ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)| Impact Factor: 7.293|

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

Exploring Generative Adversarial Networks (GANs) For Realistic Image Synthesis

Mohit Mittal

Dr. A.P.J. Abdul Kalam Technical University, Lucknow, Uttar Pradesh, India

ABSTRACT: Image synthesis is the process of generating new images from the ground up, frequently utilizing preexisting data or models. By definition, super-resolution methods produce supplementary image content and features that are not present in the original input in order to reconstruct a high-resolution image from a low-resolution source. Surpassing the performance attained with high-resolution images is a challenge when training or analyzing models with low-resolution images. It is not always easy to obtain a higher-resolution image. Recognizing and identifying objects in low-resolution images is a challenging task. As a result, it is imperative to develop a method that simultaneously enhances the resolution of the low-resolution image and enhances its quality. Generative Adversarial Networks (GANs) and other generative models are increasingly acknowledged for their capacity to accurately replicate high-resolution counterparts. This article provides a performance comparison of DCGAN and Wasserstein GAN for image synthesis via image super resolution. The experimental endeavor employs the Set5 image data set. Wasserstein GAN outperforms DCGAN in terms of PSNR, SSIM, and VIF parameters.

KEYWORDS: Generative Adversarial Networks,DCGAN, WassersteinGAN Realistic Image Synthesis, Super-Resolution, Deep Learning

I. INTRODUCTION

Image synthesis refers to the creation of new images from the ground up, often using existing data or models. Super-resolution approaches, by definition, generate supplementary image content and features absent in the original input to reconstruct a high-resolution image from a low-resolution source [1].

Each image has an own array of attributes, including the hue and intricacies of an object or scene, encoded inside its individual pixels. In fact, there is a considerable amount of additional information that may be included inside each pixel of an image. It is evident that each pixel has a distinct tint. Nonetheless, the red, green, and blue (often abbreviated as R, G, and B) constituents of a pixel dictate its hue. By mixing these three fundamental colors at varying intensities, we can represent the visible spectrum of each pixel.

The level of detail in an image is closely correlated with its resolution. A better quality image has more pixels and more detail, whereas a lower resolution image contains fewer pixels and less detail. The pixel count in a picture is one metric of its quality; however, the data contained inside those pixels is also a significant factor. A higher resolution image has a greater amount of information inside its pixels [2].

Training or analyzing models with low-resolution pictures presents challenges in surpassing performance achieved with high-resolution images. Acquiring a higher-resolution image is not always straightforward. Recognizing and identifying objects in low-resolution images is difficult. Consequently, a method to improve the quality of the low-resolution image while concurrently augmenting its resolution is necessary. This approach is referred to as image super-resolution [3].

Image super-resolution is used in media creation, medical diagnostics, satellite imaging, and surveillance. Super-resolution techniques significantly enhance low-resolution images obtained from surveillance cameras. This facilitates the execution of tasks such as facial recognition, detection, and identification with enhanced precision. In medical

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)| Impact Factor: 7.293|

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

diagnostics, super-resolution may convert low-resolution images into high-resolution ones. This is especially beneficial when acquiring high-resolution MRI images presents difficulties. Super-resolution facilitates the creation of high-resolution images from a small segment of a satellite photograph, despite the whole image being captured from a considerable distance and including thousands of pixels. The utilization of high-resolution images, which need more storage space and transmission time, is particularly advantageous for reducing their total cost. To save costs, super-resolution techniques may be used to immediately enhance low-resolution images.

Super-resolution images may be generated from either a single low-resolution picture or many low-resolution photographs; these techniques are known as Single-Image Super-resolution and Multi-Image Super-resolution, respectively. In contrast to multi-image Super-resolution, which use many low-resolution photos of the same subject to generate a high-resolution image, Single-image Super-resolution utilizes a solitary low-resolution snapshot to produce a high-resolution image. Single-image Super-resolution techniques are constrained by a limited amount of input data, potentially resulting in the reconstruction of a super-resolution image with misleading patterns that do not correspond to the original image's context [4].

Due to its extensive data set, Multi-Image Super-resolution is expected to surpass its competition. However, in several application cases with constrained resources, the computational expense may escalate significantly, rendering it unworkable. Moreover, capturing several low-resolution images of the same item is both impractical and time-intensive. Thus, the practical significance of Single-Image super-resolution surpasses the intricacy of the problem formulation [5].
It is well acknowledged that single-image super-resolution constitutes a multi-solution ill-posed problem. Due to the inability to individually resolve the missing information, it is challenging to accurately identify high-resolution features from a solitary low-resolution image [6].

This presents an intrinsic problem. [7].

Researchers have devised many algorithms and approaches to tackle this challenge, using their current comprehension of the image's structure and properties. Statistical, interpolation, prediction, patch, and edge-based methodologies are among the most conventional techniques for generating high-resolution images. Recent research have presented learning-based methodologies, such as deep learning frameworks and machine learning, that are more sophisticated and effective [8].

Currently, researchers use Convolutional Neural Networks (CNNs) to generate analogous high-resolution images. Very Deep Convolutional Networks (VDSR), Deeply Recursive Convolutional Networks (DRCN), and Super-Resolution Convolutional Neural Networks (SRCNN) are the predominant frameworks. Generative Adversarial Networks (GANs) and other generative models are becoming recognized for their ability to reliably reproduce high-resolution counterparts [9].

The primary objective of picture super-resolution is to enhance the visual appeal and functionality of low-resolution images for applications such as printing, video conferencing, and surveillance by augmenting their overall quality. Imaging system limitations, including low-resolution sensors and image compression techniques, need image super-resolution methods to recover high-resolution information that is either missing or degraded. Image super-resolution enhances the resolution of low-quality pictures, hence increasing the precision of image analysis, identification, and interpretation. It may also assist in digital photography, medical imaging, and remote sensing by enabling the generation of high-quality images from low-quality sources [10].

The quality of an image is influenced by several factors. The most common issues include suboptimal shooting conditions (e.g., fuzzy or dim photographs), insufficient lighting, lens attributes (e.g., noise, blur, or flare), and artefacts resulting from post-processing (e.g., lossy compression methods), among others. The resolution may influence the image quality. Images of low resolution fail to capture significant information due to the limited number of pixels used to represent an object. To efficiently perform tasks such as item categorization and facial recognition, Artificial

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)| Impact Factor: 7.293|

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

Intelligence (AI) addresses this difficulty by using a low-resolution image, referred to as a super-resolution image, to reconstruct a corresponding high-resolution image. Surveillance cameras and mobile phones are prevalent sources of low-resolution (LR) images, often characterized by worse quality.

The predominant methods for image enlargement include bilinear, bicubic, and nearest neighbor interpolation, which upscale a low-resolution image to a higher resolution. Rather of incorporating additional information into the image, these methods only augment existing dimensions, sometimes resulting in a blurred or pixelated appearance and diminishing its quality. Nonetheless, deep learning-based image super-resolution seeks to recover absent or degraded high-resolution details from low-resolution images. The technique involves instructing a neural network model on the complex interconnections between the two domains by training it on a collection of picture pairings of differing resolutions. Subsequently, the model may accept a low-quality input image and generate a high-resolution output. Deep learning-based picture super-resolution surpasses image enlargement in quality and aesthetics, exhibiting fewer artefacts and reduced blurriness. However, it is more time-intensive and computationally expensive because to the need for extensive training data.

A key motivator for deep learning-based image super-resolution is the enhancement of visual quality and detail in low-resolution photographs. The conversion of low-resolution images to high-resolution ones may enhance several practical applications, including digital zooming, surveillance systems, medical imaging, and the improvement of compressed picture quality.

II. RELATED WORK

The area of machine learning known as deep learning makes use of the data that is presented in order to automatically learn the input-output link. Traditional task-specific learning algorithms, on the other hand, depend on expert domain knowledge to pick acceptable handcrafted features. Deep learning approaches, on the other hand, automatically generate hierarchical representations via hidden layers. Beginning with the early strategies that were based on Convolutional Neural Networks (CNNs), there have been several efforts made to apply deep learning models for SR tasks. These attempts have progressed to more promising approaches that use Generative Adversarial Networks (GANs) [11]. Different network topologies, loss functions, learning concepts, and methodologies are the primary ways in which the deep learning-based SR algorithm family varies from one another. Other differences include the principles of learning.

It was the Super-Resolution Convolutional Neural Network, also known as SRCNN, that was the first framework to learn a comprehensive mapping from low-resolution to high-resolution images. After then, more deep convolutional neural network (CNN) techniques for improving the quality of the SR image were proposed in the research literature [11]. The vast majority of CNN-based SISR algorithms upsample the LR image that is generated by using either a direct or progressive technique. Pre-upsampling and post-upsampling are the two primary kinds of direct methods to SR. Pre-upsampling procedures are more common than post-upsampling approaches. Pre-upsampling SR approaches [12, 13] upsample the LR image by using an appropriate interpolation technique. This is done prior to the use of a convolutional neural network (CNN) for the purpose of recovering the high-frequency information that was lost in an upsampled LR picture. On the other hand, these methods need a massive amount of memory since they rely on upsampled pictures, which in turn increases the amount of computing that is required. In order to circumvent these issues, many SR techniques that are founded on post-upsampling have been developed. These approaches directly extract features from the LR observation. For the purpose of obtaining the final SR image, the features that have been recovered are upscaled at the very end of the network by using either sub-pixel [15] or transposed [14] convolution. In light of this, SR techniques that depend on post-upsampling have reduced computing costs and memory requirements in comparison to those that rely on pre-upsampling. Here, numerous strategies have been described in detail in References [11,16] in order to reduce the amount of memory and processing that the SR method requires. But both methods result in the appearance of chequerboard artifacts in the SR picture. This is due to the fact that they use a bigger number of upsampling layers in order to create the image, which results in a higher upscaling factor (i.e., ×4 ×8).

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)| Impact Factor: 7.293|

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

According to the review of the relevant literature [16], the use of a progressive upsampling process as opposed to a direct approach has the potential to enhance the quality of the SR results. In order to train their models, the CNN-based SR techniques that were stated before make use of pixel-wise loss functions (i.e., L1, L2, Charbonnier) and perform an excellent job of retrieving the high-frequency information that is absent from the LR data. The existing body of research indicates that these loss functions have the potential to improve quantitative performance, as evaluated by metrics such as the Structural Similarity Index Measure (SSIM) and the Peak Signal-to-Noise Ratio (PSNR). This is accomplished by maintaining the average of all possible estimated SR images. Despite the fact that the PSNR and SSIM measurements have great values, it is nevertheless feasible for the SR data to have a hazy appearance and a lack of high-frequency characteristics according to [8].

In an effort to circumvent this problem, the Generative Adversarial Network (GAN), which is a relatively recent tool for the SISR challenge, has been developed [6–10]. The generator network in GAN-based SR techniques is able to give SR solutions with better high-frequency features, which are subsequently sampled in natural manifolds. This is made possible by the discriminator network and proper loss functions. This, in turn, causes the generated SR photographs to seem to be extremely similar to the HR images that are based on the ground truth. In their paper [11], Ledig and colleagues presented Single Image Super-Resolution using GAN (SRGAN), which was the first SISR approach that was based on GAN. Instead of relying on pixel-wise loss functions, the SRGAN model was trained to attain state-of-the-art performance in terms of perceptual quality of SR results. This was accomplished by using a unique perceptual loss derived from the high-level feature mappings of the VGG network in conjunction with a traditional adversarial loss function. In order to further improve the SR performance of the SRGAN model, a number of GAN-based SISR models make use of a variety of training techniques and loss functions that are unique to them [6,7,8,9,10]. It was mentioned before that human perception does not match very well with traditional methods of quality measurement such as PSNR and SSIM. An SR image that has a greater PSNR and SSIM does not necessarily include more high-frequency information. This is not something that can be assumed. In order for researchers to validate the visual quality of surveillance images, they need metrics that were based on human perception. One of these metrics is called a Mean Opinion Score (MOS), and it is calculated by taking into account the points that a number of human observers have assigned to the second-order image. It is not possible for humans to provide a quality score or an error label to the recently made image in a reliable manner. The community of computer vision researchers is currently looking for an alternate method in order to achieve their goal of developing a more accurate quantitative measure that is equivalent to MOS.

III. METHODOLOGY AND RESULTS

3.1 DCGAN

A Deep Convolutional Generative Adversarial Network (DCGAN) is a type of Generative Adversarial Network (GAN) that employs deep convolutional networks to construct the generator and discriminator, thereby improving the quality and stability of image generation. In 2015, the article "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks" [17] was published by Alec Radford, Luke Metz, and SoumithChintala, which introduced DCGAN. It has paved the way for more intricate designs, such as WGAN and StyleGAN, by being one of the most potent GAN versions.

Figure 1 illustrates that GAN is composed of two neural networks. Generator: In this instance, a neural network generates fabricated data—images—from arbitrary noise. A discriminator neural network will be capable of distinguishing between generated data and genuine data from the training set. The generator and discerning are engaged in a min-max game:

- Generator: aims to deceive the discriminator by generating data that is indistinguishable from genuine data.
- Discriminator: endeavors to differentiate between synthetic and actual data.

The objective of concurrently training both networks is to have the generator produce data that is so realistic that the discriminator is unable to differentiate between genuine and fraudulent data.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)| Impact Factor: 7.293|

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

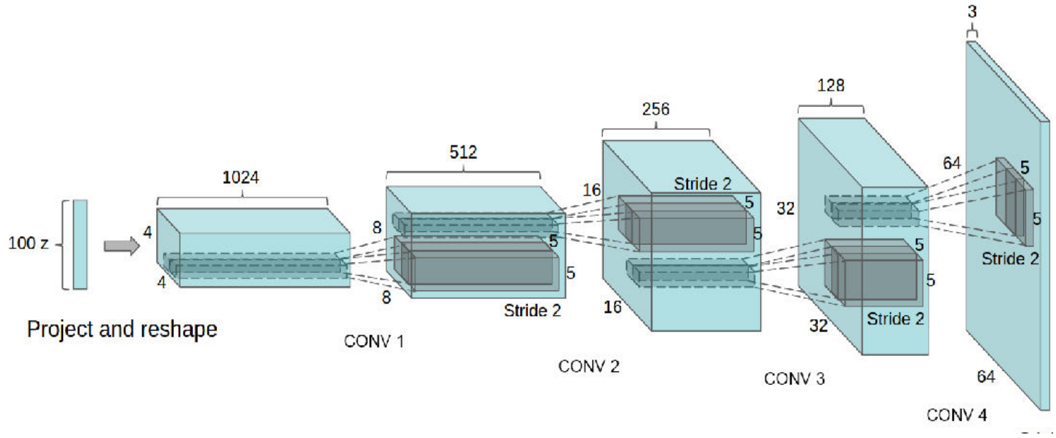


Figure 1: DCGAN Architecture [17]

The generator network generates fake images from random noise. Typically, its design comprises the following:
1. Input: Typically, input is a random noise vector that is a 100-dimensional vector derived from a normal distribution.
2. Layer 1: fully linked layer; subsequent noise reconfiguration as a tensor (e.g., 4x4x1024)
3. Convolutional layers. A sequence of transposed convolutional layers, also known as deconvolutions, is used to upsample the noise vector into a full-size image. Batch normalization and leaky ReLuare intermediate processes.
4. Output: A picture spanning -1 to 1 is generated by a tanh activation function.

Discriminator The discriminator network evaluates images to determine their authenticity or falsity. Typically, its architecture includes the following:
1. Input: a photograph, either authentic or generated by the generator.
2. Convolution Layer- The image is progressively reduced in size as hierarchical characteristics are extracted through the use of multiple convolutional layers. Utilize batch normalization and leaky ReLU between layers.
3. Output: A probability of genuine against false is produced by a sigmoid activation function.

3.2 **Wasserstein GAN**

Wasserstein GAN (WGAN) is an extension of the original **Generative Adversarial Network (GAN)** architecture that aims to address some of the key issues faced by standard GANs, particularly **training instability** and **mode collapse**. WGAN introduces a novel loss function, the **Wasserstein distance** (also known as **Earth Mover's Distance**), to improve the stability of GAN training and produce higher-quality generative models.

WGAN was introduced by **Martin Arjovsky, SoumithChintala, and Léon Bottou** in their paper *"Wasserstein GAN"* in 2017 [18]. In this explanation, we will break down the core concepts of WGAN, its architecture, its key innovations, and how it resolves some of the issues found in traditional GANs. It is shown in figure 2

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)| Impact Factor: 7.293|

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

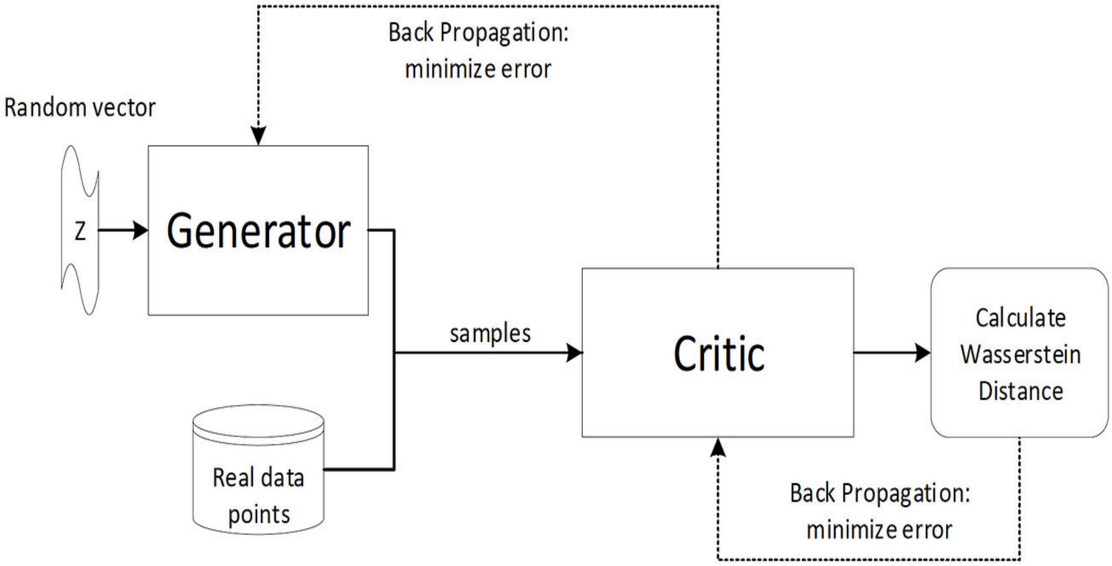


Figure 2: Wasserstein GAN

Before diving into Wasserstein GANs, it's important to briefly recap the **Generative Adversarial Network** (GAN) framework:

- **Generator (G)**: The generator produces fake data (e.g., images) from random noise (latent vectors). It learns to generate data that resembles the real data distribution.
- **Discriminator (D)**: The discriminator's job is to distinguish between real data (from the training set) and fake data (from the generator). It assigns a probability score to each sample, indicating whether the data is real or fake.

The generator and discriminator are trained simultaneously in a **minimax game**, where the generator tries to improve at fooling the discriminator, while the discriminator tries to improve at detecting fake data.

- **GAN Loss**: In the traditional GAN, the generator tries to minimize the discriminator's ability to distinguish real from fake data, while the discriminator tries to maximize its ability to differentiate.

This training procedure, however, often leads to **instabilities** (i.e., when training doesn't converge properly or the model produces poor-quality images). This problem is most often caused by the **vanishing gradient problem**, where the discriminator becomes too good at its job, and the generator receives little to no feedback for improving its output.
The core innovation behind **Wasserstein GAN** is the replacement of the traditional GAN loss function with the **Wasserstein loss** (also known as the **Earth Mover's Distance, EMD**), which measures the distance between two probability distributions.

The **Wasserstein distance** intuitively represents the minimum amount of "work" required to transform one distribution into another, by "moving" the probability mass. This is a more robust and informative metric compared to the **Jensen-Shannon divergence** used in traditional GANs. Wasserstein distance provides a smoother and more continuous loss landscape, which helps resolve many of the training issues seen in standard GANs.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)| Impact Factor: 7.293|

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

In WGAN, the **critic** (not the discriminator) learns to approximate the **Wasserstein distance** between the real and generated distributions. The critic's role is to assign a **real-valued score** to both real and fake data, and the loss function is defined as:
LWGAN=Ex∼Preal[D(x)]−Ex∼Pfake[D(G(z))]\mathcal{L}_{WGAN} = \mathbb{E}_{x \sim P_{\text{real}}} [D(x)] - \mathbb{E}_{x \sim P_{\text{fake}}} [D(G(z))]LWGAN=Ex∼Preal[D(x)]−Ex∼Pfake[D(G(z))]
Where:

- D(x)D(x)D(x) is the critic’s score for the real data xxx,
- G(z)G(z)G(z) is the generated data from the generator GGG,
- PrealP_{\text{real}}Preal and PfakeP_{\text{fake}}Pfake are the real and fake data distributions, respectively.

This loss function encourages the generator to produce outputs that **minimize the Wasserstein distance** between the generated distribution and the real distribution, effectively pushing the generated data closer to the real data.

1. **Critic instead of Discriminator**:
- In WGAN, the **discriminator** from the traditional GAN is replaced by a **critic**.
- The critic outputs real-valued scores (rather than probabilities) for both real and fake data.
- The critic's goal is to approximate the Wasserstein distance between the real and generated data distributions.
2. **Weight Clipping**:
- To enforce the **Lipschitz continuity** required for Wasserstein distance, the weights of the critic are constrained using **weight clipping**.
- This means that the weights of the critic are limited to a certain range (typically [-0.01, 0.01]). Weight clipping ensures that the critic’s function is a **1-Lipschitz function**, which is a necessary condition for the Wasserstein distance to be valid.
3. **Training Procedure**:
- **Critic**: The critic is trained multiple times per generator update. In practice, the critic is updated around 5 times for each update of the generator.
- **Generator**: The generator is updated once every few critic updates. This allows the critic to approximate the Wasserstein distance more accurately before updating the generator.

3.3 Results

Set5 [19] image dataset contains five popular images: one medium size image (‘baby’, 512 × 512) and four smaller ones (‘bird’, ‘butterfly’, ‘head’, ‘women). Set5 dataset is widely used for image super-resolution tasks due to its small size, diverse content, and availability of ground truth images.Set five images are shown in figure 3 below. Output images are shown in figure 4. Performance comparison of DCGAN and Wasserstein GAN are shown in Table 1, Table 2, figure 5 and figure 6. Wasserstein GAN is performing better than DCGAN on PSNR, SSIM and VIF parameters.

Figure 3: Set 5 Images

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)| Impact Factor: 7.293|

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

Figure 4: Output Images using the Set5 dataset. (a) Input Image (b) Output image using DCGAN (c) Output image using Wasserstein GAN

Numerous criteria are crucial when assessing super-resolution technology. These measurements provide numerical approaches for evaluating the quality of visual reconstruction. The fidelity is evaluated using the Peak Signal-to-Noise Ratio (PSNR), which quantifies the peak signal power relative to the noise power. The structural similarity index (SSIM) evaluates the correspondence between the ground truth and the reconstructed image for structure, contrast, and luminance. The term "visual information fidelity" (VIF) refers to the preservation of visual features and textures. The average squared difference between the original data and the reconstructed image is referred to as mean squared error (MSE). Numerous metrics provide a comprehensive assessment by quantifying various aspects of image quality and performance.

The primary image evaluation criteria included in our testing were the Structural Similarity Index (SSIM), Mean Squared Error (MSE), and Peak Signal-to-Noise Ratio (PSNR). PSNR assesses the greatest discrepancy between the original high-resolution image and the reconstructed SR image, whereas SSIM examines the extent of structural similarity between the two images. The mean squared error (MSE) quantifies the divergence between the actual and super-resolved pictures.

The Peak Signal-to-Noise Ratio (PSNR) is a metric for signal quality that compares the peak signal intensity to the mean squared error between the ground truth and reconstructed signals. This numerical measure evaluates the extent to which signal reconstruction or compression processes introduce noise or diminish quality. Their simplicity and clarity are the grounds for their extensive use. PSNR is valuable for assessing various compression techniques and understanding their effects on image or video quality.

The Structural Similarity Index (SSIM) assesses the extent of structural similarity between two pictures by evaluating criteria such as brightness, contrast, and structural elements. SSIM offers a more precise assessment than pixel-wise metrics by accounting for the nuances of human visual perception. It is robust because to its ability to endure variations in contrast, brightness, and other aberrations. In the context of image restoration and compression assessments, SSIM is an excellent option due to its capability to capture perceptually significant features and structural characteristics. Mean Squared Error (MSE) quantifies signal distortion or inaccuracy by averaging the squared differences between the original and reconstructed signals. This metric offers significant advantages in computing efficiency and ease of usage.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)| Impact Factor: 7.293|

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

It serves as a fundamental parameter for evaluating the performance of algorithms in image or video processing and is often used as a baseline for comparisons.

Visual Information Fidelity (VIF) assesses the quality of reconstructed signals by using visual fidelity and perceptual content, hence evaluating the retention of visual information in images or movies. VIF effectively captures essential structural nuances and visual intricacies by considering both local and global elements. In domains where visual perception is essential, such as medical imaging and video surveillance, VIF provides a valuable method for evaluating the quality of pictures or videos, considering the intricacies of human vision.

Table 1: Performance Comparison of DCGAN and Wasserstein GAN for Image 1

Model Name	PSNR	SSIM	VIF
DCGAN	31.72	0.88	0.90
Wasserstein GAN	33.95	0.91	0.92

Table 2: Performance Comparison of DCGAN and Wasserstein GAN for Image 2

Model Name	PSNR	SSIM	VIF
DCGAN	30.56	0.891	0.88
Wasserstein GAN	32.87	0.906	0.94

Figure 5: Comparison of DCGAN and Wasserstein GAN for Image 1

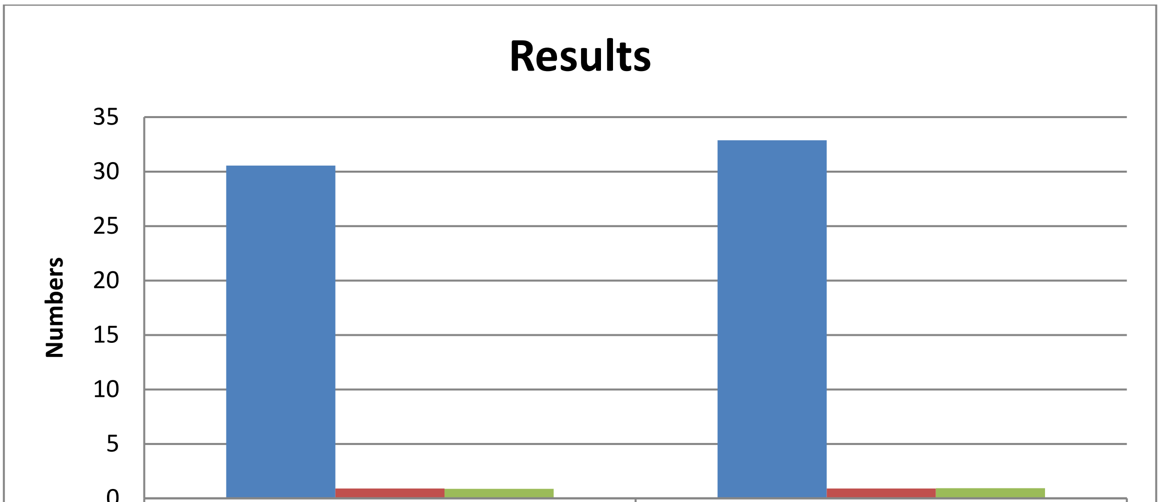
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)| Impact Factor: 7.293|

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018



	DCGAN	Wasserstein GAN
PSNR	30.56	32.87
SSIM	0.891	0.906
VIF	0.88	0.94

Figure 6: Comparison of DCGAN and Wasserstein GAN for Image

IV. CONCLUSION

Generative models, such as Generative Adversarial Networks (GANs), are increasingly capable of accurately simulating their high-resolution counterparts. The primary objective of picture super-resolution is to enhance the overall quality of low-resolution photographs, thereby enhancing their appearance and functionality for applications such as surveillance, video conferencing, and printing. The recovery of lost or degraded high-resolution information is necessitated by image super-resolution techniques, which are necessary due to imaging system constraints, including low-resolution sensors and image compression mechanisms. The lucidity of low-resolution images is enhanced by image super-resolution, resulting in more precise identification, analysis, and interpretation. Its capacity to enhance low-quality input images may prove advantageous in the fields of medical imaging, remote sensing, and digital photography. This paper contrasts the efficacy of DCGAN and Wasserstein GAN in picture synthesis by employing picture super resolution as a benchmark. The experimental research employs the Set5 picture data acquisition. Wasserstein GAN surpasses DCGAN in terms of PSNR, SSIM, and VIF metrics.

REFERENCES

[1] C. Dong, C. C. Loy, K.He, and X. Tang, "Learning a deep convolutional network for image super-resolution," *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 184–199, 2014.
[2] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *IEEE Trans. Comput. Imaging*, vol. 3, no. 4, pp. 1–11, Dec. 2017.
[3] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1646–1654, 2016.
[4] B. Yang, H. K. Wen, S. Wang, R. Clark, A. Markham, and N. Trigoni, 3D object reconstruction from a single depth view with adversarial learning, in Int. Conf. Computer Vision Workshops (ICCVW), 2017, pp. 679–688

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)| Impact Factor: 7.293|

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

[5] W. Y. Wang, Q. G. Huang, S. Y. You, C. Yang, and U. Neumann, Shape inpainting using 3D generative adversarial network and recurrent convolutional networks, in The IEEE Int. Conf. Computer Vision (ICCV), Venice, Italy, 2017, pp. 2298–2306.
[6] Q. F. Chen and V. Koltun, Photographic image synthesis with cascaded refinement networks, in The Int. Conf. Computer Vision (ICCV), Venice, Italy, 2017.
[7] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1125–1134, 2016.
[8] Y. P. Zhou and T. L. Berg, Learning temporal transformations from time-lapse videos, in Proc. 14th European Conf. Computer Vision (ECCV), Amsterdam, The Netherlands, 2016.
[9] C. Vondrick, H. Pirsiavash, and A. Torralba, Generating videos with scene dynamics, in Proc. 29th Conf. Neural Information Processing Systems, Barcelona, Spain, 2016, pp. 613–621.
[10] J. Y. Zhu, P. Krahenb ¨ uhl, E. Shechtman, and A. A. ¨ Efros, Generative visual manipulation on the natural image manifold, in Proc. 14th European Conf. Computer Vision (ECCV), Amsterdam, The Netherlands, 2016
[11] H. Zhang and X. C. Cao, Magic pencil: Generalized sketch inversion via generative adversarial nets, in Proc. SIGGRAPH ASIA 2016 Posters, Macau, China, 2016.
[12] S. Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele, and H. Lee, Learning what and where to draw, in Proc. 29th Conf. Neural Information Processing Systems, Barcelona, Spain, 2016, pp. 217–225.
[13] K. Jiwon, L. Jung Kwon, and L. Kyoung Mu, “Deeply-recursive convolutional network for image super-resolution,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1637–1645.
[14] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in European conference on computer vision. Springer, 2016, pp. 391–407
[15] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and ´ Z. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1874–1883.
[16] T. Tong, G. Li, X. Liu, and Q. Gao, “Image super-resolution using dense skip connections,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 4799–4807.
[17] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv*, Nov. 2015.[Online]. Available: https://doi.org/10.48550/arXiv.1511.06434.
[18] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv preprint arXiv:1701.07875*, 2017. [Online]. Available: https://arxiv.org/abs/1701.07875
[19] https://huggingface.co/datasets/eugenesiow/Set5