Blur

Bilateral blur [22] is a non-linear, edge-preserving, and noise-reducing blurring technique for images. It is an improved version of the existing gaussian blur technique. As in gaussian blur, using a gaussian kernel, each pixel is replaced by a weighted average of the pixels present in the kernel window, in which the weights are inversely proportional to the distance from the center of the neighborhood. Similarly, the bilateral blur begins with the linear gaussian blurring. In addition to it, the bilateral blur technique adds a total weight such that pixel values that are close to the pixel value in the center as shown in Eq. 2. are weighted more than pixel values that are more different.

$$g(x) = (f * G^s)(x) = \int f(y) G^s(x - y)\, dy \qquad (2)$$

The weights in gaussian blur depend solely on, the weight for f(y) equals $G^s(x - y)\, dy$ and is only dependent on the spatial distance that is |x-y|.

The bilateral filter adds a weighting term that depends on the total distance f(y)-f(x). g(x) is the kernel in spatial domain as given by Eq. 3

$$g(x) = \frac{\int_R f(y)(G^s(x-y)G^t(f(x)-f(y))dy}{\int_R (G^s(x-y)G^t(f(x)-f(y))\, dy} \qquad (3)$$

For colored images, the three channels use a color distance that prevents the scope of blurring from occurring in only one channel. Fig. 4 shows all different blurred techniques on one image sample out of 10 randomly selected colored images.

V. IMAGE QUALITY MATRIX

In this study, we have considered three popular image quality metrics from state-of-the-art literature such as PSNR, SSIM, and FSIM to evaluate output images from our experiments.

a) Average Blur

b) Gaussian Blur

c) Median Blur

d) Bilateral Blur

Fig 4: Different Blurring Effect on Coloured Image

A. Peak Signal to noise Ratio

PSNR [23], [24], [25] is used as an image quality metric to determine the quality of a reconstructed image compared to the original image. It is a ratio between the maximum possible power of an image to the power of the corrupting noise that affects the quality of the reconstructed image. Moreover, PSNR is derived from the Mean Squared Error (MSE), which represents the cumulative squared error between the reconstructed image and the original image as represented.

$$MSE = \frac{\sum_{M,N}[I1(m,n)-I2(m,n)]^2\{M*N\}}{M*N} \qquad (4)$$

where, M = no of rows of pixel, N = no of columns of pixel, R = 255 (number of levels) for 8-bit representation of pixels, m = row index and n = column index. Definition of PSNR is represented in Eq. 5.

$$PSNR = 10 \log_{10} \left(\frac{R^2}{MSE}\right) \qquad (5)$$

For color color images with three (RGB, one for each channel) values per pixel, the definition of PSNR is the same except the MSE is the sum over all squared value differences for each color channel divided by image size and by three. Generally, a higher PSNR value indicates that the reconstructed image has better quality compared to the original image. However, PSNR has been shown to perform poorly compared to other image quality metrics when it comes to estimating the quality of images as perceived by human eyes.

 DOI : 10.35291/2454-9150.2021.0278 © 2021, IJREAM All Rights Reserved.

B. Structural Similarity Index Measure

SSIM [26, 27, 23] is another image quality metric that is considered to be correlated with the quality perception of the Human Visual System (HVS) [28]. In SSIM, the traditional error summation is not calculated, instead, it is designed by modeling any image distortion as a combination of three factors that are loss of correlation, luminance distortion, and contrast distortion. Thus, the SSIM value tries to find the similarity in the original and reconstructed image based on the three terms which are luminance, contrast, and structural similarity. Mathematically, SSIM is defined in Eq 6.

$$SSIM(x,y) = [l(x,y)]^{\alpha} * [c(x,y)]^{\beta} * [s(x,y)]^{\gamma} \quad (6)$$

SSIM is the multiplicative combination of the three terms, where l(x,y) = luminance, c(x,y) = contrast and s(x,y) = structural similarity. The l(x,y) represents the luminance is given by Eq. 7.

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x{}^2 + \mu_y{}^2 + C_1} \quad (7)$$

First, the luminance of each signal is compared. Assuming discrete signals, this is estimated as the mean intensity, this factor becomes 1, iff $\mu_x = \mu_y$. The luminance comparison function l(x,y) is then a function of μ_x and μ_ySecond, we remove the mean intensity from the signal. The contrast factor c(x,y) is represented in Eq. 8.

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x{}^2 + \sigma_y{}^2 + C_2} \quad (8)$$

The contrast is defined as the difference between the intensities of two images. It is measured by the standard deviation of two images σ_x and σ_y. The contrast factor c(x,y) measures the similarity in the contrast of the original and reconstructed image. The factor is maximum to value 1, if $\sigma_x = \sigma_y$ and $C_3 = \frac{C_2}{2}$. The structural similarity is a measure of the correlation coefficient between the two images. It uses the covariance between the two images. Eq. 9. represents the structural similarity between input and transformed pixel.

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (9)$$

C. Feature Similarity index Measure

FSIM [29] was proposed based on the fact that the HVS observes an image based on its low-level features. It maps the features and measures the similarity between the original and the reconstructed image by comparing luminance components between the two images. The image quality assessment brings from pixel-based to structure-based stage.

FSIM is a two-stage process:

- the computation of the local similarity map.
- the pooling of the calculated similarity map into a single similarity score.

The value of FSIM is dependent on two criteria, Phase Congruency (PC) and Gradient Magnitude (GM). The PC is primary in FSIM which is a dimensionless feature that gives an understanding of the local structure of an image. It is contrast invariant and slight variation in the image. However, it can detect the features of a given image. As PC is contrast invariant, however, HVS image perception is also dependent on the contrast factor, so the criteria GM is used. GM is obtained by convolution of an image with convolution masks called gradient operators with common ones being the Sobel operator, the Prewitt operator, and the Scharr operator.

For gradient magnitude calculation, if f(x) is an image, $G_x$and G_y are virtual and horizontal gradients. Then the value of GM will be represented by Eq. 10.

$$\sqrt[2]{G_x{}^2 + G_y{}^2} \quad (10)$$

Now based on these two criteria, the FSIM value is calculated and given by Eq. 11.

$$FSIM = [S_{PC}(x,y)]^{\alpha} * [S_{GM}(x,y)]^{\beta} \quad (11)$$

FSIM is the multiplicative combination of the product of the values of PC and GM values obtained using the original and the reconstructed image. For simplicity, the constants α and β are taken equal to one. Here, the PC and GM values of the two images are compared shown in Eq. 12 and Eq. 13 respectively.

$$S_{PC} = \frac{2*PC_x*PC_y+t_1}{PC_x{}^2+PC_y{}^2+t_1} \quad (12)$$

Here, PC_x, PC_y, GM_x and GM_y refers to the phase congruency and gradient magnitude of original image, and reconstructed image, respectively. The constants t_1 and t_2 are added to avoid negative denominator.

VI. Comparative Analysis

In this section, the considered evaluation metrics performance is summarized. For experimenting with RGB images, 10 random images are used from the "Flikr8k" dataset [30]. Further, conclusions are drawn from the obtained result for each metric based on the higher the metric value the better the technique.

A. PSNR Comparison

PSNR is mostly expressed in the logarithmic decibel scale and is often used for image analysis. From Fig. 5, we can observe PSNR metric behavior for colored images. For a bit depth of 8, the practically acceptable values are between 30 and 50 dB. It can be observed easily that the bilateral blur represented by red colour stands high compared to the rest methods as depicted in Fig. 5. Consequently, through the PSNR evaluation metric, the bilateral blurring outperforms the rest methods.

 DOI : 10.35291/2454-9150.2021.0278 © 2021, IJREAM All Rights Reserved.

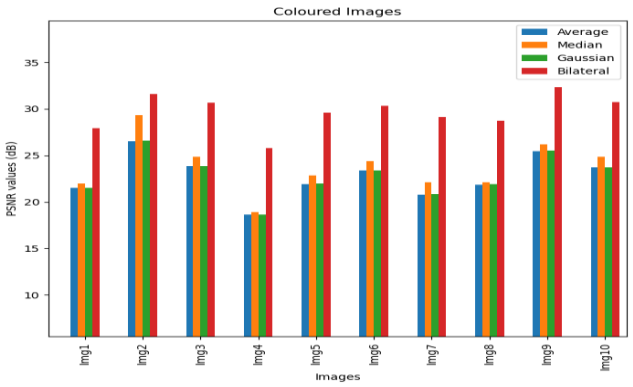

Fig 5: PSNR for coloured Images

B. *SSIM Comparison*

SSIM is a relatively better metric than PSNR [23]. Practically values lie in the range [0.2,0.8], the values close to 1 like 0.95, 0.98 represents very good image reconstruction. From Fig. 6, we can observe SSIM metric behavior for colored images. It compares L1 loss pixel by pixel. It is a referenced metric in the sense that its values are calculated based on the distortion-free image as a reference image. It can be noticed easily that the bilateral blur represented by red colour is better than the rest methods as can be seen in Fig. 6. So, through the SSIM evaluation metric, the bilateral blurring again shows excellent performance on all 10 colored images.

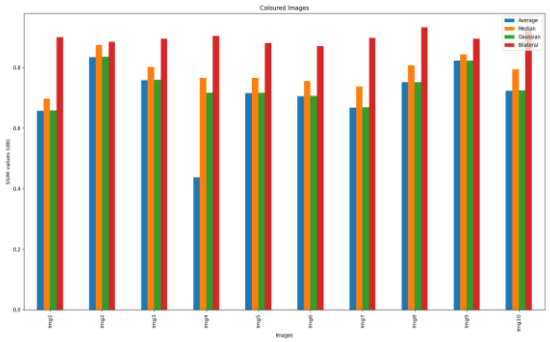

Fig 6: SSIM for coloured Images

C. *FSIM Comparison*

FSIM observes and understands an image based on its low-level features and transforms in human visual system features. Practically values lie in range [0.3,0.8], the values close to 1 like 0.95, 0.98 represent good image reconstruction. The main variant of FSIM is PC that carries contrast information and affects the HVS perception towards the image quality. Here, from Fig. 7, we can observe that the bilateral blur stands tall again compared to the rest methods. Thus, the FSIM evaluation metric also proves that the bilateral blurring technique is best than the rest on all 10 tested colored images.

VII. Analysis & Results

This study is important for image pre-processing methods to fed generator suitable LR-based technique so that it produces SR samples with enhanced quality for GAN-based architectures. Our work is performed on natural (RGB) colored images from the "Flikr8k" dataset.

The results concluded from our experiments are as follows:

1) The re-implemented four LR blurring techniques are evaluated based on a comparative analysis of potential image quality metrics like PSNR, SSIM, and FSIM. For all metrics, the higher the value, the better the image quality.

2) Through our experimental observations and comparison, the bilateral blurring method performs best on PSNR, SSIM, and FSIM metrics. The worst method is the average blur among all, followed by Gaussian blur, followed by median blur, and followed by the best bilateral blurring method as can be seen in Fig. 5, Fig. 6, and Fig. 7.

3) Moreover , SSIM and FSIM are normalized metrics while MSE and PSNR are not. If we compare image quality metrics, PSNR is not good as per the human visual perception system since it doesn't capture the structural intelligence of an image and calculates the difference between two adjacent pixels which hardly depicts the natural artifacts present in an image according to the human visual system. SSIM is the best image quality metric than other metrics since it captures the finest structural details of an image.

VIII. Conclusion

In this paper, we presented an empirical comparative study using various existing blurring techniques such as average, gaussian, median, and bilateral. We explained how each considered blurring technique works and re-implemented them. The re-implementation of each technique was performed over 10 random colored images. Next, the obtained results with our experiments are examined using popular image quality metrics named PSNR, SSIM, and FSIM respectively. Based on the achieved results and detailed comparison, we conclude that the bilateral blurring method performs better on PSNR, SSIM, and FSIM. On the other hand, the least performing method is the average blurring among all. Our empirical study creates an added value to the state-of-the-art literature by helping practitioners, researchers, and developers to choose suitable blurring methods and image quality metrics.

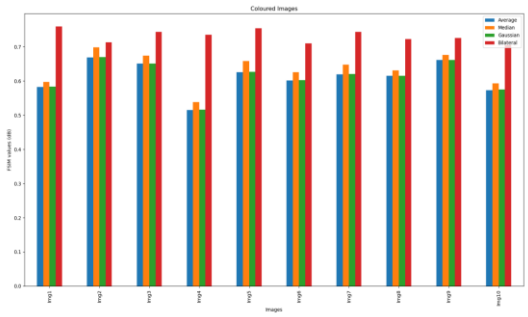

Fig7: FSIM for Colored Images

 DOI : 10.35291/2454-9150.2021.0278 © 2021, IJREAM All Rights Reserved.

REFERENCES

[1] S. Qu, R. Hu, S. Chen, J. Jiang, Z. Wang, M. Zhang, Super-resolution for face image with an improved k-nn search strategy, China Communications 13, pp. 151–161. 2016

[2] K. Gribbon, D. Bailey, A novel approach to real-time bilinear interpolation, in: Proceedings. DELTA 2004. Second IEEE International Workshop on Electronic Design, Test and Applications, pp. 126–131., 2004, doi:10.1109/DELTA.2004.10055

[3] X.-g. Zhang, A new kind of super-resolution reconstruction algorithm based on the icm and the bicubic interpolation, in: 2008 International Symposium on Intelligent Information Technology Application Workshops, 2008, pp. 817–820. doi:10.1109/ IITA.Workshops.2008.12

[4]] Z. Wang, J. Chen, S. C. H. Hoi, Deep learning for image super-resolution: A survey, 2020. arXiv:1902.06068

[5] A. Lukin, A. Krylov, A. Nasonov, Image interpolation by super-resolution (2010).

[6] C. Dong, C. C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, 2015.

[7] J. Kim, J. K. Lee, K. M. Lee, Deeply-recursive convolutional network for image super-resolution, 2016. arXiv:1511. 04491.

[8] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2790–2798. doi:10.1109/CVPR.2017. 298.

[9] J. G. Kreifeldt, "An analysis of surface-detected EMG as an amplitude-modulated noise," presented at the 1989 Int. Conf. Medicine and Biological Engineering, Chicago, IL.

[10] R. Sim, G. Dudek, Learning generative models of scene features, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, volume 1, 2001, pp. I–I. doi:10.1109/CVPR.2001.990504.

[11] C. Ledig, L. Theis, F. Huszar, 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, 2017. arXiv:1609.04802.

[12] N. C. Rakotonirina, A. Rasoanaivo, Esrgan+ : Further improving enhanced super-resolution generative adversarial network, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 3637–3641. doi:10.1109/ ICASSP40776.2020.9054071.

[13] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, D. Chao, Small-object detection in remote sensing images with end-to-end edge-enhanced gan and object detector network, 2020. arXiv:2003.09085.

[14] H. Kim, M. Choi, B. Lim, K. M. Lee, Task-aware image downscaling, in: ECCV, 2018.

[15] W. Sun, Z. Chen, Learned image downscaling for upscaling using content adaptive resampler, IEEE Transactions on Image Processing 29 (2020) 4027–4040.

[16] T. Tong, G. Li, X. Liu, Q. Gao, Image super-resolution using dense skip connections, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 4809–4817. doi:10.1109/ICCV.2017.514.

[17] L. Wang, M. Zheng, W. Du, M. Wei, L. Li, Super-resolution sar image reconstruction via generative adversarial network, in: 2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE), 2018, pp. 1–4. doi:10.1109/ISAPE.2018.8634345.

[18] E. Pardo-Iguzquiza, M. Chica-Olmo, P. Atkinson, Downscaling cokriging for image sharpening, Remote Sensing of Environment 102 (2006) 86– 98.

[19] T. Singh, Comparative analysis of image deblurring techniques, International Journal of Computer Applications 153 (2016) 39–44.

[20] C. Chang, J. Hsiao, C. Hsieh, An adaptive median filter for image denoising, in: 2008 Second International Symposium on Intelligent Information Technology Application, volume 2, 2008, pp. 346–350. doi:10.1109/IITA.2008.259.

[21] E. S. Gedraite, M. Hadad, Investigation on the effect of a gaussian blur in image filtering and segmentation, in: Proceedings ELMAR-2011, 2011, pp. 393–396.

[22] P. Kornprobst, J. Tumblin, F. Durand, Bilateral filtering: Theory and applications, Foundations and Trends in Computer Graphics and Vision 4 (2009) 1–74.

[23] A. Hore, D. Ziou, Image quality metrics: Psnr vs. ssim, in: 2010 20th ´ International Conference on Pattern Recognition, 2010, pp. 2366–2369. doi:10.1109/ICPR.2010.579.

[24] M. Vranjes, S. Rimac-Drlje, K. Grgic, Locally averaged psnr as a simple objective video quality metric, in: 2008 50th International Symposium ELMAR, volume 1, 2008, pp. 17–20.

[25] M. Cammina, R. Lancini, M. Marconi, A perceptual psnr based on the utilization of a linear model of hvs, motion vectors and dft-3d, in: 2000 10th European Signal Processing Conference, 2000, pp. 1–4.

[26] Zhou Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (2004) 600–612.

[27] S. Jiao, W. Dong, Sar image quality assessment based on ssim using textural feature, in: 2013 Seventh International Conference on Image and Graphics, 2013, pp. 281–286. doi:10.1109/ICIG.2013.62.

[28] J. Delaigle, C. Devleeschouwer, B. Macq, L. Langendijk, Human visual system features enabling watermarking, in: Proceedings. IEEE International Conference on Multimedia and Expo, volume 2, 2002, pp. 489–492 vol.2. doi:10.1109/ICME.2002.1035653.

[29] L. Zhang, L. Zhang, X. Mou, D. Zhang, Fsim: A feature similarity index for image quality assessment, IEEE Transactions on Image Processing 20 (2011) 2378–2386.

[30] M. Hodosh, P. Young, J. Hockenmaier, Flickr8k dataset, Journal of Artificial Intelligence Research 47 (2013).

 DOI : 10.35291/2454-9150.2021.0278 © 2021, IJREAM All Rights Reserved.