CHAPTER 1

Super-resolution-based GAN for image processing: Recent advances and future trends

Meenu Gupta[a], Meet Kumari[b], Rachna Jain[c], and Lakshay[c]
[a]Department of Computer Science & Engineering, Chandigarh University, Ajitgarh, Punjab, India
[b]Department of Electronics & Communication Engineering, Chandigarh University, Ajitgarh, Punjab, India
[c]Department of Computer Science & Engineering, Bharati Vidyapeeth's College of Engineering, Delhi, India

1.1 Introduction

One can think of the term machine as older than the computer itself. In 1950, the computer scientist, logician, and mathematician Alan Turing penned a paper for the generations to come, "Computing Machinery and Intelligence" [1]. Today, computers can not only match humans but have outperformed them completely. Sometimes people think about not achieving the superhumanity face recognition or cleaning the medical image of the patient accurately, even for the small algorithm, as a machine learning algorithm is the best at pattern reorganization in existing image data using features for tasks such as classification and regression. When we try to generate new data, however, the computer has struggled [2]. An algorithm can easily defeat a chess grandmaster, classify whether a transaction is fraudulent or not, and classify in a medical report whether the given medical report has any disease or not, but fail on humanity's most basic and essential capacities—including crafting an original creation or a pleasant conversation. Mahdizadehaghdam et al. [3] proposed some tests named the Imitation game, also known as the Turing Test. Behind a closed door, an unknown observer talks with two counterparts means a computer and a human.

In 2014, all of the above problems were solved when Ian Goodfellow invented generative adversarial networks (GANs). This technique has enabled computers to generate realistic data by using two separate neural networks. Before GANs, different ways have been proposed by the programmer to analyze the generated data. But the result received from the generated data was not up to the mark. When GANs were introduced the first time, it showed a remarkable result as there was no difference between the generated fake images or photograph-image and gave the same result as the real-world-like quality. GANs turn scribbled images to a photograph-like image [4].

Generative Adversarial Networks for Image-to-Image Translation
https://doi.org/10.1016/B978-0-12-823519-5.00030-0

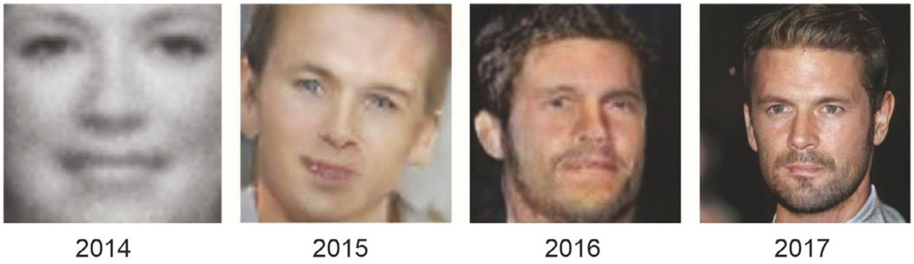

Fig. 1.1 Improves realism of the image as general adversarial networks varies [4].

In recent years, how far GANs have changed the meaning of generating or improving the real image is shown in Fig. 1.1. Fig. 1.1 was first produced by GAN in 2014 and shows how human faces continuously improve in generating fake images. The machine could produce as a blurred image, and even that achievement is celebrated as a success. In just the next 3 years, we could not classify which is fake or which qualifies as high-resolution portrait photographs [5].

GANs are a category of machine learning techniques that uses two simultaneously trained models: the first is the generator to generate fake data and the discriminator is used to discrete the raw data from the real dataset images. The word generative indicates creating new data from the given data. GAN generates the data which learn from the choice of the given training set. The term adversarial points to maintaining the dynamic between the two models that are the generator and the discriminator. Here two networks are continually trying to trick the other as the generator generates better fake images to get convincing data. The better discriminator is trying to distinguish the real data examples from the fake generated ones. The word networks indicates the class of machine models. The generator and the discriminator commonly use the neural network. As a complex, the neural network is more complex than the implementation of GAN [6].

GAN has two models. First, it works where we put the input and then we get the output. The goal is to form two models that combine and run simultaneously so that the first discriminator receives input from the real data that come from the training dataset, and the second time onward there are two input sources that are the actual data and the fake examples coming from the given generator. A random number vectors is passed through the generator. The output acquired from the generator is Fake examples that try to convince as far as possible the real data. The discriminator predicted the probability of the input real. The main purpose of creating two models separately is to overcome the problem of fake data that is generated from the training dataset. The discriminator's goal is to differentiate between the fake data generated from the generator and the real input example from the dataset. This section further discusses the training parts of the discriminator and the generator in Sections 1.1.1 and 1.1.2 [7].

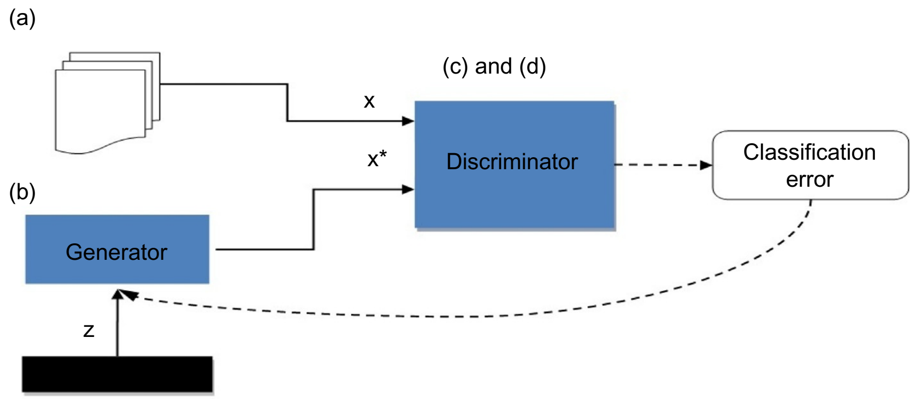

Fig. 1.2 Train the discriminator.

1.1.1 Train the discriminator

Fig. 1.2 discusses the trained model of the discriminator and the steps are as follows [8]:

(a) First, get a random real example x from the given training dataset.

(b) Now get a new random vector z and, utilizing the generator network, synthesize a fake example as x^*.

(c) Utilize the discriminator network to distinguish between x^* and x.

(d) Find the classification error and back-propagate. Then try to minimize the classification error to update the discriminator biases and weight.

1.1.2 Train the generator

Fig. 1.3 shows the trained model of the generator as you can see the labeling of these steps as follows:

(a) First, choose a random new image from the dataset as vector z, using a generator to create an x^*, i.e., a fake example.

(b) It used a discriminator to categorize real and fake examples.

(c) Find the classification error and back-propagate. Then try to minimize the classification error due to which the total error to renovate the discriminator biases and weight [9].

1.1.3 Organization of the chapter

This chapter is further classified into different sections. Section 1.2 discusses the background study of this work and also different research views. Section 1.3 discusses the SR GAN model for image processing. Section 1.4 discusses the application-based GAN case studies to enhance object detection. Section 1.5 discusses the open issues and challenges faced in the working of GAN. Section 1.6 concludes the chapter with its future scope.

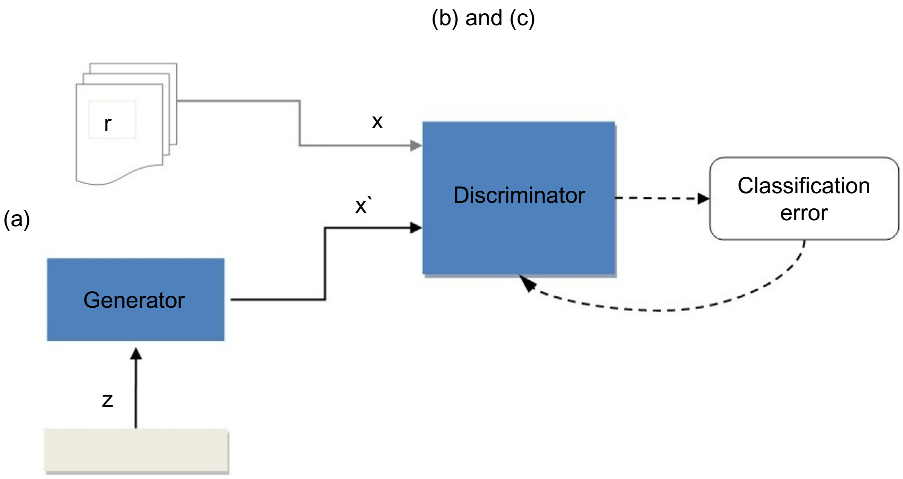

Fig. 1.3 Train the generator.

1.2 Background study

The goal of Perera et al. [10] is to determine whether the given query is from the same class or different class. Their solution is based on learning the latent representation of an in-class example using a de-noising and auto-encoder network. This method gives good results for COIL, MINST f-MINST dataset.

They give a new thinking to GANs as in face recognition we create fake images, which help to identity theft and privacy breaches. In this chapter, they proposed a technique to recognize the forensic face. They use the deep face recognition system as a core for their model and create fake images repeatedly to help data augmentation [11].

Tripathy et al. [12] present a generic face animator that controls the pose and expression using a given face image. They implemented a two-stage neural network model, which is learned in a self-supervised manner.

Rathgeb et al. [13] proposed a supervised deep learning algorithm using CNNs to detect synthetic images. The proposed algorithm gives an accuracy of 99.83 for distinguishing real images from dataset and fake images generated using GANs.

Yu et al. [14] have shown advances to complete a new level as they proposed a method to visualize forensic and model attribution. The model supports image attribution, enables fine-grained model authentication, persists across different image frequencies, fingerprint frequencies and paths, and is not biased.

Lian et al. [15] provide a guidance module that is introduced in FG-SRGAN, which is utilized to reduce the space of possible mapping functions and helps to learn the correct mapping function from a low-resolution domain to a high-resolution domain. The guidance module is used to greatly reduce adversarial loss.

Takano and Alaghband [16] proposed the SRGAN model. In this chapter, they solved the problem of sharpening the images. It can give a slight hint on how the real image looks like from the blurry image as they convert a low-resolution image to a high-resolution image.

Dou et al. [17] proposed PCA-GAN, which greatly improves the performance of GAN-based models on super-resolving face images. The model focuses on cumulative discrimination in the orthogonal projection space spanned by PCA projection to details into the discriminator.

Jiang et al. [18] proposed to improve the perception of the CT image using SRGAN, which leads to greatly enhance the spatial resolution of the image, as perception increases the disease analysis on a tiny portion of areas and pathological features. They introduced a diluted convolution module. The mean structural similarity (MSSIM) loss is also introduced to improve the perceptual loss function.

Li et al. [19] provide an improvement method of SRGAN and a solution for the problem of image distortion in textile flow detection, a super-resolution image reconstruction. Here the result of an experiment shows that the PNSR of SRGAN is 0.83 higher than that of the Bilinear, and the SSIM is higher than 0.0819. SRGAN can get a clearer image and reconstruct a richer texture, with more high-frequency details, and that is easier to identify defects, which is important in the flaw detection of fabrics.

Wang et al. [20] decided to use dense convolutional network blocks (dense blocks), which connect each layer to every other layer in a feed-forward manner as our very deep generator networks. GAN solves the problem of spectral normalization as the method offers better training stability and visual improvements.

Nan et al. [21] solved the complex computation, unstable network, and slow learning speed problems of a generative adversarial network for image super-resolution (SRGAN). We proposed a single image super-resolution reconstruction model called Res_WGAN based on ResNeXt.

Li et al. [22] discussed an edge-enhanced super-resolution network (EESR), which proposed better generation of high-frequency structures in blind super-resolution. EESR is able to recover textures with 4 times upsampling and gained a PTS of 0.6210 on the DIV2K test set, which is much better than the state-of-the-art methods.

Sood et al. [23] worked on magnetic resonance (MR) images to obtain high-resolution images for which the patients have to wait for a long time in a still state. Obtaining low-resolution images and then converting them to high-resolution images uses four models: SRGAN, SRCNN, SRResNet, and sparse representation; among them, SRGAN gives the best result.

Lee et al. [24] present a super-resolution model specialized for license plate images, CSRGAN, trained with a novel character-based perceptual loss. Specifically, they focus on character-level recognizability of super-resolved images rather than pixel-level reconstruction.

Chen et al. [25] divided the technique into two different parts: the first one is to improve PSNR and the second one is to improve visual quality. They propose a new dense block, which uses complex connections between each layer to build a more powerful generator. Next, to improve perceptual quality, they found a new set of feature maps to compute the perceptual loss, which would make the output image look more real and natural.

Jeon et al. [26] proposed a method to increase the similarity between pixels by performing the operation of the ResNet module, which has an effect similar to that of the ensemble operation. That gives a better high-resolution image.

As the resolutions of remote sensing images are low, to improve the performance we required high-level resolution. In this chapter, they first optimize the generator and residual-in-residual dense without BN (batch normalization) is used. Firstly GAN (relativistic generative adversarial network) is introduced and then the sensation loss is improved [27].

1.3 SR-GAN model for image processing

Image super-resolution is defined as an increase in the size image, but trying to not decrease the quality of the image keeps the reduction in quality to a minimum or creates a high-resolution image from a low-resolution image by using the details from the original image. This problem has some difficulties as for an input low-resolution image, and there are some multiple solutions available. SR-GAN has numerous applications like medical image processing, satellite image, aerial image analysis, etc. [28].

1.3.1 Architecture of SR-GAN

Many programs that are good, fast, and accurate get a single image super-resolution. But still, something that is missing is the texture of the original features of the image. That is the way where we recover the low-resolution image so that the image produce is not distorted. Later we recover these errors, but it is not complete all errors that are produced. The main error shows that result has a peak signal-to-noise ratio (PSNR) high thus provides good image quality results, but lacking high-frequency details. The previous result also sees the similarity in pixel space, which leads to a blurry or unsatisfying image. Due to this, we introduce SR-GAN, a model that can capture the perceptual difference in the ground truth image and the model output. Fig. 1.4 discusses the architecture of SRGAN [29].

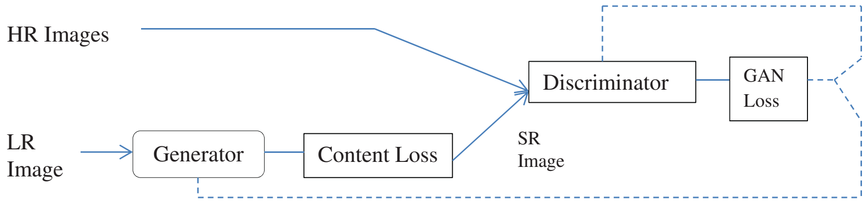

Fig. 1.4 Architecture of SRGAN [29].

The training algorithm of SRGAN is shown in the following steps:

(a) We run the HR (high-resolution) images to get sample LR (low-resolution) images. To train our dataset, we required both LR and HR images.

(b) Then allow LR images to pass through the generator, which increases the samples and provides SR (super-resolution) images.

(c) LR and HR images are classified by passing through the discriminator and back-propagated [30].

Fig. 1.5 presents the network between the generator and the discriminator. It contains convolution layers, parameterized ReLu(PrelU), and batch normalization. The generator also implements skip connections similar to ResNet [31].

1.3.2 Network architecture

Residual blocks are defined as seep learning networks that are difficult to train. The residual learning framework makes the training easier for the networks and enables them to go deep substantially, to improve the performance. In the generator, there are a total of 16 residual blocks used [32].

As in Generator 2, a subpixel is used for getting the feature map up-sampling. Every time pixel shuffle is applied it rearranges the elements of the L*B*H*r*r tensor and transforms into the rL*rB*H tensor. With increase in computation, the bicubic filter from the pipeline has been removed. We use parameterized rely on instead of Relu pr LeakyRelu. Prelude adds a learnable parameter, which leads to learning the negative part coefficient adaptively. The convolution layer "k3n64s1" represents kernel filters of 3*3 outputting channels 64 along with stride 1. Similarly, "k3n256s1" and "K9n3s1" are other convolution layers added [33].

1.3.3 Perceptual loss

As in the below equation, LSR shows the perceptual loss, and it is a commonly used model based on the mean square error. As the equation is a loss function, it calculates the loss function and gives a solution concerning characteristics. Here LSR_x is notated

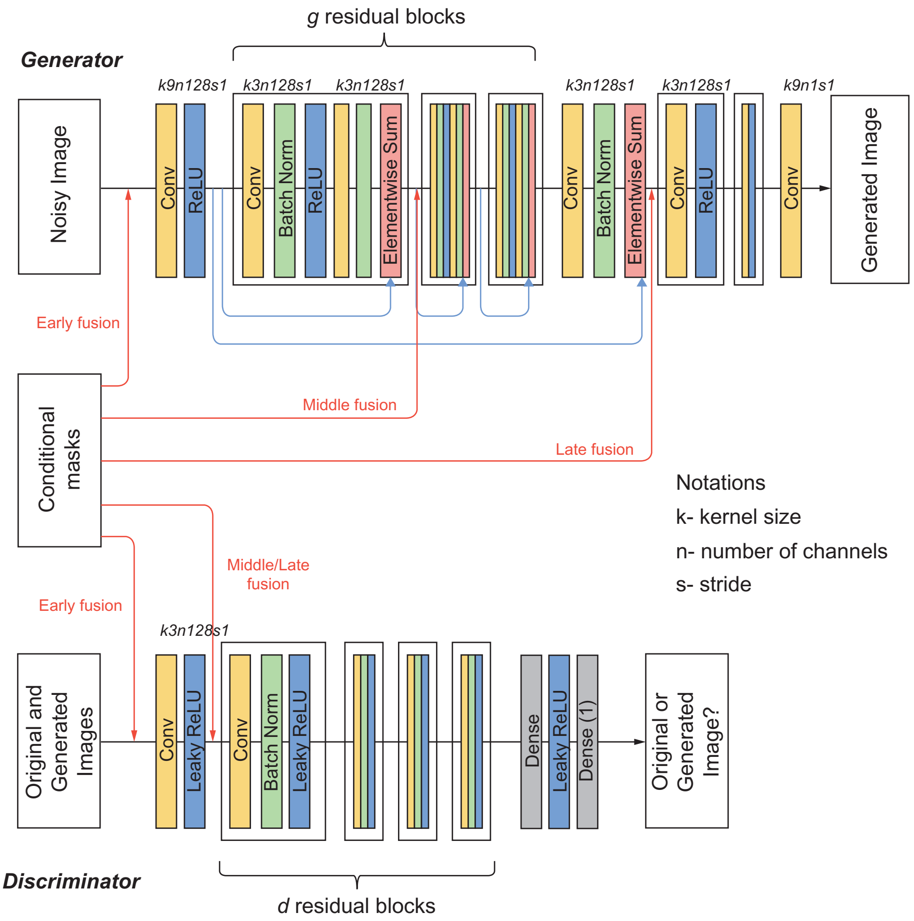

Fig. 1.5 SRGAN-based model architecture for seismic images [30].

as a content loss, and it is used as the last term is an adversarial loss as shown in Eq. (1.1). The weighted sum of both gives the perceptual loss (VGG-based content loss):

$$LSR = LSR_x + 10^{-3} LSR_{Adv} \tag{1.1}$$

1.3.3.1 Content loss

The pixel-wise mean square error loss is evaluated as

$$LSR_{MSE} = \frac{1}{r^2 LB} \sum_{x=1}^{rl} \sum_{y=1}^{rb} \left(IHR_{x,y} - G_{\theta G}\left(ILR_{x,y} \right) \right)^2 \tag{1.2}$$

Eq. (1.2) is the most generally utilized advancement focus for image SR on which many best-in-class approaches depend [34, 35]. However, while accomplishing significantly high PSNR, arrangements of MSE advancement problems frequently require high frequency, which brings about perceptually unsuitable management with highly smooth surfaces [36].

Rather than depending on losses (pixel-wise), we expand on the thoughts of Shi et al. [35], Denton et al. [37], and Ledig [4] and use a loss function work that is closer to perceptual likeness. We characterize the VGG loss dependent on the ReLU enhancement layers of the preprepared 19-layer VGG model portrayed in Simonyan and Zisserman [38]. We demonstrate the element map obtained by the *n*th convolution before the *I*th max-pooling layer inside the VGG19 loss as the Euclidean separation between the component portrayal of reproduced picture $G_{\theta G}$(ILR) as shown in the Eq. (1.3) [39]

$$LSR_{VGG/ij.} = \frac{1}{L_{i,j}B_{i,j}} \sum_{x=1}^{L_{i,j}} \sum_{y=1}^{B_{i,j}} \left(\varnothing_{i,j}\left(IHR_{x,y}\right) - \varnothing_{i,j}\left(G_{\theta G}\left(ILR_{x,y}\right)\right)\right)^2 \tag{1.3}$$

Here $L_{i,\,j}$ *and* $B_{i,\,j}$ describe the length and the width of the given feature map used in the VGG system.

1.3.3.2 Adversarial loss

The last section of this chapter discusses the content loss and also included the generative part of GAN pertaining particularly to perceptual loss. It urges our system to support arrangements that dwell on the complex of regular pictures, by attempting to fool the discriminator. The generative loss LSR_{adv} is defined on the basis of the probabilities of the discriminator $G_{\theta G}(ILR)$ overall training as shown in Eq. (1.4) [40]:

$$LSR_{adv} = \sum_{n=1}^{N} -\log D_{\theta D}\left(G_{\theta G}\left(ILR\right)\right) \tag{1.4}$$

where $D_{\theta D}(G_{\theta G}(ILR))$ is the probability that creates fake images $G_{\theta D}(ILR)$ as a real high-resolution image. For a good gradient, we limit the value $-\log D_{\theta D}$ from $\log 1 - x$ where x is the probability of creating fake images [41].

1.4 Case study

This includes the different case studies as applications of EE-GAN to enhance object detection, edge-enhanced GAN for remote sensing image, application of SRGAN on video surveillance, and forensic application and super-resolution of video using SRGAN.

1.4.1 Case study 1: Application of EE-GAN to enhance object detection

Detection performance of small objects in remote sensing images has not been more desirable than in huge size objects, especially in noisy and low-resolution images. Thus, enhanced super-resolution GAN (ESRGAN) provides significant image enhancement output. However, reconstructed images generally lose high-frequency edge data. Thus, object detection performance gives small objects decrement on low-resolution and noisy remote sensing images. Thus, residual-in-residual dense blocks (RRDB) for both the EEN and ESRGAN and EEN, for the detector system used a high-speed region-based convolutional network (FRCNN) as well as a single-shot detector (SSD) [42].

1.4.2 Case study 2: Edge-enhanced GAN for remote sensing image

The recent super-resolution (SR) techniques that are dependent on deep learning have provided significant comparative merits. Still, they remain not desirable in high-frequency edge details for the recovery of pictures in noise-contaminated image conditions, such as remote sensing satellite images. Thus, a GAN-based edge-enhancement network (EEGAN) is used for reliable satellite image SR reconstruction with the adversarial learning method, which is noise insensitive. Especially EEGAN comprises two primary subnetworks: an edge-enhancement subnetwork (EESN) and an ultra-dense subnetwork (UDSN). First, in UDSN, 2-D dense blocks are collected for feature extraction to gain an intermediate image in high-resolution result, which looks sharp but offers artifacts and noise. After that, EESN is generated to enhance and extract the image contours by purifying the noise-contaminated components with mask processing. The recovered enhanced edges and intermediate image can be joined to produce high credibility and clear content results. Extensive experiments on *Jilin-1* video satellite images, *Kaggle Open Source Data set as well as Digital globe provide* a more optimum reconstruction performance than previous SR methods [43].

1.4.3 Case study 3: Application of SRGAN on video surveillance and forensic application

Person reidentification (REID) is a significant work in forensics and video applications. Several past methods are based on a primary assumption that several person images have sufficiently high and uniform resolutions. Several scale mismatching and low resolution always present in the open-world REID. This is known as scale-adaptive low-resolution person re-identification (SALR-REID). The intuitive method to address this issue is to improve several low resolutions to a high resolution uniformly. Thus, SRGAN is one of the popular image super-resolution deep networks constructed with a fixed upscaling parameter. But it is yet not suitable for SALR-REID work that requires a network not only to synthesize image features for judging a person's identity but also to enhance the capability of scale-adaptive upscaling. We group multiple SRGANs in series to supplement the ability of image feature representation as well by plugging in an identification network. Thus, a cascaded super-resolution GAN (CSRGAN) framework with a unified formulation can be used [44].

1.4.4 Case study 4: Super-resolution of video using SRGAN

SRGAN techniques are used to improve the image quality. There are several methods of image transformation where the computing system gets input and sends it in the output image. GAN is the deep neural network that consists of two networks, discriminator and generator. GANs are about designing, such as portrait drawing or symphony composition. SRGAN gives various merits over methods. It proposes a perceptual loss factor that comprises the merits of content and adversarial losses. Here the discriminator block discriminates between real HR images from produced super-resolved images [45], while the generator function is used for propagating model training. Adversarial loss function utilizes a discriminator network that is trained to discriminate already between the two pictures. However, content loss function utilizes perceptual similarity despite the pixel space similarity. The superior thing about SRGAN is that it produces the same data as real data. SRGANs learn the representations that are internal to produce upscale images [46]. The neural network is faithful in photo-realistic textures that are recovered from downgraded images. The SRGAN methods demit with a high peak to signal noise ratio but also give high visual perception and efficiency. Joining the adversarial and perceptual loss will produce a high-quality, super-resolution image. Moreover, the training phase perceptual losses evaluate image similarities robustly compared to per-pixel losses. Further, perceptual loss functions identify the high-level semantic and perceptual differences between the generated images [42].

1.5 Open issues and challenges

When we train our GAN models, we suffer many major problems. Some problems are nonconvergence, model collapse, and diminished gradient unbalanced between the two models. GAN is sensitive toward the hyper-parameter factors. In GAN, sometimes the partial model is collapsed [45]. The gradient corresponds to ILR approaches zero, and then our model is collapsed. When we restart our model, the training in the discriminator detects the single-mode impact. The discriminator will take charge and change a single point to the next most likely point [46].

Overfitting is one of the main challenges as the balance between the generator and the discriminator. Some programmers give the solution. Someone proposes to use cost function with a nonvanishing gradient instead. Nonconvergence occurs due to both low and high mesh quality [47].

As we cannot apply GAN on static data due to a more complex convolution layer being required as the real and fake static data, we have not classified the data. There are some results theoretically but cannot be implemented [7].

Again, alongside various merits of GANs, there are yet open challenges that require to be solved for their medical imaging employment. In cross-modality image and image reconstruction synthesis, most tasks still adopt traditional shallow reference advantages like PSNR, SSIM, or MAE for quantitative analysis. However, these measures do not relate to the image's visual quality, e.g., pixel-wise loss direct optimization generated a blurry result but gave higher numbers compared to using adversarial loss [48]. It provides great difficulty in interpreting these horizontal comparison numbers of GAN-based tasks, particularly when other losses are presented. One method to reduce this issue is to utilize downstream works like classification or segmentation to validate the quality of the produced sample. Some other method is to recruit domain experts but this method is time-consuming, expensive, and hard to scale [49].

Today, we have applied GAN for more than 20 basic applications. All the applications have a broad area where GAN is applied, as most important are satellite images where GAN is best for training and testing of images. In medical images like MRI and X-ray images as they are of low resolution, the images and edges are not sharp enough, due to which extraction of more features is not possible with the help of SR-GAN and EE-GAN GAN [50].

1.6 Conclusion and future scope

In the past years before the discovery of GANs, image processing of satellite images or medical X-ray images was quite hard for feature extraction purposes. Classification is also somewhat hard due to the presence of a high error rate at the time. In a single image, every 1px represents at least 10m, due to which feature extraction is significantly reduced. As the images are of low quality, the objects are blurry to get the high-resolution image, due to which SR-GAN is used. As both the models run at the same time, it greatly reduces the training time. GAN is used to generate fake data in today's world. Hence, many algorithms are proposed, which make fake things appear real. GAN has several other applications, including making recipes, songs, fake images of a person, generating Cartoon characters, generating new human poses, face aging, and photo blending. These are the areas generally used in the present scenarios where GAN is freely applied. In future, by using GAN, we can create videos of robot motion and train a robot for progressive enhancement. Some researchers are working on the Novo generation of new molecules for extracting the desired properties in silica molecules. Many of the researchers are also working on the application of autonomous driving of a self-driving car using GAN.

References

[1] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, J. Jiang, Edge-enhanced GAN for remote sensing image superresolution, IEEE Trans. Geosci. Remote Sens. 57 (8) (2019) 5799–5812.

[2] V. Ramakrishnan, A.K. Prabhavathy, J. Devishree, A survey on vehicle detection techniques in aerial surveillance, Int. J. Comput. Appl. 55 (18) (2012).