



Generic image application using GANs (Generative Adversarial Networks): A Review

S. P. Porkodi¹ · V. Sarada¹ · Vivek Maik¹ · K. Gurushankar²

Received: 6 April 2022 / Accepted: 5 September 2022 / Published online: 30 September 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

The generative adversarial network (GAN), which has received considerable notice for its outstanding data generating abilities, is one of the most intriguing fields of artificial intelligence study. Large volumes of data are required to develop generalizable deep learning models. GANs are a highly strong class of networks capable of producing believable new pictures from unlabeled source prints and labeled medical imaging data is scarce and costly to produce. Despite GAN's remarkable outcomes, steady training remains a challenge. The goal of this study is to perform a complete evaluation of the GAN-related literature and to present a succinct summary of existing knowledge on GAN, including the theory following it, its intended purpose, potential base model alterations, and latest breakthroughs in the area. This article will aid you in gaining a comprehensive grasp of GAN and provide an overview of GAN and its many model types, as well as common implementations, measurement parameter suggestions, and GAN applications in image processing. It will also go over the several applications of GANs in image processing, as well as their benefits and limitations, as well as its prospective reach.

Keywords Artificial intelligence · Generative adversarial networks · Image processing · Neural network · Supervised learning · Semi-supervised learning

1 Introduction

The capacity of computer systems to behave, think, and make choices like humans has been one of the most significant and remarkable advances in the area of computer science and this is said to be machine learning technology. Various algorithms have been created over time to create machines and computer systems that can mimic human brains, and a variety of programming languages have been used to implement these algorithms.

Many advances in the sphere of machine learning, particularly deep learning, were discovered when more computing or processing capacity becomes accessible. deep learning makes it easier to extract relevant, abstract, and high-level features from input data for usage as classifiers and detectors. This methodology is often known as learning by representation and it is interpreted from how human minds think and work. The principle of a generative model (deep learning-based models) is the focus of generative adversarial networks (or GANs). The topic of image synthesis has gotten a lot of attention. It is a phrase for the process of creating an image using the image's veiled and exposed characteristics. GANs are commonly applied to the field of imaging algorithms in general due to their shown ability to function effectively with images. GANs are made up of two models that are trained against each other at the same time. Historically, Markov chains and highest probability estimation were used to construct models of GAN like the Restricted Boltzmann system (Fischer and Igel 2012) and the variational auto encoder (Kingma and Welling 2013). They are modeled based on the distribution of input data which leads to the estimation of the generated data, but their output and results suffer because of their low generalization

✉ V. Sarada
saradav@srmist.edu.in

S. P. Porkodi
ps4195@srmist.edu.in

Vivek Maik
vivekma@srmist.edu.in

K. Gurushankar
gurushankar01051987@gmail.com

¹ Department of ECE, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 603203, India

² Higher Medical and Biological School, South Ural State University, Chelyabinsk, Russian Federation

capacity. To recover the problem in 2014, Goodfellow et al. (2014) proposed GAN, a new theory in the field of generative models. It is made up of a generator and a discriminator network, which are rivals who are always trying to outperform one another while improving themselves. GAN was created to help people understand joint probability distributions.

The Generator's job is to generate new data points depending on the distribution of existing input sampled data points, with the deception that the generated sample points are correct. The task of the Discriminator is to call the Generator's bluff by detecting the sampled data as artificially produced or obtained from real data. It is the equivalent of two rivals playing a zero-sum game. Back-propagation (Rumelhart et al. 1986) is used to train models, and dropouts are deleted (to avoid overfitting). The core idea of GAN is a derivative from a two-person game of zero-sum in which a person's gain or loss is perfectly matched by the gain or loss of the other person. GANs are similar in this aspect that the generator and a discriminator both learn at the same time. The generator creates fresh data samples when attempting to capture the likely distribution of actual samples. The discriminator is typically a binary classifier, which accurately separates individual samples from manufactured samples. In addition, the generator and discriminator will be built using conventional deep neural network architecture (Goodfellow et al. 2016; Radford et al. 2015). The best strategy for GANs is to play a minimax game to reach Nash equilibrium (Ratliff et al. 2013), where the generator has optimally captured the sampling distribution of real data. Historical prospects of GAN-based image processing are discussed in this article. Section 2: GAN Overview. The numerous types of GAN models are discussed in Sect. 3. Section 4 goes over some of the most prevalent GAN applications in image processing, and Sect. 5 goes over some of the more advanced GAN applications. The merits and downsides of GANs are discussed in Sect. 6. Section 7 gives GANs Limitation. A conclusion and possible scope remarks are included in Sect. 8. Figure 1 depicts the whole survey analysis.

2 Review of GAN concepts and categories

GAN is a sort of architecture that stages two neural layers as adversaries to each other to produce new synthetic sample data, that very much depict real sampling data and runs high probability to be taken as real inputs. Repeatedly used in the formation of images, videos, and speech. GANs are particularly well suited to image processing because of their excellent performance rate in picture tasks. They're regarded to be the most efficient image generating procedure, and they're used in a wide range of applications (Kumar and Dhawan 2020; Pan et al. 2019). This section covers the fundamentals

of GAN architecture, goal functions, latent space, and GAN problems. The two-person least-max null-sum game is a crucial characteristic of GAN. In this game, one person receives compensation at the expense of the other competitor's loss. The game contestants represent the discriminator and generator networks. One of the basic goals of the discriminator is to detect whether a sample belongs to a true or false distribution (Goodfellow et al. 2014; Kumar and Dhawan 2020). Meanwhile, the generator is attempting to deceive the discriminator by creating an incorrect trial distribution. The discriminator assesses how likely or unlikely a particular sample is to be a genuine sample. The sample is more likely to be representative of the population if the likelihood value is higher. The sample is fraudulent if the value is close to zero. A probability value close to 0.5 indicates that the best solution is generated and depicts the lack of distinction between real and synthesized sample data.

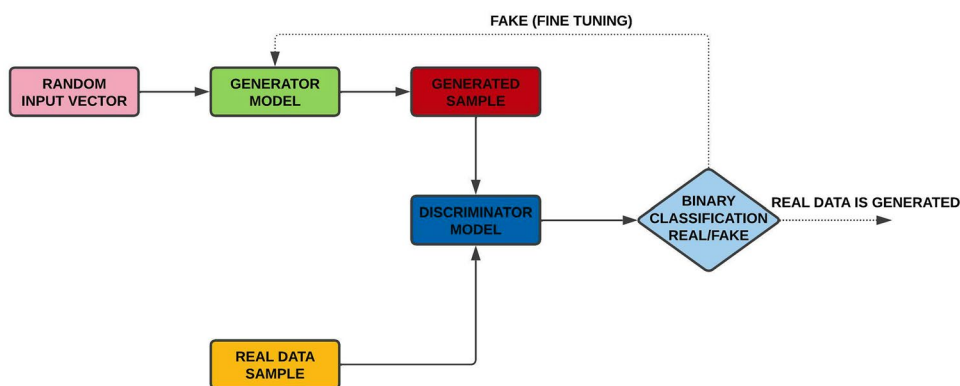
GAN's overall architecture is depicted in Fig. 2. As seen the dual network of generator and discriminator make up GAN. Over time of generator inception, its capacity to produce credible data increases rapidly. The produced instances are used as negative training examples by the discriminator and with time the discriminator becomes well adept at distinguishing between fictitious and genuine data from the generator. If the generator delivers improbable results, the discriminator penalizes it.

The use of random noise is much recommended to create graphics. Z is the symbol for random noise. The images created by the noise are saved in the G format (z). Gaussian noise, with its normal distribution, is the most common input. Both the networks in GAN need to be recursively adjusted in training and updated progressively. The made-up character of the discriminator can estimate the original distribution of any given image. For a given image X the $D(X)$ represents a unit probability for genuinity and a zero probability for a fake one. The generative modeling goal is to fit the $p_{data}(x)$ and $p_g(x)$ real data distributions. As a result, for training generative models, minimizing discrepancies between two distributions is critical (Goodfellow et al. 2014). JSD ($p_{data} || p_g$) calculated by the discriminator is reduced by regular GAN (Hong et al. 2019). Researchers recently discovered that, instead of JSD, different distances or divergence measurements can be used to increase the GAN's accuracy. In this part, we'll look at how to use different distances and objective functions to calculate the difference that exists among the real data distributions. Latent space, also known as embedding space, stores a compact representation of data. If we tried to modify or describe any features of a picture such as a posture, an age, an appearance, or an image's object all in the spatial domain, could be challenging because of the high dimensionality and distribution space (Lin et al. 2018). As such part taking in the latent space is a much more feasible option since the latent



Fig. 1 Outline of the survey. It consists of three important parts such as generative adversarial network, along with its different types of GAN models and also application of GAN

Fig. 2 General architecture of GAN



representation transmits basic properties of the input image compactly. This section examines how GAN expresses goal qualities in latent space and how the GAN system might benefit from a variational strategy. Even when trained on multi-model data, GANs have the drawback of producing homogeneous samples. When GANs are trained on data of handwritten ten-mode digits, for example, G may be incapable of producing any digits (Goodfellow 2016). This is referred to as the mode collapse problem, and much literature has been proposed on overcoming this problem. Additionally, rather than a fixed-point convergence, G and D can oscillate during training. When one player becomes more effective than another, the system can become unstable due to vanishing gradients. D rapidly develops the ability to differentiate between genuine and fabricated samples, although the created samples are initially of poor quality. As a result of this, the productive sample probability will be closer to zero, resulting in a very small $\log(1-D(G(z)))$ gradient (Zhu et al. 2017). This demonstrates that G will not update if D lacks gradients. Additionally, it is critical to choose hyperparameters including momentum, batch size, and learning rate carefully to ensure that GANs training converges.

3 Different types of GAN models

Several modifications to the original GAN model have been developed since its introduction, resulting in a variety of GAN models (Durgadevi et al. 2021). These variants include changes made for a specific application or a problematic assertion, such as style conversion from one image dataset to another image dataset (Zhu et al. 2017), image improvement (Ledig et al. 2017), complete and incomplete face images (Chen et al. 2018; Li et al. 2017), and producing an image from text (Reed et al. 2016). In 2014, Goodfellow et al. (2014) introduced it as a baseline notion for GANs for the first time. This is the most basic sort of GAN. The Generator and Discriminator in this case are basic multi-layer perceptron's. The vanilla GAN (Jiang et al. 2021) was designed

to work by deriving samples from a specified data distribution without explicitly modeling the underlying probability density function. The early GAN designs employed vanilla GAN. The proposed Vanilla GAN, generator and discriminator were tested using the Toronto Face Dataset (TFD), MNIST handwritten digit dataset, and CIFAR-10 natural image dataset. The Jensen-Shannon divergence, or measure of comparison between two distributions, might become constant because the probability distributions of real and fraudulent data overlap so little, leading to the vanishing gradient problem. It does not function effectively when dealing with more complex problems. Radford et al. (2015) proposed a GAN based on convolutional layers (DCGANs) (Dewi et al. 2021; Cheng et al. 2021) with certain special case assumptions including:

- All hidden layers that are fully connected are deleted.
- Fractional strided convolutions are employed in the generator instead of pooling layers. Pooling layers are replaced with convolutional strides in the discriminator.
- Both the generator and the discriminator use batch normalization.
- The ReLU (rectified linear unit) activation function is used in the pre-output layer of a generative model, and on the rest of the layers, leaky ReLU activation is deployed.
- The performance of the DCGAN models was assessed using the LSUN, SVHN, CIFAR10, and Imagenet1k datasets. By initially employing DCGAN as a feature extractor and then fitting a linear model on top of those features, the effectiveness of unsupervised representation learning was assessed.

Arjovsky et al. (2017) in their work stated that WGAN could easily overcome the vanishing gradient problem. The discriminator must be trained for longer iterations than the generator in order to train a WGAN model. In other words, you need to train the discriminator across a number of iterations for each generating iteration. For 10,000 generator iterations, train with a mini batch size

of 64. The use of Earth-Mover distance as a substitute to the Jensen-Shannon divergence for probability distribution comparison to those of fake or induced data. The critic function f is used to define the discriminator. It's based on the Lipschitz constraint. Although WGAN (Wang et al. 2019) is useful for reliably training GANs, it still creates low-quality data samples and fails to converge on occasion. An enhanced version of WGAN was proposed by Gulrajani et al. (2017). A GAN which is based on clipping of weights can be used to confine the critic with Lipschitz restrictions, causing training failure, they discovered. They obtained their results and concluded that unusual behavior develops. Rather than clipping weights, the Lipschitz constraint was introduced, which imposes the norm penalty on a gradient of the critic function (f) in respect to distribution. In comparison to the WGAN weight clipping variation, this technique converges faster and with less distinction between real and generated samples. The sigmoid cross-entropy loss function is employed during the backpropagation step in classic GANs. The vanishing gradient problem will nonetheless occur during the learning process as a result of this loss function. As a result of this challenge, Mao et al. (2017) created the LSGAN technique. The loss feature used is the least square loss feature. Minimizing the Pearson divergence results from minimising the goal function of LSGAN. During the learning process, LS-GANs (Wang et al. 2021) produce higher-quality images and are more trustworthy than traditional GANs. In a fundamental GAN, the training requires a single class label to identify the real or generated data source. Odena et al. have (Odena 2016) proposed the use of the class label (say N) for actual data and conditioning discriminator D as a Semi-GAN (SGAN). In SGAN, training a classifier model with a labeled and unlabeled input. In Keras, there is at least three ways to implement the supervised and unsupervised discriminator models utilised in the semi-supervised GAN. D will divide data into one of the $N+1$ classes after training and G will use the additional class to identify the source of data. Due to its trained and classified characteristics, this technique produces a more accurate classification that yields high-quality samples than ordinary GAN. The generator in a traditional GAN is merely provided latent space. The conditional GAN modifies this, as described by Mirza and Osindero (2014), by adding an extra parameter (label y) to the generator in addition to latent space and training it to generate related images. The discriminator is given true images and labels as input to better discern genuine images. This model is shown to generate digits that are identical to those in the MNIST dataset when given class labels (0, 1, 2, 3...9). It's known as a Conditional GAN as a result of this. The generator analyses by translating latent function vectors to

actual data probability distributions in traditional GANs. It does not, however, have a good way of mapping actual data to latent data. Bidirectional GANs were proposed by Donahue et al. (2016) (BiGANs). By mapping the real data probability distribution to latent space, it aids in learning how to delete relevant characteristics. The goal is to create a GAN capable of learning rich representations for us in applications such as unsupervised learning. According to Denton et al. (2015) the images are produced in a coarse to fine fashion using a framework of the Laplacian pyramid and a convolutional network cascade. They were able to leverage the multipurpose structure of authentic images by developing a sequence of generative models that captured the visual structure of the Laplacian pyramid at a different level. A Laplacian pyramid is identical to a Gaussian pyramid in appearance, but it saves the disparity images of the obscured versions between each stage. Makhzani et al. (2015) developed an adversarial autoencoder-based GAN which can conduct variational inference on the autoencoder's hidden code vector by comparing its aggregated posterior with a prior distribution. The AAE is trained using two criteria in adversarial autoencoding: a reconstruction error using the conventional objective and aggregate posterior-based adversary training. Once the training is complete the encoder moves the data distribution to the prior distribution, while the decoder uses a deep generative model to learn how to map before the distribution. According to Im et al. (2016), illustrated in this recurrent generative model, unfolding the optimization using gradient produced periodic computations that generated visuals by progressively appending on t to a visual canvas. In this scenario, the convolutional network's "encoder" extracts representations of the current "canvas". The codes which get generated, along with the reference image's code, are given into a "decoder," which determines whether the "canvas" should be modified or not. GANs with an information-theoretic extension described by Chen et al. (2016) can acquire disentangled functionality in an unsupervised manner. Because they directly reflect the significant aspects of a data instance, disentangled representations are effective for tasks like facial identification and object recognition. InfoGANs (Ye 2022) purpose is to maximize the mutual information between small fixed selections of GAN's noisy observation variables, which differs from its goal of learning meaningful representations. A disentangled representation directly displays the prominent aspects of a data item which can be beneficial for tasks like face and object identification. In this case, InfoGANs change GAN's goal of learning meaningful representations by maximising the mutual information between a fixed small selection of GAN's noise parameters and observation. Table 1 shows a comparative

Table 1 Comparison of various types of GAN based on different objective and performance metrics (Hitawala 2018)

Criteria	GAN	DCGAN	WGAN	WGANGP	LSGAN	SGAN	CGAN	BIGAN	LAPGAN	AAE	GRAN	INFOGAN
Learning	Supervised	Unsuper-vised	Unsuper-vised	Unsuper-vised	Supervised and semi-supervised	Semi-super-vised	Supervised	Supervised and unsuper-vised	Supervised	Supervised, semi-supervised and unsupervised	Supervised	Unsupervised
Architecture	Multilayer perceptron	CN with constraints	WGAN critic or the GAN discriminator	Lipschitz constraint	Probably approximate correct (PAC)-style theorem	Constraint based and CNN	Multilayer perceptron	Deep multilayer neural networks	Laplacian pyramid of CN	Autoencoders	Recurrent CN with constraints	Multilayer perceptron
Objective methodology	G's value function is reduced, whereas D's is increased	Learning the hierarchy of representations from sections of objects to scenes in both G and D	Computing the difference between the real and created distributions	Use a critic function on the input instead of clipping weights.	There was no discernible deterioration or mode collapse of the output images when the batch normalizing layers were removed	Enhance the quality of the generated samples and shorten the generator's training time	Conditional on more information, decrease value function for G and increase value function for D	Learn characteristics for similar semantic tasks and apply them in uncontrolled situations	The fashion image is created on a coarse-to-fine scale	Inferencing the prior distribution and the posterior of the autoencoders hidden code vector	Gradual adjustments to a "canvas" are used to produce images	To learn disentangled representations, maximize mutual information
Gradient updates	1 step for G and k steps for D in SGD	SGD with Adam optimizer for both G and D	Earth Mover distance	Critic function	cone shaped with vanishing gradient	base line classifier with no generative component	1 step for G and k steps for D in SGD	--	--	SGD with reconstruction and regularization step	Both G and D are SGD upgrades	G and D are both SGD upgrades.
Performance comparison metrics	Log-likelihood	Accuracy and error rate	Improve the stability of learning	Improvement to the WGAN	To maximize the image generation performance	Log-likelihood	Log-likelihood	Accuracy	Human estimation and log-likelihood	Log-likelihood and error-rate	Generative adversarial metric (proposed)	Information metric and representation learning

examination of several forms of GANs utilizing various criteria (Hitawala 2018).

4 Application of GAN

Since GAN is capable of generating realistic samples from a given input latent space, it can be considered an extremely efficient and useful generative model. We are not required to know the exact distribution of real-world data or to draw any additional statistical inferences (Alqahtani et al. 2021). These advantages have resulted in the widespread use of GAN in several academic and technological fields (You et al. 2022). We take a look at a few computer vision applications that have been published and refined in the literature. These examples were chosen to demonstrate several methods for manipulating, interpreting, and characterizing images using GAN-based representations, and do not reflect the full range of GAN applications. This section discusses in-depth the applications of GANs (Aggarwal et al. 2021) in image processing.

4.1 Image generation with enhanced quality

The majority of current GAN research has been devoted to improving the quality and utility of picture creation skills. In a course to fine way, the LAPGAN model was extended with a CNN cascade to generate images within a Laplacian pyramid structure (Donahue et al. 2016). Zhang et al. (2019) developed the self-attention based GAN (SAGAN) for image generation problems, which enables long-range dependency modeling through attention. In contrast to standard convolutional GANs, which create high-resolution information from just locally distributed points in a lower resolution feature map. SAGAN, on the other hand, is fascinated by the information that may be gleaned from a mixture of stimuli from all feature placements. On the difficult ImageNet dataset, the SAGAN was able to show state-of-the-art performance, increasing beyond the highest inception score from 36.8 to 52.52 and shrinking the Frechet Inception difference from 27.62 to 18.65. Huang et al. (2017) instead of using lower resolution images, GANs use intermediate representations. This technique has been proven to be effective, and it is currently a commonly used method for boosting image quality. By giving additional mark information as input to both G and D networks, LAPGAN has expanded the conditional version of the GAN model; this method has shown to be beneficial and is now a regular practice for increasing image quality. The GAN conditioning technique was later expanded to encompass natural language.

As demonstrated by Nguyen et al. (2016), a gradient increase in the generator networks latent space enhances

multi-neuron activation in a distinct classifier exciting technique to synthesize fresh images. This approach was further developed in Nguyen et al. (2017) by incorporating a latent code, which enhanced sample consistency, precision, and variety, resulting in a new generative model that creates images with a resolution of 227×227 , which is superior to prior generative models. This is true for each and every one of the 1000 ImageNet forms.

For generative adversarial networks, Salimans et al. (2016) provided a set of innovative structural properties and planning strategies (GANs). The emphasis of the authors is on two GAN applications: semi-supervised learning and the creation of visually realistic images. They didn't want to create a model that assigned a maximum likelihood, and they didn't want it to learn without labels. On MNIST, CIFAR-10, and SVHN (street view house numbers), the authors applied unique methodologies to get state-of-the-art semi-supervised classification results. The exceptional quality of the images produced was confirmed by a visual turing test. The suggested model generated the MNIST dataset, which no one can distinguish apart from real data, as well as CIFAR-10 samples with a human error rate of 21.3 percent.

4.2 Image super resolution

The term “super resolution” refers to a variety of upscaling techniques for video and images. The trained model contains image real data while sampling, which leads to the creation of a high resolution image from a lower resolution image (Wang et al. 2019). Wang et al. (2018) found that the visual efficiency of SRGAN is increased by combining three major SRGAN aspects - structural network design, antagonistic and perceptual loss—to create an enhanced SRGAN (ESR-GAN). The residual dense block (RRDB) was the primary unit used to create networks without batch normalization. They also adjusted the relativistic GAN principle such that the discriminator can predict relative realness instead of absolute value. In the end, perception loss has been exacerbated by activating functionality before texture recovery and brightness consistency, recommending a better restructuring of texture and consistency monitoring. The suggested ESRGAN achieves consistent visual consistency with more practical and realistic textures than SRGAN and has won first place in the PIRM 2018-SR Challenge with the highest perceptual index (region 3).

Karras et al. (2017) proposed a new approach for generative adversarial networks has been made. The key idea behind this study is to gradually improve the precision of both the generator and discriminator networks: we start with a low resolution and gradually add more layers that model finer and finer information as training progresses. This

speeds up and stabilizes the planning process, allowing us to create image graphs of exceptional quality.

4.3 Image inpainting

Visual inpainting is a reorganizing strategy for missing image data sections to prevent observers from identifying that they have been restored. It is often used to eliminate undesirable artifacts from images or to restore the degraded areas of historical or artifact pictures. Edge Connect suggested by Nazeri et al. (2019) is a 2-stage adversary paradigm that includes the network of image completion and edge generators. The edge generator prepares edge hallucination (normal as well as irregular), and the image completion network uses these hallucinated edges as a priority to fill missing regions. We test our model from the beginning to the end using publically accessible data sets like CelebA, Places2, and Paris Street View. Yu et al. (2018) developed a deep model-based generative method, which not only synthesizes single image/image structures but also uses image attributes around it to improve predictions as a reference during training of networks. During the experiment, the approach is a CNN feed (convolutionary neural network) that can handle the images in random and variable-sized places with many holes. Yeh et al. proposed a new approach to paint semantic images (Yeh et al. 2017). The researchers have viewed semantic painting as a limited picture creation problem with existing generative modeling developments. In this situation, an opponent network (Goodfellow et al. 2014; Radford et al. 2015) has developed a deep generative model and is now trying to encode the corrupted picture which is 'closest' to the image in the latent space. The signal is then reproduced with the encoding by the generator. A weighted background loss is used to make the corrupted image conditional, while an earlier loss is used to penalize illogical images.

4.4 Object recognition

Object detection is a method of detecting actual objects such as faces, bikes, and buildings in pictures or films. Object identification algorithms commonly employ extracted features and learning techniques to identify individual object-type instances. All driver aid systems (ADAS) use image recovery, security, monitoring, and sophisticated driver assistance. It is typically difficult to detect small things because of their low resolution and brilliant representation. Li et al. (2017) have been developing a modern Perceptual Generative Adversarial Network (Perceptual GAN) to improve small object recognition, minimizing the representational gap between small and large things. Its generator learns to deceive a competitor through perceived weak little

object representations that are close enough to true enormous items. In the meantime, the discriminator competes with the generator to grade the created representation and imposes a visual criterion on the generator which is important for the detection of representations of tiny objects.

4.5 Generation and prediction of video

Computer vision is a big issue in understanding object motions and scenic dynamics. A model of how scenes convert is included both video recognition (e.g., classification of action) and video generation (e.g. future prediction). On the other hand, the construction of a dynamic model is difficult because of the great range of shapes that objects and surroundings could take. Mathieu et al. (2015) employed a convolutive network trained on an input sequence to construct likely frameworks. To address the internal biases of the standard Multi-Scale Features (MSF), three separate and complementary techniques of feature-learning were developed: a multidisciplinary structural design, an adverse training approach, and a differential image gradient feature. To overcome the conventional MSF erroneous predictions are considered. They compare the predictions with many previously published results by using recurring neural networks and the UCF101 data set.

To distinguish scenes, Vondrick et al. (2016) employed a video network with a spatio temporarily coevolutionary structure. Experiments indicate that this method provides simple basic guidelines to make short films at a full-frame rate up to a second and to anticipate the future of static images. Experiments and views reveal that the model analyses significant components for comprehension of internal behavior at minimum power, and scene dynamics provide an attractive signal for the learning of representations. Tulyakov et al. (2018) design for video production was inspired by the Motion and Content of the Generative Adversarial Network (MoCoGAN). A random vector sequence is mapped in the recommended structure to make a video. The contents and the motion component are contained in every random vector. The motion part is stochastically implemented while the content part is constant. The authors have devised a fresh adverse way of learning which uses picture and video discriminators to uncontrollably learn movement and content breakdown. The usefulness of the proposed method has been demonstrated by early findings on a range of tough data sets as well as qualitative and quantitative comparisons to the state of the art approaches.

4.6 Generation of anime character

The costs for developing games and designing animations are costly since it requires many producer artists to carry out relatively repeated work. Automated Anime Characters

(Jin et al. 2017) are created and colored by GAN. The model consists generator and discrimination system of various layers, batch normalization, ReLU, and avoidable connections.

Chen et al. (2018) developed an approach that would be useful in computer vision and graphics, that would turn real-world image graphs into cartoon visuals. The proposed approach is CartoonGAN, a generative cartoon-style opponent network (GAN). This straightforward procedure uses unparalleled images and cartoon images for preparation. The two new losses to address considerable stylistic differences in the image and cartoon are proposed: (1) the loss of semantic content, designed as a scant regularisation of the huge level function maps in the VGG network, and (2) the loss of edge-promoting opponents to keep the edges clear. Jin et al. (2017) offered a strategy that combines a simple dataset with a wide range of GAN training approaches. GAN may be utilized for the creation of automated anime characters. The developers were able to construct a model with realistic anime faces.

4.7 Image to image transformation

Conversion of an input image into an output image is a typical challenge in computer graphics, image recognition, and computer vision (Torbunov et al. 2022). For this purpose, the conditional opponent networks are ideal. This family of problems is solved with the Pix2pix model (Yeo et al. 2022). CycleGAN (Zhu et al. 2017) has expanded this method with an insufficient cycle continuity that seeks to maintain the original image following a transformation and reversal cycle. Matching pictures are no longer necessary in this formulation for training. This speeds the processing of data and expands the application opportunities of the method. For example, the transmission of artist's styles (Li and Wand 2016) uses an incomparable library of painters and nature image graphs to build visions like Picasso or Monet. According to Chen et al. (2018), the generative network is divided into two networks, each dedicated to a single sub-task. The focus projected video network includes spatial focus diagrams and a transition network for object translation. The sparseness of the attention map formed by the attention network is recommended to focus more attention on things of interest. Prior to and after object modification, attention mappings should be consistent. Furthermore, if picture segmentation annotations are provided, the trained attention network will receive extra instruction. The proposed method would increase the quality of created images by teaching suitable concentration, emphasizing the importance of investigating attention in object transformation. The uncontrolled multimodal picture-to-image translation system was developed by Huang et al. (MUNIT) (Huang et al. 2018). The authors assumed that the picture display can be divided into a domain invariant content code and

a domain-specific style code. We mix the content code of an image with a random style code that is chosen to transform it to another domain from the target domain styles. The structure proposed was examined and a variety of analytical results were produced. Extensive comparisons with state-of-the-art techniques have shown significant advantages in the proposed framework. Single GAN is a unique approach for executing image-to-image translations via several domains using only one generator, according to Yu et al. (2018). In order to assure efficient translation, they used domain code to track the various generating actions directly and to include many optimization objectives. Experiments on a wide number of unpaired datasets reveal that our approach excels in translating between domains.

4.8 Text to image transformation

With modern performance, a synthesis of text-to-image is a challenge for many improvements. The synthesis of the defined techniques produces a rough outline of the image presented, but it does not express the real meaning of the text. Sample accuracy has been suggested by Fedus et al. (2018). GANs are explicitly opposing networks that enhance the capability of a generator to produce high-quality models. In image production, they achieved a lot of success. They built an actor critique, CGAN, who fills the gaps in the lack of meaning. They confirmed qualitatively and quantitatively that, relative to a highly likely model, this yields ever more natural, conditional and unconditional texts.

The GAN architecture has been used to synthesize images from word explorations, according to Reed et al. (2016). For instance, a pigeon is described in the text subtitle as 'white with some black on his head.' The trained GAN says that "Wings and a long orange beak" can create a series of pictures corresponding to the description. In addition to the text definitions as a condition, the Generative GAWWN Framework provided an immersive interface where huge images could be progressively built up using Adversarial Where Network conditions (GAWWN) for the image positions (Reed et al. 2016) with text definitions of the sections and user-supplied bounding boxes.

4.9 Human pose estimation

The method of measuring a person's pose is known as human pose estimation. Body structure (pose) derived from a single, traditionally pic, monocular one of the most important aspects of human pose assessment is issued in computer vision that has been researched for a long time more than 15 years. Ge et al. (2018) suggests the usage of the Feature Distilling Generative Adversarial Network (FD-GAN) to learn identity-related and to present unrelated representations. The proposed system is based on a Siamese configuration with

multiple novel individuals and discriminators of identification. They also presented a new approach to the integration of poses, requiring an identical look of the images produced by similar individuals. After learning utilizing unrelated attributes, recommendations, no auxiliary information, and increased computational costs are projected during testing.

4.10 De-occlusion

Occlusion happens when one object blocks a three-dimensional image of another object. “De-Occlusion” means the removal of an obstruction that obstructs the vision of an object. In addition, Wu et al. (2019) recommended a method to synthesize individual images labeled automatically and to use them to enhance the number of samples in datasets per identity. The author used rectangles of blocks to destroy the individual’s random portions. They suggested a GAN model to synthesize images that are equivalent to but not like the original pictures using coupled occlusion and original photos. They then used de-occluded photos to add to the workout samples by identifying them with the raw photographs. The amplified data sets have been used for the basic model.

In order to tackle pedestrian occlusion and lack of resolution, Fabbri et al. (2017) has proposed the use of a deep convolutionary generative model (DCGAN) The recognition network of attributes, the reconstruction network, and the super-resolution network are all sub-networks in your model. The final attributes of the categorization system were estimated by the authors by combining global and local portions. The deep features were deleted with ResNet50 and the relevant score was obtained using the average global pooling. The final prediction value is derived by combining these values. In order to overcome the occlusion and low-resolution problems, they proposed the usage of a deep generative opposed network (Goodfellow et al. 2014) to create reconstructed and super-defined images. Their model recognized the properties of a multi-label classification network using photos that were pre-created.

4.11 Text mining

Text mining, also known as information data mining, is the process of structuring unstructured text data in order to find insightful patterns and new information. Employing sophisticated analytical techniques like Nave Bayes, Support Vector Machines (SVM), and other deep learning algorithms, organizations may explore and discover hidden correlations within their unstructured data.

Yang and Edalati (2021) proposed that Schools and universities have switched to online teaching from on-campus teaching due to the COVID-19 pandemic, and mining students’ reviews towards online courses become critical in helping teachers and schools understand students’ feedback

and need as well as improving online teaching quality. But dataset imbalance is a quite often problem for sentiment classification within the education domain, which means there are much fewer neutral and negative reviews than positive reviews. The highly imbalanced dataset problem would influence the performance of sentiment classification models. We wanted to employ SOTA (State of the Art) GAN models to create content and then apply deep learning and machine learning to examine the influence of synthetic text creation on the sentiment classification job of the highly unbalanced dataset. Two SOTA category aware GAN models are trained with the imbalanced dataset. Both GAN models are trained with 250 epochs. We compared metrics results and generated samples of these two samples on three different datasets mentioned above. Finally, the category-aware GAN (CatGAN) model with a multilevel evolutionary algorithm is chosen to create text to balance the highly unbalanced training dataset for sentiment classification since it can generate higher-quality text without sacrificing text variation. The imbalanced and synthetic balanced datasets are obtained from the last experiment step. Same machine learning algorithms and deep learning models are trained on synthetic balanced and imbalanced dataset from the different dataset, respectively. The results indicate that compared with the original imbalanced dataset, the performance on accuracy and F1-score of the model trained on synthetic balanced dataset from CatGAN text generation model, is improved. Specifically, accuracy is increased from 2.039 to 4.822 percent for CR23k and CR100k dataset, whereas F1-score is increased from 2.79 to 9.208 percent for CR23k (Course Reviews) and CR100k dataset. Also, the results show that the improvement for CR100k is higher than CR23k. Also, the average performance improvement for deep learning is higher than machine learning algorithms. Due to time limitation, we have not extended our experiments on more complex sentiment analysis deep learning models such as aspect based sentiment analysis model to see how those more sophisticated models would behave on the synthetic balanced dataset. Nevertheless, these four models are the necessary parts for most NLP deep learning models used for sentiment analysis. So we infer that the performance improvement of these four models would more or less improve the performance of models with more complex architectures. Besides, just GAN text generation models are exploited while some newest transformer based text generation model such as GPT-3 (Generative Pre-trained Transformer) has not been tested yet, and the experiments are limited within the education domain. In the future, researchers could exploit different type of text generation and more complex sentiment analysis models in order to have a complete picture of the impact of synthetic text generation on the sentiment classification task of the highly imbalanced dataset. Besides, researchers can also try to construct a new

Table 2 Summary of GANs application

References	Architecture	Inference	Application
Zhang et al. (2019)	SAGAN	Inception Score: 52.52	Generation of high quality image
Huang et al. (2017)	LAPGAN	NA	Generation of high quality images
Wang et al. (2018)	SRGAN	PSNR: 30.28	Image super resolution
Nazeri et al. (2019)	Edge connect	PSNR:25.28	Image Inpainting
Li et al. (2017)	PGAN	NA	Object detection
Tulyakov et al. (2018)	MoCoGAN	NA	Video prediction and generation
Chen et al. (2018a, 2018b, 2018c)	Cartoon GAN	NA	Generation of anime character
Huang et al. (2018)	MUNIT	NA	Image to image transformation
Reed et al. (2016)	GAWWN	NA	Text to image transformation
Ge et al. (2018)	FD-GAN	mAP(Mean average precision) Market-1501 : 77.7 CUHK03 : 91.3 DukeMTMC-reID:64.5	Human pose estimation
Fabbri et al. (2017)	DCGAN	NA	De-occlusion
Yang and Edalati (2021)	SOTAGAN	F1-score: CR100k-9.21	Text mining

sentiment analysis model that can avoid the influence of a highly imbalanced dataset. The mentioned GAN applications are summarised in Table 2.

5 Advanced GANs application

GANs have been the only generative algorithms to provide excellent results; they therefore opened up numerous new research areas, and GANs ultimately are acknowledged as perhaps the most notable research in machine learning in recent years. There's many domains for which GANs shall soon be employed, including producing infographics from text, coming up with website designs, compressing data, discovering and developing new drugs, creating text and music, and several other things. In fields where computer vision plays a significant role, such as photography, image editing, and gaming, among others, GANs are employed because they learn to detect and differentiate pictures. Unsupervised neural nets, such as generative adversarial networks, train by examining data from a specified dataset to produce new picture patterns. As a result, they find use in sectors that depend on computer vision technologies, such as: strengthening cybersecurity (Yinka-Banjo and Ugot 2020), Employing artificial intelligence, neural networks, and generative adversarial networks is expected to have a significant positive impact on a number of industries, including healthcare and pharmaceuticals. Generative adversarial networks have a lot to offer the video game industry. Thanks to GANs, the job of developers and designers will be shortened. GANs may be used to automatically create the 3D models needed for cartoons, animated films, and video games (Fadaeddini et al. 2018). Transferring satellite images to Google Maps (Song et al. 2021), detail editing from day to night and vice versa ,

Changing black-and-white photos to color, Converting sketching into color photos. One intriguing application will be seen in the dentist department, where researchers are believed to be fabricating dental crowns with the great assistance of GANs, which will speed up the procedure for the patient because a process that previously took weeks can be done with high accuracy in just a few hours.

6 GAN's advantages and disadvantages

In this segment, different types of GAN models are compared and contrasted. Early GANs, such as Vanilla GAN and Conditional GAN, focused solely on supervised learning, but as shown in Table 2, this was eventually expanded to incorporate semi-supervised and unsupervised learning as well. Later adversarial designs incorporated convolutional networks, WGAN critique, Lipschitz limit, Probably Approximate Correct (PAC)-style theorem, autoencoders, and deep neural networks to replace multilayer perceptrons. Furthermore, both the generator and discriminator networks were trained using Stochastic Gradient Descent based optimization in the great majority of simulations. In all incarnations, the primary aim of any adversarial network remains a two-player mini-max game. Several models also included secondary objectives such as feature learning and representation learning via similar semantic exercises, with the learned features eventually being employed for categorization or identification in unsupervised contexts. Models like LAPGAN and GRAN have produced a sequential production of pictures by the generator using Laplacian pyramids and recurrent networks.

In addition, prior simulations relied on the measurement of the good fit for model assumptions, which were later discovered to be incorrect approximations in subsequent

iterations. Instead, the accuracy and error rates of a model were used to assess its impact. The Generative Adversarial Metric, presented by GRAN, is a new metric for calculating GAN's efficiency that no other generative model has yet to employ. Table 2 shows the results. In addition to manual assessment and evaluation of samples generated by GANs' Generator module, quantitative measurements like Inception Score (IS), PSNR, and mAP (mean Average Precision) are employed and discussed. Fig 3 shown below explains the library search outcomes (Aggarwal et al. 2021).

The main benefit of GAN is that it avoids the requirement to define the probability distribution structure of the generator model. As a result, GAN stays away from tractable density forms, which are effective for classifying complex and high-dimensional distributions. GAN has the following advantages over other models with a well-defined probability density (Karras et al. 2017). It can simultaneously sample created data. Due to its autoregressive nature, the $pg(x)$ of the pixel CNN (Karras et al. 2017), PixelRNN (Van Oord et al. 2016), and the WaveNet (Oord et al. 2016) are decomposed into a product of the conditional distributions previously established values. For example, the autoregressive models construct an image pixel per pixel, and before the value of the previous pixel, the probability distribution of the next pixel cannot be known inherently. As a result, the generation of high-size data such as speech synthesis is generally slow to handle (Oord et al. 2016). The GAN generator, by contrast, is a basic feed network from Z-to-X. The data, as with self-regressing models, is generated all at once instead of pixels by pixels. Therefore, GAN may simultaneously synthesize samples, speed upsampling and permit the use of GAN in a wider range of real-world applications.

GANs are commonly used in the manufacture of generative models. GANs are a generative algorithm to tackle spontaneously data production problems, GANs can be used. As the used architecture of the neural network does

not restrict generation, the range of data samples created is greatly broadened, particularly for the high-dimensional production of data. Also, the builder of the model has additional freedom because the neural network set-up can contain several loss functions. Return spread may generally be employed to train GANs and the training criterion is being applied by two adverse networks. The planning does neither employ the outdated Markov chain modal nor approximation inferences. There is no lower dynamic variation limit which decreases training complexity while improving training efficiency dramatically; instead, GANs can sample and predict new samples in real-time, thus improving sample output. The samples created are more diverse as the process of adverse training does not directly double or mean genuine data. The GANs formed in motion can simply be understood by people. GANs, for instance, produce incredibly sharp and lifelike images. Finally, GANs appears to be a manner to supply data that can be used by people.

GANs are advantageous and informative for semi-monitored learning and help in the building of generative structures (Wu et al. 2022). The GANs learning methodology gives no data labels except for the data source. While GANs are not intended for semi-supervised learning, they can use their training methodology to unmarked pre-training data. GANs can be trained with vast amounts of unlabelled data, then use a limited number of labelled data to create a discriminatory classification and regression model based on the unlabelled data interpretation of the trained GANs. The GAN algorithm was created to solve the minimal generator/discriminator game. Although numerous experiments have been carried out to examine the convergence and nature of the Nash balance in the GAN game, GAN training is extremely surprising and difficult to achieve. GAN uses the gradient descent approach to solve the minimax for generators and discriminators in an iterative way. The Nash balance is the parameters point where the costs of the discriminator

Library Search Outcomes

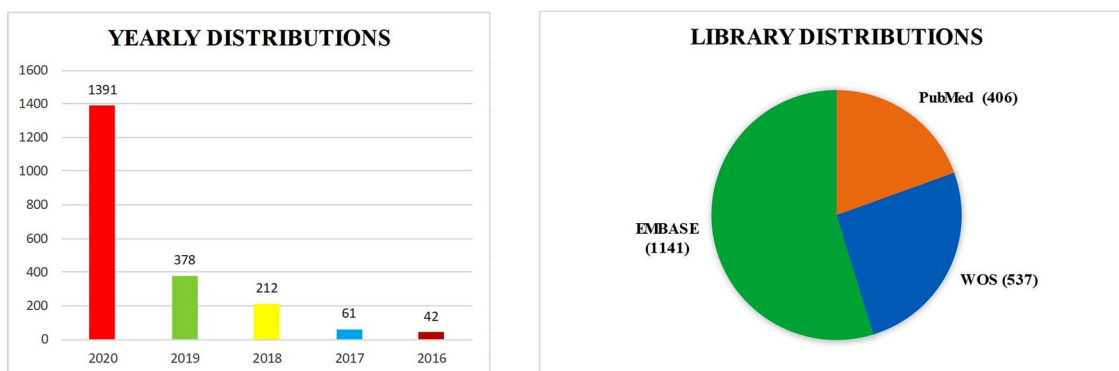


Fig. 3 Library search outcomes: yearly distribution (left) and Library distribution (Right)

and the generator are lower with regards to their parameters for the cost function $V(G; D)$. The discriminator's cost function will be reduced, but the generator costs will be raised and vice versa. This would allow the GAN game to converge. Mode collapse is another significant problem for GAN. Since the collapse of the mode restricts the capacity of GAN to be varied, this interconnection is detrimental in real-world applications. The task of the generator is to trick the discriminator, not to depict the multimodality of distribution of actual data. Several research studies have been conducted with new components or a new item function (Zhu et al. 2017; Huang et al. 2017) to try and overcome the collapse of modes. However, the mode collapse remained a problematic challenge for GAN to tackle in the event of a highly dynamic and multi-modal real data distribution.

7 GAN's limitation

GANs have addressed many generative model challenges and inspired other AI approaches, however they still have limits. GANs use adversarial training, although the models converge and the presence of an equilibrium point have yet to be shown. It is difficult to acquire satisfactory training outcomes unless the training procedure assures the symmetry and alignment of both adversarial networks. However, because the coordination of the dual adversarial networks is difficult to manage, the trained model may be unpredictable. Furthermore, being generative models based on neural networks, GANs share a similar flaw with neural networks (i.e., poor interpretability). Furthermore, despite the diversity of the samples generated by GANs, the collapse mode still occurs (Zhang 2021). To overcome this issues many measures and techniques are on process.

According to portrait analysis, the images produced by GAN appear to be photographs of actual people. People have expressed worry over the possible use of human image synthesis using GAN by scammers, resulting in the production of fraudulent images and videos. Defense Advanced Research Projects Agency (DARPA's) Media Forensics initiatives assist in combatting such bogus media profiles created by GANs, and numerous regulations are established and will be enforced by 2020.

8 Conclusion and future scope

The purpose of this article is to summarise and analyze the history of GANs, the basic theory, characteristics, changes, measures, implementations, disadvantages, and prospective scope. Furthermore, the GAN literature is summarised and interpreted. A range of GAN implementations is demonstrated in this article. New and upgraded solutions to new

and current GAN problems must be addressed to increase the efficiency of GANs. While the GAN field is an attractive topic of study, its own set of obstacles includes unstable planning, non-convergence, and according to evaluation methodology, the requirement for more computer resources and the complexity of the model. In summary, GAN is an important and beneficial area of research with many applications, although extra work must be undertaken to tackle the current issues because of its relatively short span since inception. New research is underway to address the weaknesses of GANs. For instance, WGAN can partially resolve both collapse mode and instability issues. The difficulty of avoiding a collapse of GANs mode, therefore, remains unresolved. There is also research on the essence of Nash balance and the concept of GAN model convergence. GANs are extensively utilized in computer vision, but in other areas such as natural language processing, they are less widely used. Differences in image and non-image data qualities lead to this difficulty. Since GANs can be used for a range of fascinating applications in a range of areas, research is still ongoing in this sector along with ways of improving GAN quality and performance.

Author contributions The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding No funds, grants, or other support was received.

Data availability Not applicable.

Declarations

Conflict of interest The authors have no financial or proprietary interests in any material discussed in this article. The authors declare that they have no conflict of interest.

Informed consent Not applicable.

References

- Aggarwal A, Mittal M, Battineni G (2021) Generative adversarial network: an overview of theory and applications. *Int J Inform Manag Data Insights* 1(1):100004
- Alqahtani H, Kavakli-Thorne M, Kumar G (2021) Applications of generative adversarial networks (GANS): an updated review. *Arch Comput Methods Eng* 28(2):525–552
- Arjovsky M, Chintala S, Bottou L (2017) Wasserstein generative adversarial networks. In: *International Conference on Machine Learning*, pp 214–223. PMLR
- Chen X, Duan Y, Houthoofd R, Schulman J, Sutskever I, Abbeel P (2016) Infogan: interpretable representation learning by information maximizing generative adversarial nets. In: *Advances in neural information processing systems*, pp 2171–2180

- Chen Y, Lai Y-K, Liu Y-J (2018a) CartoonGAN: Generative adversarial networks for photo cartoonization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 9465–9474
- Chen Z, Nie S, Wu T, Healey CG (2018b) High resolution face completion with multiple controllable attributes via fully end-to-end progressive generative adversarial networks. [arXiv:1801.07632](https://arxiv.org/abs/1801.07632)
- Chen D, Xu D, Li H, Sebe N, Wang X (2018c) Group consistent similarity learning via deep crf for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 8649–8658
- Cheng K, Zhu L, Yao C, Yu L, Wu X, Zheng X, Wang L, Lin F (2021) Degan based spectrum sensing data enhancement for behavior recognition in self-organized communication network. *Chin Commun* 18(11):182–196. <https://doi.org/10.23919/JCC.2021.11.013>
- Denton EL, Chintala S, Fergus R et al (2015) Deep generative image models using a laplacian pyramid of adversarial networks. In: Advances in neural information processing systems, pp 1486–1494
- Dewi C, Chen R-C, Liu Y-T, Tai S-K (2021) Synthetic data generation using degan for improved traffic sign recognition. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-021-05982-z>
- Donahue J, Krähenbühl P, Darrell T (2016) Adversarial feature learning. [arXiv:1605.09782](https://arxiv.org/abs/1605.09782)
- Durgadevi M, et al. (2021) Generative adversarial network (gan): a general review on different variants of gan and applications. In: 2021 6th International Conference on Communication and Electronics Systems (ICCES), pp 1–8. IEEE
- Fabbri M, Calderara S, Cucchiara R (2017) Generative adversarial models for people attribute recognition in surveillance. In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp 1–6. IEEE
- Fadaeiddini A, Majidi B, Eshghi M (2018) A case study of generative adversarial networks for procedural synthesis of original textures in video games. In: 2018 2nd National and 1st International Digital Games Research Conference: Trends, Technologies, and Applications (DGRC), pp 118–122. IEEE
- Fedus W, Goodfellow I, Dai AM (2018) MaskGAN: better text generation via filling in the_. [arXiv:1801.07736](https://arxiv.org/abs/1801.07736)
- Fischer A, Igel C (2012) An introduction to restricted Boltzmann machines. *Iberoamerican congress on pattern recognition*. Springer, pp 14–36
- Ge Y, Li Z, Zhao H, Yin G, Yi S, Wang X et al (2018) Fd-gan: Pose-guided feature distilling gan for robust person re-identification. In: Advances in neural information processing systems, pp 1230–1241
- Goodfellow I (2016) NIPS 2016 tutorial: Generative adversarial networks. [arXiv:1701.00160](https://arxiv.org/abs/1701.00160)
- Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems 3:2672–2680
- Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press, Cambridge (EE, UU.)
- Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC (2017) Improved training of Wasserstein GANs. In: Advances in neural information processing systems, pp 5767–5777
- Hitawala S (2018) Comparative study on generative adversarial networks. [arXiv:1801.04271](https://arxiv.org/abs/1801.04271)
- Hong Y, Hwang U, Yoo J, Yoon S (2019) How generative adversarial networks and their variants work: An overview. *ACM Comput Surv (CSUR)* 52(1):1–43
- Huang X, Li Y, Poursaeed O, Hopcroft J, Belongie S (2017) Stacked generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 5077–5086
- Huang X, Liu M-Y, Belongie S, Kautz J (2018) Multimodal unsupervised image-to-image translation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 172–189
- Im DJ, Kim CD, Jiang H, Memisevic R (2016) Generating images with recurrent adversarial networks. [arXiv:1602.05110](https://arxiv.org/abs/1602.05110)
- Jiang Y, Chang S, Wang Z (2021) Transgan: Two pure transformers can make one strong gan, and that can scale up. In: Advances in neural information processing systems, CoRR, vol. abs/2102.07074
- Jin Y, Zhang J, Li M, Tian Y, Zhu H, Fang Z (2017) Towards the automatic anime characters creation with generative adversarial networks. [arXiv:1708.05509](https://arxiv.org/abs/1708.05509)
- Karras T, Aila T, Laine S, Lehtinen J (2017) Progressive growing of gans for improved quality, stability, and variation. [arXiv:1710.10196](https://arxiv.org/abs/1710.10196)
- Kingma DP, Welling M (2013) Auto-encoding variational bayes. [arXiv:1312.6114](https://arxiv.org/abs/1312.6114)
- Kumar S, Dhawan S (2020) A detailed study on generative adversarial networks. In: 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp 641–645. IEEE
- Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, et al. (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 4681–4690
- Li C, Wand M (2016) Precomputed real-time texture synthesis with markovian generative adversarial networks. In: European Conference on Computer Vision, pp 702–716. Springer
- Li J, Liang X, Wei Y, Xu T, Feng J, Yan S (2017) Perceptual generative adversarial networks for small object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1222–1230
- Li Y, Liu S, Yang J, Yang M-H (2017) Generative face completion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 3911–3919
- Lin Z, Khetan A, Fanti G, Oh S (2018) Pacgan: the power of two samples in generative adversarial networks. In: Advances in neural information processing systems, pp 1498–1507
- Makhzani A, Shlens J, Jaitly N, Goodfellow I, Frey B (2015) Adversarial autoencoders. [arXiv:1511.05644](https://arxiv.org/abs/1511.05644)
- Mao X, Li Q, Xie H, Lau RY, Wang Z, Paul Smolley S (2017) Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp 2794–2802
- Mathieu M, Couprie C, LeCun Y (2015) Deep multi-scale video prediction beyond mean square error. [arXiv:1511.05440](https://arxiv.org/abs/1511.05440)
- Mirza M, Osindero S (2014) Conditional generative adversarial nets. [arXiv:1411.1784](https://arxiv.org/abs/1411.1784)
- Nazeri K, Ng E, Joseph T, Qureshi FZ, Ebrahimi M (2019) Edgeconnect: Generative image inpainting with adversarial edge learning. [arXiv:1901.00212](https://arxiv.org/abs/1901.00212)
- Nguyen A, Dosovitskiy A, Yosinski J, Brox T, Clune J (2016) Synthesizing the preferred inputs for neurons in neural networks via deep generator networks. In: Advances in neural information processing systems (NIPS), pp 3387–3395
- Nguyen A, Clune J, Bengio Y, Dosovitskiy A, Yosinski J (2017) Plug & play generative networks: Conditional iterative generation of images in latent space. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 4467–4477
- Odena A (2016) Semi-supervised learning with generative adversarial networks. [arXiv:1606.01583](https://arxiv.org/abs/1606.01583)

- Oord Av.d., Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A, Kavukcuoglu K (2016) Wavenet: A generative model for raw audio. [arXiv:1609.03499](https://arxiv.org/abs/1609.03499)
- Pan Z, Yu W, Yi X, Khan A, Yuan F, Zheng Y (2019) Recent progress on generative adversarial networks (GANs): a survey. *IEEE Access* 7:36322–36333
- Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. [arXiv:1511.06434](https://arxiv.org/abs/1511.06434)
- Ratliff LJ, Burden SA, Sastry SS (2013) Characterization and computation of local nash equilibria in continuous games. In: 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp 917–924. IEEE
- Reed SE, Akata Z, Mohan S, Tenka S, Schiele B, Lee H (2016) Learning what and where to draw. In: *Advances in neural information processing systems*, pp 217–225
- Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H (2016) Generative adversarial text to image synthesis. In: *International Conference on Machine Learning*, pp 1060–1069. PMLR
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323(6088):533–536
- Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X (2016) Improved techniques for training gans. In: *Advances in neural information processing systems*, pp 2226–2234
- Song J, Li J, Chen H, Wu J (2021) Mapgen-gan: a fast translator for remote sensing image to map via unsupervised adversarial learning. *IEEE J Sel Topics Appl Earth Observ Remote Sens* 14:2341–2357
- Torbunov D, Huang Y, Yu H, Huang J, Yoo S, Lin M, Viren B, Ren Y (2022) Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation. [arXiv:2203.02557](https://arxiv.org/abs/2203.02557)
- Tulyakov S, Liu M-Y., Yang X, Kautz J (2018) Mocogan: Decomposing motion and content for video generation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 1526–1535
- Van Oord A, Kalchbrenner N, Kavukcuoglu K (2016) Pixel recurrent neural networks. In: *International Conference on Machine Learning*, pp 1747–1756. PMLR
- Vondrick C, Pirsivash H, Torralba A (2016) Generating videos with scene dynamics. In: *Advances in neural information processing systems*, pp 613–621
- Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y, Change Loy C (2018) Esrgan: Enhanced super-resolution generative adversarial networks. In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pp 0–0
- Wang H, Wu W, Su Y, Duan Y, Wang P (2019) Image super-resolution using an improved generative adversarial network. In: 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), pp 312–315. IEEE
- Wang Q, Zhou X, Wang C, Liu Z, Huang J, Zhou Y, Li C, Zhuang H, Cheng J-Z (2019) Wgan-based synthetic minority over-sampling technique: improving semantic fine-grained classification for lung nodules in ct images. *IEEE Access* 7:18450–18463
- Wang J, Chang X, Wang Y, Rodríguez RJ, Zhang J (2021) Lsgan-at: enhancing malware detector robustness against adversarial examples. *Cybersecurity* 4(1):1–15
- Wu D, Zhang K, Zheng S-J, Hao Y-T, Liu F-Q, Qin X, Cheng F, Zhao Y, Liu Q, Yuan C-A et al (2019) Random occlusion recovery for person re-identification. *J Imaging Sci Technol*. <https://doi.org/10.2352/j.imagingsci.technol.2019.63.3.030405>
- Wu J, Gu Y, Li N, Huo T, Zhang J (2022) Design and implementation of marine datasets based on dcgan network. In: 2022 International Seminar on Computer Science and Engineering Technology (SCSET), pp 250–254. IEEE
- Yang R, Edalati M (2021) Using gan-based models to sentimental analysis on imbalanced datasets in education domain. [arXiv:2108.12061](https://arxiv.org/abs/2108.12061)
- Ye Y (2022) Generative adversarial networks. In: *Conference on Computer Vision and Pattern Analysis (ICCPA 2021)*, vol. 121580, p 2
- Yeh RA, Chen C, Yian Lim T, Schwing AG, Hasegawa-Johnson M, Do MN (2017) Semantic image inpainting with deep generative models. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 5485–5493
- Yeo Y-J, Sagong M-C, Park S, Ko S-J, Shin Y-G (2022) Image generation with self pixel-wise normalization. [arXiv:2201.10725](https://arxiv.org/abs/2201.10725)
- Yinka-Banjo C, Ugot O-A (2020) A review of generative adversarial networks and its application in cybersecurity. *Artif Intell Rev* 53(3):1721–1736
- You A, Kim JK, Ryu IH, Yoo TK (2022) Application of generative adversarial networks (gan) for ophthalmology image domains: a survey. *Eye Vis* 9(1):1–19
- Yu X, Cai X, Ying Z, Li T, Li G (2018) Singlegan: Image-to-image translation by a single-generator network using multiple generative adversarial learning. In: *Asian Conference on Computer Vision*, pp 341–356. Springer
- Yu J, Lin Z, Yang J, Shen X, Lu X, Huang TS (2018) Generative image inpainting with contextual attention. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 5505–5514
- Zhang K (2021) On mode collapse in generative adversarial networks. In: *International Conference on Artificial Neural Networks*, pp 563–574. Springer
- Zhang H, Goodfellow I, Metaxas D, Odena A (2019) Self-attention generative adversarial networks. In: *International Conference on Machine Learning*, pp 7354–7363. PMLR
- Zhu J-Y., Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp 2223–2232

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.