CHAPTER 3

Generative adversarial networks and their variants

Er. Aarti
Department of Computer Science & Engineering, Lovely Professional University, Phagwara, Punjab, India

3.1 Introduction of generative adversarial network (GAN)

Generative adversarial networks [1] were projected to figure out the weaknesses of other generative structures also proven successful in the field of unsupervised learning. GAN has acquired wide consideration in the AI area for their capability to learn high-dimensional and complex real information circulation. In particular, they do not depend on any suppositions about the distribution and can create pictures that look genuine such as latent space in a basic way. This ground-breaking property drives GAN to be applied to different applications, for example, image translation, image synthesis, domain adaptation, image attribute editing, and other scholastic fields [2]. The most persuasive explanation that GANs are broadly considered, created, and utilized is a result of their prosperity. GANs have had the option to create photographs so sensible that people can't tell whether they are scenes, items, and individuals that do not exist in reality [3]. Generating a picture from a given text depiction has two objectives: visual authenticity and semantic consistency. Although huge advancement has been made in creating high-caliber and outwardly sensible pictures utilizing generative adversarial networks, ensuring semantic consistency between the text depiction and visual substance stays very challenging. Various interesting applications of GANs are image-to-image translation, super-resolution, semantic-image-to-photo translation, generation of new human poses, photos to emojis, photograph editing, face aging, photo blending, and many more.

In a game-theoretic scheme, the generator system is required to contend against an adversary by completing the objective, as generative adversarial networks depend on this scheme. Adversarial games are the domain of AI where two or more agents play opposite to each other. GANs are an exciting recent innovation in deep learning. GANs are one of the new state-of-the-art neural networks that can be used to do many things. Recovering corrupted data, text-to-image generation, and many more endless applications generative adversarial network has.

Generative models can be thought of as containing more information than their discriminative counterparts since they also are used for *discriminative tasks* such as *classification* or *regression*. The adversarial modeling structure is generally straight to apply when both

Generative Adversarial Networks for Image-to-Image Translation
https://doi.org/10.1016/B978-0-12-823519-5.00003-8

the frameworks are multilayer perceptrons. No doubt, adversarial networks act as a general-purpose solution to image-to-image translation issues. These systems not only acquire the information regarding mapping from an intake picture to yield picture but also get a loss activity to prepare for this mapping. The probability distribution can be duplicated by GAN so that they could, therefore, utilize loss activity, which depicts the distance among the dissemination of the information produced through the GAN and the dispersion of the original information.

GANs are a way to deal with generative modeling by utilizing DL strategies, for example, CNN. An improved method called deep convolutional GAN, or DCGAN prompted increasingly stable models. These days, most of the GANs are at least loosely dependent on the DCGAN design. It is also one of the variants of GAN. Generative modeling is an unsupervised task in artificial intelligence which contains automatically searching and learning the patterns or regularities as intake information. The framework is utilized to produce new structures that possibly can be taken from the initial dataset [3]. GANs design for naturally preparing a generative framework by considering the independent issue as supervised and utilizing both generative and discriminative structures. Fig. 3.1 shows the design of the generative adversarial network. (See Table 3.1.)

It is a deep-learning system and one of the most encouraging techniques for independent learning in complex dissemination. Deep-learning techniques can be utilized as generative structures. Two mainstream models incorporate the restricted Boltzmann machine (RBM) and the deep belief network (DBN). Two present-day instances of deep-learning generative framework algorithms incorporate the variational autoencoder (VAE) and the GAN [3]. GANs are a special case of generative models which are able to predict features in a much better way due to the adversarial training.

They are a smart way of preparing a generative framework by presenting the issue as a supervised learning issue with two submodels. The architecture of GAN goes through two components in the system: generative and discriminative models. Both of these models are prepared altogether by an adversarial procedure. Each model can be any neural network, such as a convolutional neural network (CNN), a recurrent neural network (RNN), or a long short-term memory (LSTM).

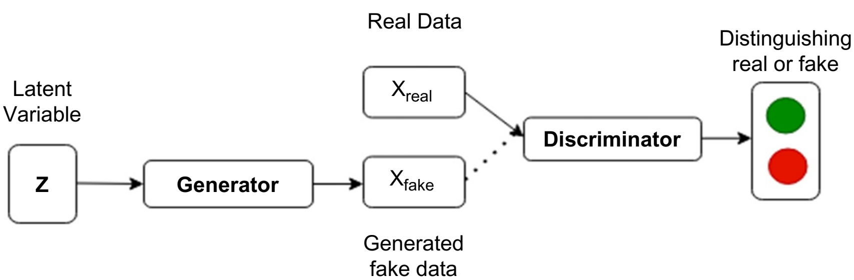

Fig. 3.1 GAN [4].

Table 3.1 Comparison between various GAN-based approaches.

Methods	Input	Output	Characteristics	Loss function	Resolution	Code
SRGAN [18]	Image	Image	High upscaling factor	Adversarial + feature	Arbitrary	TF
FCGAN [52]	Face	Face		Adversarial + distance	128 × 128	
VGAN [47]	Noise vector	Video		Adversarial	64 × 64	T
TGAN [43]	Noise vector	Video	Temporal generator	Adversarial	64 × 64	Ch
VariGAN [41]	Human + view	Human	Coarse to fine	Adversarial + distance	128 × 128	
StackGAN [71]	Text	Image	High-quality	Adversarial	256 × 256	
cycleGAN [19]	Image	Image	Unpaired data	Adversarial + cycle consistency	256 × 256	T + PT + TF
pix2pix [54]	Image	Image	General framework	Adversarial + distance	256 × 256	T + PT
Age-cGAN [64]	Face + age	Face	Identity preserved	Identity preserving		
Context Encoder [40]	Image + holes	Image		Adversarial + distance	128 × 128	T + TF
TP-GAN [68]	Face	Face	Two pathway	Adversarial + distance + identity preserving + tv + symmetry	128 × 128	

In code column T, TF, Ch, PT denotes Torch, Tensorflow, Pytorch, Chainer, respectively.

The training process of the G and D network is called adversarial preparation. G and D structures are prepared together in an adversarial fashion to improve each other as well as adjust the parameters for G to minimize $\log(1 - D(G(z))$ and for D to reduce $\log D(X)$ [5] while competing for the two-player min-max game with value function $V(G;D)$.

$$\min_G \max_D V(G, D) = \min_G \max_D E_{x \sim pdata}[\log D(x)] + E_{z \sim pz}[\log(1 - D(G(z)))] \quad (3.1)$$

where $V(G,D)$is a binary cross-entropy function mainly utilized in a binary classification problem, $D(x)$ is a multilayer perceptron, $p_z(z)$ is the distribution of input noise variables, and $p_{data}(x)$ and $p_z(z)$ in Eq. (3.1) denotes the real data probability distribution defined in the data space X and the probability distribution of z defined on the latent space, respectively. Gmaps z from Z into the element of X, whereas D takes an intake x and distinguishes whether x is a real sample or a fake sample generated by G [6].

3.1.1 Generative model (GM)

A *generator* model mainly figures out how to make pictures that look genuine. At the time of training, the *generator* continuously turns out to be better while the creation of pictures seems genuine. It takes a static length arbitrary vector as intake and produces an example in the area, as shown in Fig. 3.2. The vector is drawn arbitrarily from gaussian dissemination and utilized to seed the generative procedure. After preparation, points in multidimensional vector space will compare with the issue space, framing a restricted portrayal of the information circulation. The vector space is known as a latent space that contained inactive factors. A latent variable is an arbitrary variable that is significant for an area, however not straightforwardly recognizable. They are often implied as compression or projection of information conveyance. On account of GANs, it applies context to points in a selected latent area to such an extent that new points drawn from that area can be given to the generator model as intake and used to produce new and distinctive yield models [3].

The main purpose of the generator is to deceive the discriminator and generates new conceivable models from the problem area through machine mostly in picture whereas the discriminator figures out the false data made by the generator and determines whether the picture is authentic or machine generated [7]. In the first GAN hypothesis, G and D are not required to be neural systems and only required to have the option to fit the comparable generation and discriminant capacities. However, deep neural systems are commonly utilized as G and D. Both can be a nonlinear mapping function, such as a multilayer perceptron.

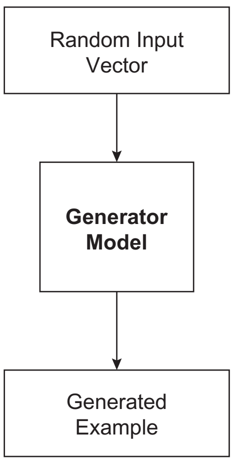

Fig. 3.2 GAN generator model.

3.1.2 Discriminator model (DM)

A *discriminator* model attempts to differentiate pictures as either authentic or false while training it turns into better at revealing that separation. It takes a model as input (genuine or created) from the space and presents a parallel class mark of genuine or produced, as shown in Fig. 3.3. The genuine model originates from the preparation dataset. The discriminator is a typical classification model. The discriminator model is disposed of as interest on the generator after the preparation procedure.

The procedure arrives at balance when the discriminator can no longer recognize genuine pictures from fakes. GANs possibility for both great and terrible is tremendous, in light of the fact that they can make sense of how to imitate any scattering of data. In this way, it can make universes amazingly such as own in any region: music, pictures, composition, and speech. They are robot specialists, and their output is imperative even solid. However, they can likewise be utilized to make false media content and is the innovation supporting Deepfakes. They have been utilized for some applications, particularly for picture blend on account of their capacity to produce high-quality pictures. In recent years, various variations of GAN have been projected, and they generated excellent outcomes for image generation. GANs relate to the arrangement of generative structures, which imply that they can create new content [8]. GAN does not work with a definite thickness function [9]. In game-theoretic methodology, it figures out how to produce from preparing dispersion through two-player game. Samples are best but can be precarious and unsteady to prepare with no inference queries. GANs depend on a hypothetical game situation wherein the generator system must challenge against an adversary, and legitimately creates tests. The discriminator network and adversary are the ones who

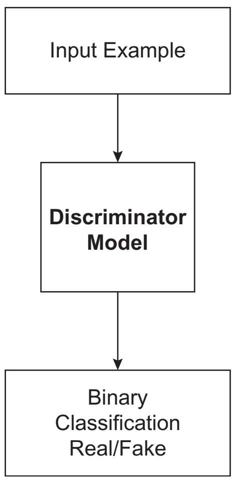

Fig. 3.3 GAN discriminator model.

try to recognize tests drawn from the preparation information and the generator [3]. The common game preparation produces a sensibly decent outcome. An amazing GAN application requires a reasonable preparation strategy; alternatively, the yield might be unsuitable because of the flexibility of the neural system model [7].

Advantages

1. It is a better modeling of data distribution.
2. In theory, GANs can train any type of generator network. Other frameworks require generator networks to have some specific form of functionality, such as the output layer being Gaussian.
3. There is no need to use the Markov chain to repeatedly sample, without inferring in the learning process, without complicated variational lower bounds, avoiding the difficulty of approximating the difficult probability of calculation.

Disadvantages

1. It is hard to train, unstable. Proper synchronization is required between the generator and the discriminator, but in actual training, it is easy for D to converge and G to diverge. D/G training requires careful design.
2. It has mode collapse issue. The learning process of GANs may have a missing pattern, the generator begins to degenerate, and the same sample points are always generated, and the learning cannot be continued.
3. It cannot solve inference queries such as p(x).

3.2 Related work

Goodfellow et al. [1] portrayed the GAN architecture in 2014 and discussed the nonsaturating loss function. It also provides the derivation for the optimal discriminator and demonstrates the effectiveness empirically on the MNIST, TFD, and CIFAR-10 image datasets. Radford et al. [10] introduced a class of deep convolutional GANs (DCGANs) that imposes empirical constraints on the network architecture to solve the problem of potential instability during training. Salimans et al. [11] provided a set of tools to avoid instability and mode collapsing, which includes historical averaging, minibatch discrimination, one-sided label smoothing, feature matching, and virtual batch normalization. Che et al. [12] used regularization methods for the objective to avoid the problem of missing modes. Arjovsky et al. [13] suggested minimization of the Wasserstein-1 or Earth-Mover distance among generator and data distribution with theoretical reasoning. In a follow-up paper, Gulrajani et al. [14] projected an enhanced approach for training the discriminator—termed critic by Arjovsky et al. [13]—which behaves stably, even with deep ResNet architectures. GANs have mostly been investigated on pictures, showing significant success with tasks such as image generation [15–17], image superresolution [18], style transfer [19, 20], and many others.

3.3 Deep-learning methods

There has been a gigantic advancement in framework demonstration and perception after presenting the advanced models for deep learning (DL). DL techniques quickly developed and extended applications in different logical and engineering areas. Deep learning is a growing area of AI (ML) research. It includes various concealed layers of artificial neural systems. The methodology applies nonlinear alterations and structures the deliberations of a high level in the huge collection of data. The current improvements in deep-learning structures inside various fields have just given huge commitments in AI. Current analysis has applied deep learning as the principal tool for digital image processing. A convolutional neural networks (CNN) is used for Iris recognition considered as more powerful in comparison with customary Iris sensor [21]. Deep learning is a subset of the field of ML, which is a subfield of AI [22]. Health informatics, bioinformatics, safety, energy, economic, security, urban informatics, hydrological systems modeling, and computational mechanisms are the advanced application field of DL [23]. Deep-learning techniques are quickly advancing for better performance.

Recently, DL algorithms have come out of AI and soft computing strategies. From that point, a few DL algorithms are currently acquainted with mainstream researchers and used in different application areas. Nowadays, their use has evolved into fundamental because of its knowledge, effective learning, precision, and strength in the model structure. Deep-learning strategies are quickly developing. Some of them have progressed to be had practical experience in a specific application area. Literature incorporates sufficient survey papers on the advancing designs in particular usage areas, such as superresolution imaging, multimedia analytics, cardiovascular image analysis, transportation systems, radiology, medical ultrasound analysis, 3D sensed data classification, activity recognition in radar, sentiment classification, renewable energy forecasting, image cytometry, 3D sensed data classification, text detection, apache-spark, and hyperspectral [24–29]. The convolutional neural network, recurrent neural network, de-noising autoencoder, deep belief network, and long short-term memory techniques have been recognized as the most famous deep-learning strategies [23].

3.3.1 Convolutional neural network

It is one of the well-known structures of deep-learning procedures. It includes three sorts of a layer with various pooling, convolutional, and completely associated layers shown in Fig. 3.4. There are two phases for the preparation procedure in each CNN, the feed-forward, and the back-propagation phase. GoogLeNet [30], AlexNet [31], ZFNet [32], ResNet [33], and VGGNet [34] are the most widely recognized CNN designs. In spite of the fact that it is basically known and commonly utilized for image processing applications.

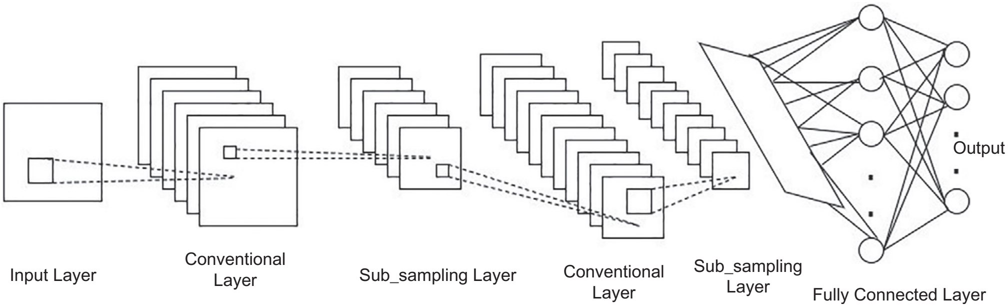

Fig. 3.4 CNN architecture [23].

3.3.2 Recurrent neural network (RNN)

It is moderately current deep-learning strategy. RNN is intended to perceive groupings and patterns, for example, handwriting, text, speech, and many more applications [23]. It has advantages in the structure of cyclic associations which utilize repetitive calculations to successively measure the intake information [35]. It is essentially an ideal neural system which has been extended beyond time through edges that feed into whenever step into rather than stepping into the following layer in a similar time. Every past source of input information is carried a state vector in concealed units, and further such vectors are used to process the yields. The expert systems, hydrological prediction, economics, energy, and navigation are its present applications. Fig. 3.5 depicts the architecture of RNN.

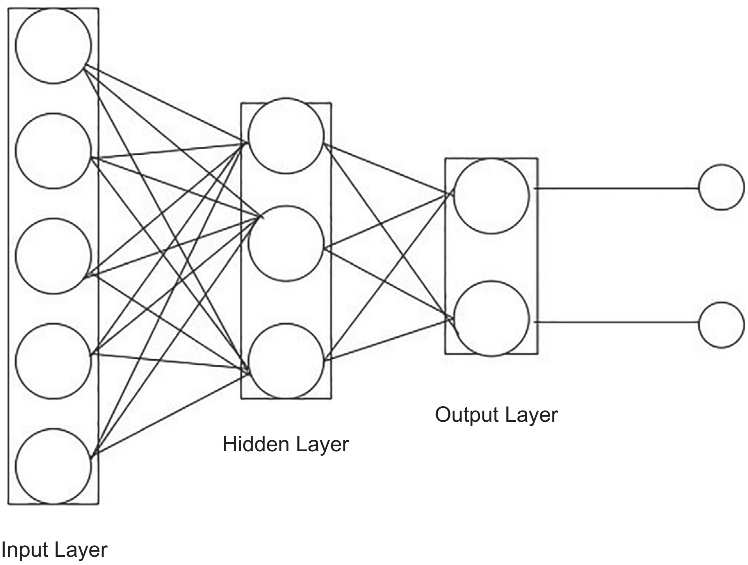

Fig. 3.5 RNN architecture [23].

3.3.3 Deep belief network (DBN)

It is recognized as a composite multilayered neural system which includes undirected and coordinated associations. It is utilized for high structural manifolds learning of information. The strategy consists of various layers which include associations among the layers with the exception of associations between units inside every fold. It also contains restricted Boltzmann machines (RBM) that are prepared in an insatiable way [36] in which each layer connects with both the past and resulting layer [37, 38]. The structure is comprised of a feed-forward system and a few folds of RBM as characteristic extractors [39]. The two layers of an RBM [40] are hidden and visible layers. Fig. 3.6 depicts the design of the DBN strategy.

Deep belief network is one of the most dependable DL strategies having computational proficiency and high precision [23]. Human emotion discovery, time arrangement expectation, sustainable power source forecast, cancer diagnosis, and financial estimating are the public application area.

3.3.4 Long short-term memory

It is an RNN technique that advantages input associations to be utilized as a general-purpose computer. The technique is used for two arrangements such as patterns recognition and image processing applications. Mainly, it consists of three central parts which

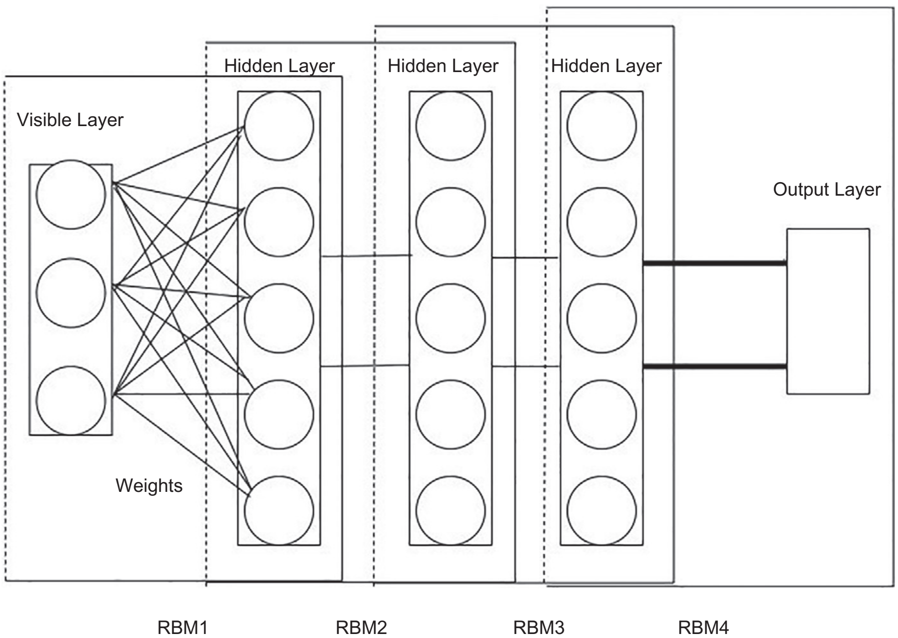

Fig. 3.6 DBN architecture [23].

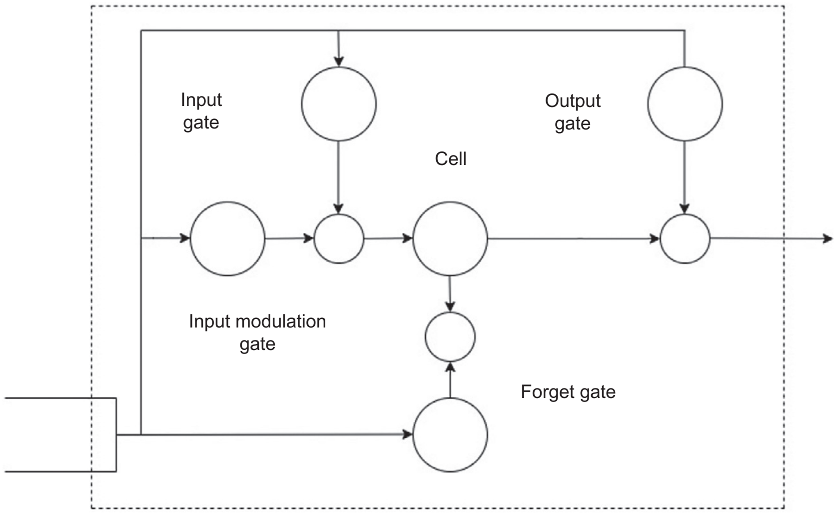

Fig. 3.7 LSTM architecture.

include information, yield, and forget doors that can be controlled on choosing when to allow the data to come inside the neuron and also to recollect what was figure out during the last time step. As it chooses, whole this that relies on the present intake is one of the fundamental qualities of the LSTM technique [23]. Fig. 3.7 presents the design of the LSTM technique.

LSTM has demonstrated incredible possibilities in various environmental areas such as hydrological prediction, hazard modeling, air quality, and geological modeling. LSTM design may be appropriate for some application areas because of its speculation capacities such as solar power modeling, energy demand and consumption, and wind energy industry [23].

3.4 Variants of GAN

With the advancement of technology, various improvements are made to the variants of GAN.

3.4.1 Vari GAN

Vari GAN represents variational GAN [41] which was proposed to create multiview individual pictures from a solitary perspective. This GAN replaces a coarse-to-fine manner. Vari GAN has been made out of three systems: a coarse image generator, a fine picture generator, and a restrictive discriminator. The coarse image generator G_C utilizes a restrictive VAE design [7] where VAE represents variational autoencoder. With an input picture i and an objective view v, a low-quality picture was created independently with the objective view i-v (low quality). The fine picture generator G_F is made out of double-way U-Net [42] design. The U-Net is named after its symmetric shape. This maps i-v (low quality) to a high-quality picture conditioned on the input picture. Discriminator D looks at the high-quality picture adapted on the input picture. G_F and discriminator are jointly prepared with a target function comprising a content loss and an affective loss estimating L1 distinction between (i-v) high quality picture and ground truth [10].

3.4.2 TGAN

TGAN represents a temporal generative adversarial net that was suggested by Saito et al. [43] for video generation. It comprises a discriminator, a temporal, and a picture generator. The temporal generator delivers a grouping of inactive frame vectors $[V_1^1, V_2^1, V_3^1, \ldots, V_S^1]$ from an arbitrary variable V_0, where S is the count of video frames. The picture generator takes V_0 and a frame vector Z_t^1 $(0 < t < S + 1)$ as intake and generates the t-th video frame. Here additionally, the discriminator accepts the entire video as intake and attempts to recognize it from genuine ones. TGAN follows WGAN [44] for stable preparation, however, applying further particular clipping value rather than weight clipping to the discriminator [10].

Recently, Temporal GAN (TGAN) [45] manages the instability in video generation by sending a frame-wise generation framework. A generative model is utilized to sample frames for image generation; a temporal generator preserves temporal consistency and controls this model. This model separates essential pieces of a video as a frontal area from background or dynamic from static patterns to manage the instability of preparing GANs. It accepts a latent space of pictures and considers that a video clip is produced by navigating the points in the dormant space. Video clips of various lengths relate to dormant space directions of multiple lengths.

3.4.3 Laplacian pyramid of generative adversarial network (LAPGAN)

Denton et al. [17] projected the creation of pictures in a coarse-to-fine manner utilizing a cascade of convolutional GANs having the structure of a Laplacian pyramid with N levels. This method utilizes multiple numbers of the generator and discriminator system and different levels of the Laplacian pyramid. A GAN is prepared by downsampling the picture at first at each phase of the level, N, and then it is again upscaled at each layer in a backward pass where a noise vector is mapped to a picture from the Conditional GAN with the coarsest quality until it reaches its original size. At each degree of the pyramid with the exception of the coarse stone, a different CGAN is prepared that considers the yield picture in the coarser level as a restrictive variable to produce the picture at this stage. This approach is mainly used because it can create pictures with higher quality in a coarse-to-fine manner [10]. This methodology permitted to exploitation the multiscale model of regular pictures, assembling a progression of generative models, each catching picture structure at a specific degree of the Laplacian pyramid which is made from a Gaussian pyramid utilizing upsampling $u(.)$ and downsampling $d(.)$ capacities. Assume $G(I) = [I_0; I_1; \ldots; I_K]$ be the Gaussian pyramid where $I_0 = I$ and I_K are k rehashed utilization of $d(.)$ to I. At that point, the coefficient h_k at level k of the Laplacian pyramid is given by the difference among the neighboring levels in the Gaussian pyramid, expanding the little one with $u(.)$.

$$h_k = L_k(I) = G_k(I) - u(G_{k+1}(I)) = I_k - u(I_{k+1}) \tag{3.2}$$

Laplacian pyramid coefficients $[h_1; \ldots; h_k]$ reconstruction can be performed by backward recurrence given as follows:

$$I_k = u(I_{k+1} + h_k) \tag{3.3}$$

So, a set of convolutional generative models G_0; G_1; … G_k, is used while preparing a LAPGAN where each of which captures the dispersion of coefficients h_k for various phases of the Laplacian pyramid. The generative structures are utilized to generate h_k's during reconstruction. Modification of Eq. (3.2) is given as follows:

$$I'_k = u((I'_{k+1}) + h'_k) = u(I'_{k+1}) + G'_k(z_{k,}\ u(I'_{k+1})) \tag{3.4}$$

Training image I is used for the construction of a Laplacian pyramid. A stochastic choice is made at each level regarding the coefficient h_k construction with the usage of standard procedure or produces by G_k [46].

D and G compete for the two-player minimax game with value function $V\ (G;D)$:

$$\min_G \max_D V(D, G) = E_{y, x \sim pdata(y, x)}[\log(d(y, x))] + E_{x \sim p_x, z \sim p_z(z)}[\log(1 - D(G(z, x), x))] \tag{3.5}$$

LAPGAN is a tandem system in which a set of pictures are adjusted orderly as per their quality from less to more. Based on a low-quality sample, it first produced a low-quality picture and then considered intake with a higher quality picture to the successive level. At each level, the generator corresponds to a discriminator that determines whether an intake picture is authentic or fake. The quality of the output image will be greatly improved and more authentic after many times of feature extraction. It is more advisable for high-quality pictures because it is trained under supervised learning.

The advantages of LAPGAN are easy to approach, learn residuals such as different distributions can be learned at each stage by the generator and passed as supplementary information to the next layer, step-by-step independent training, and increase the ability of GAN. In addition, it also joins CGAN to change unsupervised methodologies into supervised learning with significant performance advancement. The disadvantage is that it must be trained under supervision.

3.4.4 Video generative adversarial network (VGAN)

Video GAN (VGAN) framework proposes the utilization of independent streams for creating frontal area and background. Vondrick et al. [47] hypothesized that a video clip is a point in a latent space and suggested GANs generating video [9] with a spatiotemporal convolutional design in 2016. It adjusts the DCGAN model to predict future frames, create videos, and classify human actions. VGAN is a GAN for video in which it is considered that the entire video is joined by a stationary background scene and a dynamic foreground clip. The background is produced as a picture and afterward duplicated over time. A mutually prepared cover chooses among foreground and background to produce videos. So as to urge the system to utilize the background stream, sparsity is added earlier to the mask during learning. Henceforth, it considers a two-stream generator where the intake is a noise vector to both of them. The stationary background picture with 2D convolutional layers is produced with the effort of the background stream while the moving foreground generator attempts to create the 3D foreground video cube and the relating 3D forefront cover, with spatial-temporal 3D CNN layers predicts conceivable future frames. The discriminator considers the entire produced video as intake and attempts to recognize from the original video. Since VGAN considers video as a 3D cube that requires huge storage space and tests suggested this framework can likewise produce small videos up to a second at full frame rate better than basic baselines [21]. Also, investigations and perceptions describe the inside model that learns valuable highlights for perceiving activities with negligible oversight, recommending scene elements are a promising sign for portrayal learning.

A few attempts to approach the video generation issue were made through GANs [1]. However, past work has concentrated generally on small patches and assessed them for video grouping. This system is also to learn mapping from the dormant space to video clips. Yet, expecting a video clip is a point in the inert space that superfluously expands the intricacy of the issue, since videos of a similar activity with various execution speeds are presented by various focuses in the inactive space. In addition, this presumption forces each created video clip to have a similar length, while the length of real-world video clips varies. No doubt, GANs to video production is considered troublesome since the video has an additional temporal measurement involving a lot bigger calculation and storage cost. It is additionally not minor to keep temporal cognizance.

3.4.5 Superresolution GAN (SRGAN)

It takes a low-quality picture as intake and produces an upsampled picture with 4* upscaling quality. The main objective of SR is to enhance the quality of the low-tested picture that is upsampling the given picture. Basically, this issue is not well presented in light of the fact that the recovered high-quality picture misses high recurrence data during the upscaling of the image, particularly for huge upscaling factors. Numerous other deep-learning-based strategies [4, 45, 48] were projected to handle this issue, but those could not perform well with very low-tested pictures. This superresolution GAN utilizes deep-learning ideas to give higher quality pictures. During the process of training, a high-quality picture is always changed over into a low-quality picture by downsampling. The generator of the GAN is answerable for changing over the low-quality picture to high-quality picture, and the discriminator is liable for arranging the produced pictures [21].

Ledig et al. [18] projected SRGAN that considers a low-quality picture as input and produces an upsampled picture with 4* quality. The system design of SRGAN implemented by Ledig et al. supersedes the rules of DCGAN [36] architecture. The design of the generator utilizes both convolutional and residual networks [21]. The target function incorporates an adversarial and furthermore a feature loss rather than pixel-wise mean-squared error loss [9] to upgrade the authenticity of the renewed image and understand the 4* upscaling recreation. It also uses affective loss which is a component extricated by the convolutional neural system. By contrasting the highlights of the created picture and the attributes of the objective picture after convolutional neural system, the produced picture and the objective picture are increasingly the same in linguistics and pattern [49]. The feature loss is evaluated as separation among feature maps of the produced expanded picture and the factual picture, where the feature maps are removed from a preprepared VGG19 system by feeding the picture into it. Examinations depict that SRGAN has better execution at the best available methods on the collection of data for the public [21].

Loss is calculated as a weighted combination of regularization, adversarial, and content loss where function measures the difference in the two high-resolution images. SRGAN generator G takes low-quality image I_{LR} and outputs its high-quality image I_{SR}. θ_G are the parameters of G.

$$\hat{\theta}_G = \arg \min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} l^{\mathrm{SR}}\left(G_{\theta_G}\left(I_n^{\mathrm{LR}}\right), I_n^{\mathrm{HR}}\right) \tag{3.6}$$

SRGAN discriminator D classifies whether a high-quality image is I_{HR} or I_{SR}. θ_D is the parameter of D.

$$\min_{\theta_G} \max_{\theta_D} E_{I^{\mathrm{HR}} \sim ptrain(I^{\mathrm{HR}})}\left[\log D_{\theta_D}\left(I^{\mathrm{HR}}\right)\right] + E_{I^{\mathrm{LR}} \sim pG(I^{\mathrm{LR}})}\left[\log\left(1 - D_{\theta_D}\left(G_{\theta_G}\left(I^{LR}\right)\right)\right)\right] \tag{3.7}$$

Wang et al. [50] proposed an enhanced SRGAN which advanced the adversarial loss, the system design, and the affective loss.

3.4.6 Face conditional generative adversarial network (FCGAN)

FCGAN is face conditional GAN which focuses on facial picture SR. Berthelot et al. [51] projected BEGAN, which aims to try to maintain a balance that can be adapted for the trade-off among variety and trait. Huang projected FCGAN [52] that concentrates on facial picture SR. Inside the system design, both generator and the discriminator utilize a decoder, an encoder alongside skip associations. It creates excellent outcomes with 4* scaling factor. In preparation, the target function incorporates a loss, i.e., content loss, which is evaluated by the L1 pixel-wise dissimilarity between the produced upsampled picture and the ground truth.

3.5 Applications of GAN

The significant function of GAN is the systems that create cases with a similar dispersion as genuine information, for example, producing photo-realistic pictures. GANs can likewise be utilized to handle the issue of inadequate preparation of cases for supervised or semi-supervised learning. As of now, a favorable use of GAN is computer vision which includes pictures and video, for example, image-to-image translation, video generation, generation of cartoon characters, text-to-image translation, and many more. In this segment, the application scope of GANs is discussed [49]. GANs have some genuinely helpful practical applications, which incorporate the following.

A. The application in the image

- Image generation

  Generative systems can be utilized to create reasonable pictures after being prepared for sample pictures. For instance, to produce new pictures of dogs, a GAN can be prepared on thousands of samples of pictures of dogs. When the preparation has been completed, the generator system will have the option to create new pictures that are not quite the same as the pictures in the preparation set. Image generation is utilized in social media, marketing, entertainment, logo generation, and so on. Hanock et al. [53] projected composite GAN, which creates fractional pictures by various generators and lastly combined the whole picture.

- Image-to-image translation

  It is utilized to change over pictures taken in the day to pictures taken around evening time, to change over portrayals to artistic creations, to style pictures to look such as Picasso or Van Gogh works of art, to change over airborne pictures to satellite pictures consequently, and to change over pictures of ponies to pictures of zebras. These utilization cases are ground-breaking since they can spare time. Phillip et al. [54] exhibited GAN's, precisely pix2pix method for the image-to-image translation undertakings. Jun et al. [19] presented the renowned cycle GAN as well as the setup of noteworthy image-to-image translation models. Cycle GAN is a significant application framework of GAN in the field of a picture. It depends on two sorts of pictures that need no matching. A crying face can be transformed into a laugh through composition or zebra to the horse. Star GAN is a further advancement of Cycle GAN, where solidarity is taken to prepare a single classification for the next class. Star GAN is used to change the smiling look into a crying look, alongside a collection of appearances, for example, shock, disappointment, and so on.

- High-resolution picture generation

GANs can assist in creating high-quality pictures taken from low-quality camera pictures without losing any necessary details. Superresolution is a field in which GAN depicts a very remarkable outcome with commercial chances [55]. This can be valuable on websites. The utilization of GAN for SR tackles the inadequacies of the ordinary strategies, which includes the DL techniques, with absences of high recurrence data. Customary deep CNN can enhance the imperfection by choosing the target function. GAN can likewise take care of this issue and acquire fulfilling observation [49]. Christian et al. [18] show the utilization of GANs, explicitly SRGAN framework, to produce yield pictures having enriched pixel quality and sometimes even more. Huang et al. [56] utilize GAN to make variants of photos of personal appearances. Subeesh et al. [57] provide a case of GAN to make high-quality photos, concentrating on the road scene.

- Photo inpainting

The fundamental idea of this application is to fill the gaps of a picture. Numerous deep-learning procedures have come to tackle this issue, and the significant task is to fill the enormous gaps of a picture to make an ideal one. There are convolutional systems for picture inpainting however these are bad at filling the gaps with appropriate highlights, and henceforth generative models are utilized for searching the relevant highlights which are to be filled with, and these highlights are known through the preparation process [21]. Pathak et al. [50] have projected another technique for picture inpainting called context encoders which depend on convolutional systems prepared mostly to produce pictures at a discretionary. So these systems need to comprehend both full images and pictures with holes to recognize the highlights with which need to supplant with. The method proposed by Pathak et al. depends on encoder-decoder design. That framework is fit for taking pictures with input size 128×128 with gaps. The yield of that proposed framework is either the gap of the picture or the whole picture. The gap of the picture size will be 64×64, and the full picture is 128×128.

GANs can assist in recovering those areas in the picture that has some missing parts. Deepak et al. [40] portrayed the utilization of GAN, explicitly context encoder to execute photo inpainting that is covering a region of a photo which was expelled for unknown reasons. Raymond et al. [58] used GAN to fill in and fix purposefully corrupted photos of the human face. Yijun et al. [59] likewise used GAN for inpainting and remaking harmed photos of personal appearances [60].

- Generation of realistic photograph

  Andrew et al. [61] demonstrated the creation of synthetic photos with BigGAN strategy, which are in every practical sense undefined from authentic photographs.
- 3D object generation

  3D objects can be created with GAN [55]. Jiajun et al. [62] showed a GAN for producing new three-dimensional new items such as car, sofa, chair, and table. Matheus et al. [63] used GAN to produce 3D models that provide two-dimensional pictures of items from various points of view [60].
- Face aging

  The fundamental point of this is to create a human picture at some age. On the off chance, if the present age of an individual is 20 years, the GAN is utilized to create a picture of that individual at 40 years. Face aging techniques change a facial picture to another age, while as yet keeping character [21]. A large portion of the GAN utilized for face aging includes conditional GAN. The primary point is to produce a picture with an objective mark age from a given initial face picture. This can be extremely valuable for both the surveillance and entertainment businesses. It is especially helpful for face verification since it implies that an organization does not have to change its security frameworks as individuals get older. An Age-cGAN [64] system can create pictures at various ages, which then could be utilized to prepare a reliable model for face confirmation. Grigory et al. [64] utilized GAN to create photos of faces having various evident ages, such as from young to old one. Zhifei et al. [65] utilized a GAN-dependent strategy for de-aging the photos of different faces.
- Generate photos of the human face

  Tero et al. [66] exhibited the creation of conceivable, reasonable photos of individual faces. It is reasonable to call the striking outcome because of genuine looks. In that capacity, the consequences got a lot of media consideration. Face generation is usually prepared on examples such as big name, implying that components of current superstars are in the produced faces, causing to appear to be recognizable, however not precisely. Their techniques were likewise used to show the generation of items and scenes. Few instances were utilized from this paper in a 2018 report to exhibit the quick advancement of GANs from 2014 to 2017 [60].
- Generation of new human poses

  Liqian et al. [67] gave a case of creating current photos of individual structures with recent postures.
- Face frontal view generation

  Rui et al. [68] showed the utilization of GAN for creating front-view photos of individual faces provided photos taken at some particular point. The created fronton photographs can be utilized as intake is the concept behind it for face verification or face identification framework.

- Generation of cartoon character

Yanghua et al. [69] showed the preparation and usage of a GAN for creating anime characters' faces which are Japanese comic book characters. Motivated by the anime models, many individuals have attempted to develop Pokemon characters, for example, the poke GANventure and produce the Pokemon with DCGAN task having constrained achievement [22].

B. The Application with the Video

- Video synthesis

GANs can likewise be utilized to produce videos. They can create content in less time than if somehow managed to make content physically. They can also improve the efficiency of filmmakers and furthermore engage specialists who need to build innovative videos in their available time. Carl et al. [47] portray the utilization of GAN for video forecast, explicitly foreseeing as long as a moment of video frames with progress, principally for stationary components of the picture.

- Video frame prediction

It represents determining the future frame regarding the current frames [21]. Mathieu et al. [70] firstly used GAN preparation for video prediction in which the generator can produce the last frame of the video dependent on the prior arrangement of the frames, and the discriminator is utilized to finish up the frame. All the frames aside from the last frame are genuine pictures. The discriminator can adequately utilize the data of the time measurement and furthermore helps to make the produced frame predictable with all the past frames is its advantage. Test outcomes depict that the frames are clearer than the other algorithms created by confrontation preparation.

C. Application of human-computer interaction

- Text-to-image synthesis

It is the earlier application of domain-transfer GAN. No doubt, generating multiple pictures from text details is an intriguing use case of GANs. This can be useful in the film business, as a GAN is equipped for creating new information relied on some content that can be made up. In the comic industry, it is conceivable to naturally create arrangements of a story. Han et al. [71] exhibited the utilization of GAN, explicitly the Stack GAN to create practical appearing photos from textual portrayals of necessary items such as flying creatures and blossoms.

- Auxiliary automatic driving

Santana et al. [72] actualized the assisted automatic driving with GAN. Initially, a picture is created, which is reliable with the appropriation of the official movement scene picture, and afterward, a progress framework is prepared dependent on the cyclic neural system to anticipate the following movement pictures.

3.6 Conclusion

Nowadays, GANs are one of the most fascinating thoughts for many researchers to work on it and suggesting various models based on GAN with regard to computer engineering. Generative adversarial networks and their variants are the most promising generative approaches in the discipline of computer vision. In this chapter, a comprehensive review of GAN and their variants are provided. It can be seen that the latest variants of GAN are unsupervised and more stable than the previous models that can produce realistic content and texture details, which will be an advantage to various applications such as superresolution, image inpainting, etc. They are also applicable in different areas such as image classification, image-to-image translation, recovery of corrupted data, text-to-image generation, and many more endless applications. Comparison is also done between various GAN-based methods.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems, Curran Associates, 2014, pp. 2672–2680.

[2] A. Mosavi, S. Ardabili, A.R. Várkonyi-Kóczy, List of Deep Learning Models. (2019), https://doi.org/10.20944/preprints201908.0152.v1 (Preprint).

[3] J. Brownlee, A Gentle Introduction to Generative Adversarial Networks (GANs), Retrieved from https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/, 2019, July 19.

[4] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution. in: Computer Vision—ECCV 2014, 2014, pp. 184–199, https://doi.org/10.1007/978-3-319-10593-2_13.

[5] M. Mirza, S. Osindero, Conditional Generative Adversarial Nets, (2014)ArXiv: abs/1411.1784.

[6] Y. Hong, U. Hwang, J. Yoo, S. Yoon, How generative adversarial networks and their variants work. ACM Comput. Surv. 52 (1) (2019) 1–43, https://doi.org/10.1145/3301282.

[7] K. Sohn, X.C. Yan, H. Lee, Learning structured output representation using deep conditional generative models, in: Proceedings of the 28th International Conference on Neural Information Processing Systems—vol. 2 (NIPS'15), MIT Press, Cambridge, MA, USA, 2015, pp. 3483–3491.

[8] J. Le, The 10 Deep Learning Methods AI Practitioners Need to Apply, Retrieved fromhttps://medium.com/cracking-the-data-science-interview/the-10-deep-learning-methods-ai-practitioners-need-to-apply-885259f402c1, 2020, May 10.

[9] Y. Hong, U. Hwang, J. Yoo, S. Yoon, How generative adversarial networks and their variants work: an overview, ACM Comput. Surv. 52 (2019) 1–43.

[10] W. Sun, B. Zheng, W. Qian, Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis. Comput. Biol. Med. 89 (2017) 530–539, https://doi.org/10.1016/j.compbiomed.2017.04.006.

[11] T. Salimans, I.J. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved Techniques for Training GANs, (2016)ArXiv: abs/1606.03498.

[12] T. Che, Y. Li, A.P. Jacob, Y. Bengio, W. Li, Mode Regularized Generative Adversarial Networks, (2016)ArXiv: abs/1612.02136.

[13] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein Gan, (2017)arXiv preprint arXiv: 1701.07875.

[14] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved Training of Wasserstein GANs, (2017)ArXiv: abs/1704.00028.

[15] T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive Growing of GANs for Improved Quality, Stability, and Variation, (2018)ArXiv: abs/1710.10196.