



Potential of generative adversarial net algorithms in image and video processing applications– a survey

Akanksha Sharma¹ · Neeru Jindal¹ · P. S. Rana²

Received: 13 July 2019 / Revised: 16 April 2020 / Accepted: 9 July 2020 /
Published online: 24 July 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Generative Adversarial Network (GAN) has gained eminence in a very short period as it can learn deep data distributions with the help of a competitive process among two networks. GANs can synthesize images/videos from latent noise with a minimized adversarial cost function. The cost function plays a deciding factor in GAN training and thus, it is often subjected to new modifications to yield better performance. To date, numerous new GAN models have been proposed owing to changes in cost function according to applications. The main objective of this research paper is to present a gist of major GAN publications and developments in image and video field. Several publications were selected after carrying out a thorough literature survey. Beginning from trends in GAN research publications, basics, literature survey, databases for performance evaluation parameters are presented under one umbrella.

Keywords Cost function · Image processing · Video processing · Performance parameters

1 Introduction

Generative Adversarial Networks have become extremely popular in artificial intelligence and deep learning field. Along with the development of new models, numerous applications of

✉ Neeru Jindal
neeru.jindal@thapar.edu

Akanksha Sharma
asharma2_me17@thapar.edu

P. S. Rana
psrana@gmail.com

¹ Department of Electronics and Communication Engineering, Thapar Institute of Engineering and Technology, Patiala, India

² Department of Computer Science Engineering, Thapar Institute of Engineering and Technology, Patiala, India

GANs have also been proposed. The spectrum of task-specific applications is so wide that it ranges from jellyfish swarm detection to testing of gear safety in vehicles. These claims of popularity are supported by the number of research publications in this field as shown in Fig. 1. Although other generative models like variational autoencoders are also available, GANs have many advantages. GANs can handle sharp estimated density functions, eliminate deterministic bias, and generate desired samples and good compatibility with the neural architecture. Especially in the computer vision field, GANs enjoy success in many applications e.g. Image generation, image to image translation, image super-resolution and shadow GANs etc. [35, 36, 101]. Due to the availability of vast information on GANs, this paper has narrowed down its focus on GANs in the field of image/video processing. An inherent relation between image and video is that an image is a motionless picture that does not change with time, whereas a video has third dimension time and comprises moving images as shown in Figure 2.

This paper provides a critical study on all the major developments in the fields of image and video processing that utilize generative adversarial networks. Despite vast applications, the main purpose of GANs is as a target density estimator. GANs can implicitly learn the distribution of the original set and generate samples from the estimated distribution. The reason why GANs are so efficient in image synthesis is that learning/ estimating data distribution for higher dimensions is a tedious task, which requires the construction of likelihood functions. Currently, there is an enormous requirement of image/video algorithms using GANs as artificially generated images demand is increasing in deep learning simulations given by WANG et al. [36] and Gonog [35]. Figure 3 shows paper publications from 2014 to 2020 April on year basis, for image /video processing using generative adversarial networks. Papers were recognized through online manual search including all journals, conferences, review and others using keywords “image processing using generative adversarial networks” and “video processing using generative adversarial networks” separately on the semantic

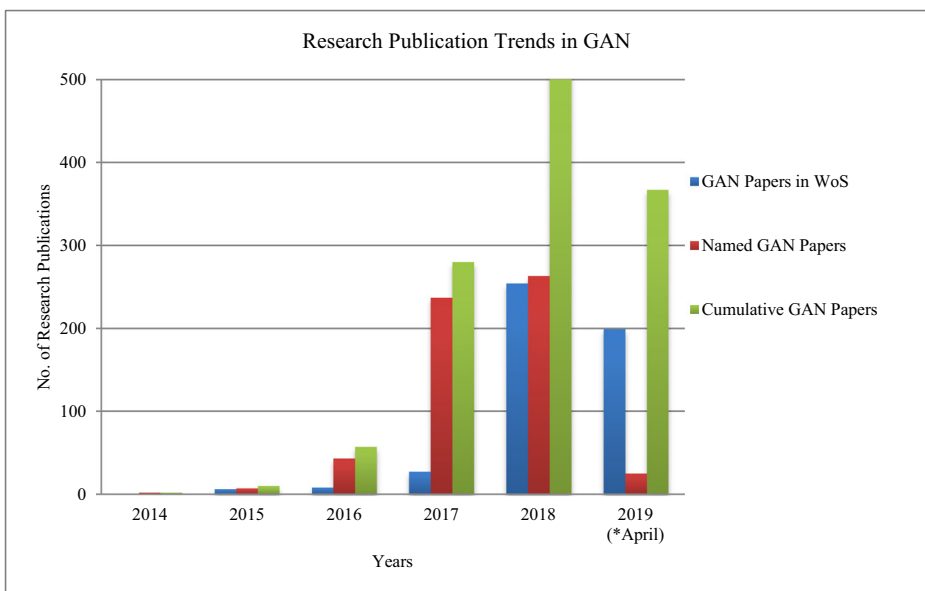


Fig. 1 Trends in research concerning generative adversarial networks have shown exponential growth

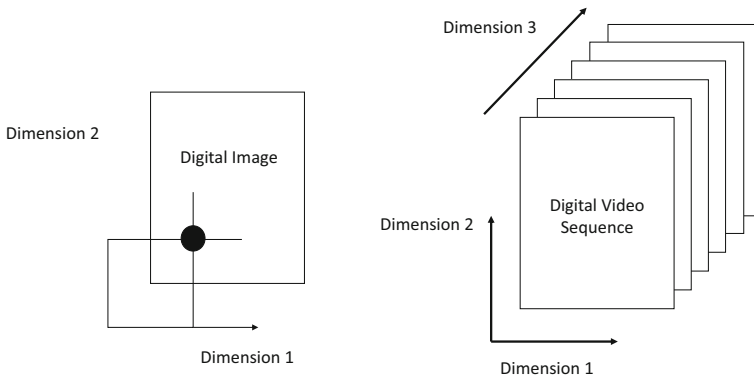


Fig. 2 Image and video dimensionality

scholar search engine [41, 44–46]. Semantic Scholar is AI-backed search engine for articles and highlights the important and significant papers [25, 27, 32, 34].

There has been substantial progress in image/video applications using GANs due to their advantages and disadvantages [47, 89, 90, and]. To summarize, the contributions of the paper are:

- To provide a vital understanding of the different images as well as video processing GANs, their architectures, cost functions, models that exist or are developing for driving today’s world.
- In scenarios where one is interested to measure and compare the quality of the GANs; quantitative and qualitative measures of GANs are effectively presented with their pros and cons under one umbrella in this paper.
- An extensive demonstration of image/video real-world applications using GANs over human activities and finally concluded with future challenges of GANs.

1.1 Basic building blocks and architecture of GAN

xGenerative Adversarial Network consists of the generator (G) and its discriminator (D) as shown in Fig. 4. The generator provides fake data by taking noise vectors as input. This fake

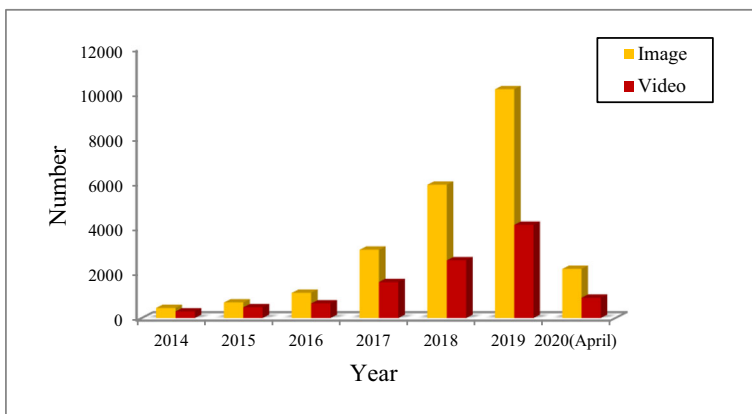


Fig. 3 Image and video GANs publications from Semantic Scholar

data is given to the discriminator with the training data. The discriminator works as a simple classifier to classify the samples original or fake from the input. This is executed by assigning probabilities to the samples [51, 62, 64, 65]. The real sample has a probability of ‘1’ whereas the forged sample has ‘0’ probability. The information in the form of the gradient is backpropagated to generator networks. This task helps the generator to learn the features of the training dataset and in turn, it generates samples images that are equivalent to the statistical distribution of original distribution.

In the next step, the discriminator accepts both the original data and the generated fake samples. Now, it performs the task of decision making whether the images are fake or not, and this learning process repeats on. The discriminator estimates the ratio of densities and passes it to the generator in the form of a gradient. The features are learned cooperatively, oscillating between two blocks, the generator and its discriminator. In the starting of the min-max game, winning for discriminator is an effortless task. But as the training process continues, the game becomes more challenging for both the player because the generator starts converging towards original data [78, 80, 96, 98, 108].

1.2 Input, training and cost functions

1.2.1 Input noise

Multiple approaches are used to provide the inputs to Generator Network. The inputs to the generator can be provided throughout to any layer in the model. G can be divided into two vectors and the first noise vector can be provided at the first layer, while the second noise

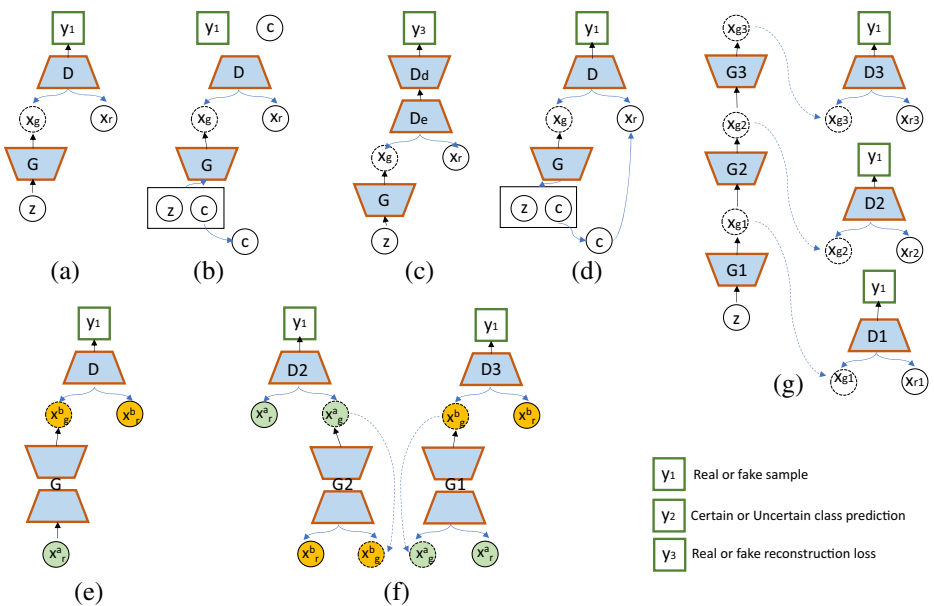


Fig. 4 Schematic view of major GAN models. (a) Original GAN model [9, 42] (b) Info GAN [20] (c) BEGAN [11] / EBGAN [161] (d) Conditional GAN [93] (e) Pix2Pix [56] (f) Cycle GAN [164] and (g) LAPGAN [30]. ‘z’ is the noise vector, c is condition labels, G is generator and D is discriminator. Subscript g refers to the generated sample and subscript r refers to the real sample

vector can be provided at the last layer. Another approach is to provide multiple random noise vectors to the inner hidden layers of the generator network.

1.2.2 Training

The training steps for GAN are summarized in Fig. 5. GAN is a structured probabilistic model. It consists of latent variables (noise) ‘ z ’ and apparent variables x (original distribution). The G function (the generator) takes ‘ z ’ (the noise vectors) as its input and uses $\theta^{(G)}$ as parameters, while the D function (the discriminator) takes ‘ x ’ (data samples from fake and original distribution) as input and uses $\theta^{(D)}$ as parameters.

The answer to the optimization of the game between the discriminator and the generator is finding its minimum value, i.e. an optimal point in parameter zone where all the other parameters have an equal or higher cost. It is solved by using Nash equilibrium, local minima $J^{(D)}$ related to $\theta^{(D)}$ and local minima of $J^{(G)}$ with respect to $\theta^{(G)}$ is calculated as a possible solution. By training the discriminator, an estimate is obtained for $p_{\text{data}}(x)/p_{\text{model}}(x)$ for each point in the x domain. Computation of the vast majority of divergences and their gradients are enabled by this ratio. GANs base their supervised form of learning on this ratio to make approximations. Simultaneous stochastic gradient descent is applied for training each step. Two small (mini-batches) are selected, one from ‘ x ’ and the other one from z . Thus, any of the gradient-based optimization algorithms can be used to update the two gradient steps simultaneously. The most commonly used optimization algorithm is Adam.

1.2.3 Role of cost functions

The cost functions of both D and G play a crucial role in the training of the GAN model. The discriminator cost function is defined in such a way that it maximizes the probability of the sample being counterfeit, while the generator cost function maximizes the probability that the generated sample is real. The vanilla GAN model explored three approaches for generator cost function, namely, the min-max approach, the heuristic approach and the maximum likelihood approach. Newer GAN models have explored different mechanisms to calculate the distance between original and generated distribution. The original GAN paper used KL divergence and Jensen-Shannon divergence [42]. Changing the cost function leads to different training behaviour and outcomes, hence researchers have explored several methods like Earth-Movers distance [4], χ^2 distribution [89], etc. Cost functions of major GAN variants are concised in Table 1.

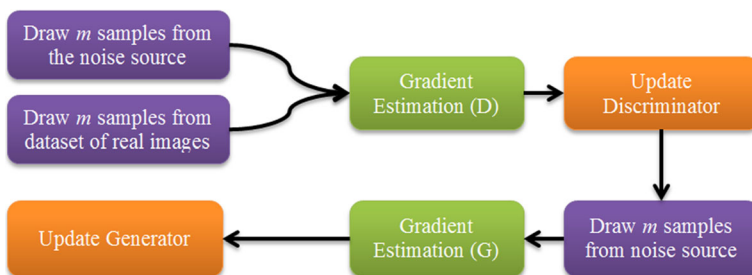


Fig. 5 Steps of GAN model training

2 Literature survey

It was Goodfellow et al. [42] who proposed the GAN firstly with the objective of an algorithm that is efficient in replicating the data distribution of original data using maximum likelihood approximation and Jensen-Shannon divergence. In practice, the two blocks, G and D were two convolutional neural networks; the former performed the task of generation while the latter performed the task of classification of samples into fake and original by allocating them probabilities between 0 and 1.

But soon after, many types of research independently discovered that GAN models lack training stability and suffer from mode collapse [122, 123, 133]. DCGAN [107] model was the first one to offer a simple solution. Instead of using the sigmoid activation function, it used the ReLU function. DCGAN also stitched the already popular deep convolutional nets [2] with the GAN model to further improve the image quality.

As generated image quality improved, more emphasis was laid on conditional image generation. Utilizing the initial GAN model, Mirza et al. [93] came up with a conditional GAN model to generate samples based on class labels. CGAN can also learn multi-modal data distribution and can be used to generate descriptive tags that are not part of the training labels. Taking inspiration from CGAN, InfoGANs [20] focused on learning detangled representation in an unsupervised manner. It works on the principle of maximizing mutual information by optimizing a lower bound of mutual information.

As using ReLU activation proved insufficient to stabilize the training, researches explored different cost functions and divergence functions. Denton et al. [30] took inspiration from the Laplacian pyramid framework and implemented it using a cascade of the convolutional network. The Laplacian pyramid helps in generating images with rough texture and appearance, and as the pyramid progresses, finer details are added. LS GAN introduced by Mao et al. [89, 90] used a new loss function for its GAN model, called the least-squares loss function. It is a distinct case of minimizing the Pearson χ^2 divergence. It also works on pulling the threshold or the decision boundary closer to the model of original data samples.

A major landmark was achieved by Salimans et al. [113], where the authors proposed several techniques to combat training instability, non-convergence, and mode collapse. Major GAN models and their details are summarized in Table 2. The training issues faced by researchers are summarized in table number 3 (Table 3).

WGAN [4] was a pioneer in GAN models, not only did it use the Wasserstein metric (called the earth movers distance); it also provided the concept of weight clipping. It continuously estimates the EM distance, which is useful in debugging hyper-parameter searches. WGAN's improved version [47] provided optimization in the form of a gradient penalty. The authors found that weight clipping leads to difficulty in optimization. Due to k-lipschitz constraints, WGAN models leaned towards learning extremely simple functions and ignored higher moments. This issue was solved by using the gradient penalty in accordance with lipschitz continuity. Using the k-lipschitz continuity, where k is lipschitz constant, the issue of complicated optimization was resolved. Other changes that were made are cancelation of batch normalization for discriminator model and using the two-sided penalty in order to keep k close to 1. Improved WGAN proved to be much more stable for training GANs.

WGANs paved the way for more GAN models to utilize EM distance. Berthelot et al. [11] proposed a new method where loss is derived from EM distance to train auto-encoder based

Table 1 Summary of major Discriminator cost functions

GAN Variant	Discriminator Loss Function	Generator Loss Function
Vanilla GAN [42]	$L_D^{GAN} = -E_{x \sim p_{data}}[\log(D(x))] - E_{z \sim p_{disc}}[\log(1 - D(G(z)))]$	$L_G^{GAN} = -E_{x \sim p_z}[\log(D(G(z)))]$
cGAN [93]	$L_D^{cGAN} = -E_{x \sim p_{data}}[\log(D(x, c))] - E_{z \sim p_z}[\log(1 - D(G(z), c))]$	$L_G^{cGAN} = -E_{x \sim p_z}[\log(D(G(z), c))]$
WGAN [4]	$L_D^{WGAN} = -E_{x \sim p_{data}}[D(x)] + E_{z \sim p_z}[(1 - D(G(z)))]$	$L_G^{WGAN} = -E_{x \sim p_z}[(D(G(z)))]$
WGAN-GP [47]	$L_D^{WGAN-GP} = L_D^{WGAN} + \lambda E_{x \sim p_{data}}[(\ \nabla D(x) + (1 - \alpha G(z))\ - 1)^2]$	$L_G^{WGAN-GP} = -E_{z \sim p_z}[(D(G(z)))]$
EBGAN [161]	$L_D^{EBGAN} = E_{x \sim p_{data}}[(D_{AE}(x))] + E_{z \sim p_z}[\max(0, m - D_{AE}(G(z)))]$	$L_G^{EBGAN} = E_{x \sim p_z}[(D_{AE}(G(z)))]$
BEGAN [11]	$L_D^{BEGAN} = E_{x \sim p_{data}}[(D_{AE}(x))] - k_f E_{x \sim p_z}[D_{AE}(G(z))]$	$L_G^{BEGAN} = E_{x \sim p_z}[(D_{AE}(G(z)))]$
Info GAN [20]	$L_D^{InfoGAN} = L_D^{GAN} - \lambda L_1(c, c')$	$L_G^{InfoGAN} = L_G^{GAN} - \lambda L_1(c, c')$
LSGAN [89]	$L_D^{LSGAN} = -E_{x \sim p_{data}}[(D(x) - 1)^2]$	$L_G^{LSGAN} = -E_{x \sim p_z}[(D(G(z)) - 1)^2]$
Improved LSGAN [90]	$L_D^{Improved LSGAN} = \frac{1}{2} E_{x \sim p_{data}}[(D(x) - 1)^2] + \frac{1}{2} E_{z \sim p_z}[(D(G(z)) + 1)^2]$	$L_G^{Improved LSGAN} = \frac{1}{2} E_{z \sim p_z}[(D(G(z)) - 1)^2]$
DRAGAN [63]	$L_D^{DRAGAN} = L_D^{GAN} + \lambda E_{z \sim p_z + N(0, \sigma)}[(\ \nabla D(G(z))\ - 1)^2]$	$L_G^{DRAGAN} = E_{z \sim p_z}[\log(1 - D(G(z)))]$
RaGAN [57]	$L_D^{RaGAN} = E_{x \sim P} [f_1(C(x_f) - E_{x_f \sim Q} C(x_f))] + E_{x_f \sim Q} [f_2(C(x_f) - E_{x_f \sim P} C(x_f))]$	$L_G^{RaGAN} = E_{x_f \sim P} [g_1(C(x_f) - E_{x_f \sim Q} C(x_f))] + E_{x_f \sim Q} [g_2(C(x_f) - E_{x_f \sim P} C(x_f))]$

Table 2 Training Challenges in Generative Adversarial Networks

Training Issue	Description
Mode Collapse	It happens when G is not able to generate multiple modes present in training dataset. It is observed when multiple input noise vectors are delineated to the same point in the latent space. This results in a lack of variety in the fake samples.
Non – Convergence	The convergence of a GAN model is subject to assumptions like the Infinite capacity of the model, A convex cost function and Enough number of training samples.
Vanishing Gradients	As the generated model improves after every step, the two data distributions start to overlap. Vanishing Gradients are a consequence of Jensen-Shannon divergence saturation and result in very small gradients or complete absence of gradients.
Performance Evaluation	GANs suffer from a lack of performance evaluation methods. Several performance evaluation methods like the Inception Score, Fréchet Inception Score (FID), etc. have been used till now.

GAN called BEGAN. It is a new equilibrium enforcing method which provides a new convergence method, called proportional control theory, with a robust GAN architecture. Another model that used GAN alongside auto-encoder was EBGAN [161], where discriminator is used as energy function which allocates low energy to the zones which are closer to original data distribution and high energy elsewhere.

Improved WGAN model heavily influenced researchers to come up with new normalization and penalty methods. One such model was proposed by Kodali et al. [63] where GAN training was studied as regret minimization, where a new gradient penalty method called Deep Regret Analytic GAN, was used. Miyato et al. [94] suggested the spectral normalization technique that embodies weight normalization which has been regarded as easy to include in existing models as compared to weight normalization, weight clipping, and gradient penalty.

Another improvement was proposed by Alexia Jolicoeur-Martineau [57] in the form of relativistic GAN and relativistic averaged GANs. The author argued that the probability of real data being real should be decreased as the generator becomes more efficient. Experimental results showed that RaGANs with gradient penalty outperformed WGAN-GP and reduced the time to reach the state-of-the-art result by four times. RaGAN even produced better quality pictures from a very small set ($n = 2011$), which is not possible with LSGAN.

3 Current trends

The basic objective of generative adversarial networks is image generation. However, GANs have been molded to perform a plethora of applications in several fields. Due to flexible nature, numerous GAN models have been combined with other techniques such as attention [158], relativism [57], etc. to perform specific tasks. In most cases, GANs have outperformed existing models. In the present work, the emphasis is laid on applications of GAN in the field of image/video applications, which are discussed henceforth [138, 147].

3.1 Image processing

GANs are basically synthetic image generation algorithms. Over the years, they have been used to perform several tasks in the field of image processing.

Table 3 Summary of Major GAN Models along with their datasets, performance parameters, and contributions

Publication	GAN Name	Specialty	Dataset Used	Performance Parameters	Contribution
Goodfellow et al. [42]	Vanilla GAN	Image Generation	MNIST, Toronto Face Database, CIFAR	Parzen window-based log-likelihood approximations	First to introduce GANs and generate synthetic images
Mirza et al. [93]	cGAN	Conditional Image Generation	MNIST, MIR Flickr 25,000	Parzen window based log likelihood approximations	Introduced conditional image synthesis in GANs
Denton et al. [30]	LapGAN	Image Generation	Cifar10, STL, LSUN	Parzen window-based log-likelihood approximations	Used Laplacian architecture to progressively build better images
Radford et al. [107]	DCGAN	Image Generation	Imagenet 1 k, LSUN, Faces, Cifar10, SVHN	Classification Accuracy	To extract more feature information, VGG network was inducted into G and D of GANs, use of ReLU
Chen et al. [20]	Info-GAN	Image Generation based on info maximization	MNIST, CelebA, SVHN, 3D faces, 3D chairs	Human Evaluation	Mutual information maximization, better representation of distangled spaces
Zhao et al. [161]	EBGAN	Image Generation	MNIST, LSUN, CelebA, ImageNet	Inception Score (IS)	Use of Hinge Loss, Autoencoder as discriminator
Mao et al. [89]	LSGAN	Image Generation	LSUN, HWDB1.0, MNIST, Chinese Characters	Human Evaluation	Use of Least Squares loss for computation of divergence and quality improvement.
Berthelot et al. [111]	BEGAN	Image Generation	CelebA	IS	Balances the power of the D against the G, image diversity & image quality
Ajrovsky et al. [4]	WGAN	Image Generation	LSUN, CIFAR10, Toy DB	IS	Use of Wasserstein Metric (Earth Movers distance)
Gulrajani et al. [47]	Improved WGAN	Image Generation	LSUN, CIFAR10, Toy DB	IS	Introduction of Gradient penalty to combat effects of weight clipping in WGAN model
Mao et al. [90]	Improved LSGAN	Image Generation	Cats, LSUN, MNIST, ImageNet	Frchet Inception Distance (FID)	Introduction of Gradient penalty to combat effects of weight clipping in the LSGAN model
Zhang et al. [158]	Self-Attention GAN	Image Generation	ImageNet	IS, FID	Use of Self attention mechanism to obtain higher details for objects in images
Alexia Jolicœur-Martineau [57]	RatGAN	Image Generation	Cats, Cifar10	FID	Use of a relativistic D and G along with spectral normalization

3.1.1 Image generation

As the primary task of GAN is image generation, it has been summarized in Table 4. The rest of the applications are discussed in upcoming sections.

3.1.2 Inpainting

Image inpainting is used to fill relevant data in images which are incomplete, damaged or missing information without any prior knowledge. Among the most popular GANs which incorporate inpainting are PatchGAN and its derivatives. Demir et al. [28] proposed PG GAN which combined two previous versions of GAN, namely Global or GGAN and patchGAN. PGGAN shares its layers with both these GAN networks but later bifurcates to produce two adversarial losses that feed the generator network. Yu et al. [148] presented a new method for inpainting using gated convolution which has masks and inputs. It uses a dynamic feature selection mechanism. Introducing a new GAN model called SNPatchGAN as the authors used spectral normalized discriminators. It can eliminate undesirable or distracting objects, improve image layouts, clean-up watermarks and generate a new one as well. Later Yu et al. [149] proposed another GAN model that was capable of utilizing surrounding image features for better predictions. It is capable of handling multiple, multi-size holes at arbitrary locations. The above mentioned GAN models are used for 2D image inpainting. Wang et al. [132] proposed a 3D GAN to carry out 3D shape inpainting, which utilized all the tools including 3D Encoder - Decoder, GAN along with a LRCN. The 3D models are processed in the form of 2D shapes. While the GAN framework captures the global 3D structure, the LRCN produces the finer details.

3.1.3 Image translation

The endeavor to achieve translation requires more effort because the context has to be preserved while doing so. Copious amounts of algorithms have been deployed to achieve near-perfect translation. Inspired from CGAN, pix-2-pix is capable of generating images from label information, reconstruction of objects from edge information, filling and coloring the image, etc. It has become popular among researchers and has been implemented on different types of data, as it is capable of adapting its loss according to the task at hand. It produces extraordinary results, especially for those translation tasks which involve highly structured graphical outputs.

Image to image translation is implemented by mapping between the images by training the set of image pairs. Zhu et al. [164] presented a novel algorithm to translate the images when paired samples are not available. Cycle uniformity loss was used to implement this as this type of mapping is under-constrained. Inverse mapping was also used in the same algorithm. This GAN is known as Cycle GAN due to its cycle consistency loss in action. It also performed object transfiguration task, photo enhancement, season and style transfer, etc. Although compelling results were obtained, they were not uniform for all cases and applications. It succeeds when color and texture changes are required. However, in tasks that require geometric changes, the performance was subpar and left scope for further improvement. Another popular GAN model for cross-domain style transfer is DISCO GAN [60]. It addressed the task of discovering the cross-domain relation when provided with unpaired data. Using the discovered cross-domain relations, disco GAN successfully transferred style from one domain

Table 4 Summary of Major GAN Models used for Image Translation

Publication	GAN Name	Specialty	Datasets Used	Performance Parameters	Contribution
Yi et al. [146]	Dual GAN	Image to Image Translation	Photo Sketch Day Night Label Facades Aerial Maps	Amazon Mechanical Turk (AMT) studies	Employs two (generators and discriminators) to attain translation, heavy computation
Zhang et al. [155]	StackGAN	Text to Image Translation	CUB200, Oxford102, MS COCO	Inception Score (IS), Frchet Inception Distance (FID)	Two-stage image synthesis with two discriminators and two generators; stage I gives rough shape, stage II adds details
Kim et al. [60]	Disco GAN	Image to Image Style Transfer	CAR, FACE, Chairs, Handbag, Shoes	Root-Mean-Square Error (RMSE)	Complete style transfer from one domain to other, irrespective of geometric dissimilarities
Zhang et al. [156]	StackGAN++	Text to image Translation	CUB200, Oxford102, MS COCO	IS, FID, High Resolution (HR)	Two-stage image synthesis with various Generators and Discriminators
Isola et al. [56]	pix2pix	Image to Image Translation	Monet, horse to zebra, Cezanne, Yosemite summer to winter apples 2 oranges	AMT studies, Per-pixel acc. (ppA), Per-class acc. (pcA), class Intersection-over-Union (IOU)	Image-image translation on paired datasets
Zhu et al. [164]	Cycle GAN	Image-image Translation	Monet, horse to zebra, Cezanne, Yosemite summer to winter apples 2 oranges	AMT, Fully Convolutional Network FCN, ppA, pcA, class IOU	Image-image translation on unpaired/unaligned datasets in cross-domain
Wang et al. [167]	pix2pixHD	Image to Image Translation	Monet, Van Gogh, Cezanne, Ukiyo-e horse to zebra Yosemite summer to winter apples 2 oranges	Mean IOU, Pixel Acc %	Improved resolution on results produced by pix2pix model

to the other. It preserved the key attributes, like the identity of the face and orientation. However, it does not handle mixed modalities well. It worked without open pair labels and then learns to relate the datasets from diverse domains.

3.1.4 Super resolution

Super-resolution is used to enhance the visual quality and details in an image given in Table 5. Ledig et al. [66] solved the problem of improving the texture details on a large scale while super resolving. The authors use perceptual loss which is the blending of content and adversarial loss. Content loss played a critical part in super-resolution as it takes into account the perceptual similarity rather than pixel similarity. Wang et al. [129] made significant improvements in SRGAN and introduced Enhanced SRGAN. The authors changed both the loss functions as well as the network architecture of SRGAN. They implemented the technique used in RaGAN to use a relativistic discriminator. While updating the network architecture, they used Residual in Residual Dense Block as the building block. To achieve better brightness consistency, improving perceptual loss and texture recovery, features extraction is performed before activation in ESRGAN. This provided improved visual quality with new realistic textures.

3.1.5 Segmentation

Segmentation is performed in image processing and multiple algorithms have been put to use in this regard. Ehsani et al. [105] took a new leap with SeGAN in a bid to complete the appearance of occluded objects from scenes. To complete this task, the knowledge of “which pixels to paint?” and “what color to paint them?” is crucial. These two questions result in the segmentation of the invisible parts and then the generation of invisible parts. SeGAN optimizes both these tasks jointly. SeGAN is trained on synthetic photorealistic images and can reliably segment natural images. It is also capable of depth layering using the occluder occluded relationships. A comparison of Typically used Image GANs is given in a summarized way in Table 6.

3.1.6 Real-world image applications

There is a rapid growth of applications deploying real-time image applications using generative adversarial networks. GAN can generate real-like image samples using random latent vector z . There is no need to know real data distribution and assumption of mathematical conditions. These are the main reasons that GAN is used in several academic/engineering/ almost in every field. Several researchers have presented prominent applications of GANs in real-world image processing applications. [2, 3, 77, 167].

High-resolution human face images can be generated from low-resolution images while up-sampling and using the trained model inferring the photo-realistic details. Bulat et al. [15] proposed real-world face super-resolution because many existing methods remain to fail to generate good results while implemented on real-world low-resolution and low-quality face images. To solve this problem, a two-stage process was presented, in which firstly a High-to-Low Generative Adversarial Network (GAN) is trained for degradation and downsample high-resolution images were required during training. After this, the output of the network is given to Low-to-High GAN to train image super-resolution using these time extirpated low and high-

Table 5 Summary of major GAN models used for Image Resolution Improvement

Publication	GAN Name	Specialty	Datasets Used	Performance Parameters	Contribution
Ledig et al. [66]	SRGAN	Super Resolution of a single image	DIV2K, Set5, Set14, BSD100, BSD300	Peak Signal-to-Noise ratio (PSNR), Structural similarity index (SSIM), Mean opinion score (MOS), Mean-square error (MSE), VGG22, VGG54	The first study to achieve single image super resolution, successful but contains stray artifacts
Karras et al. [59]	ProGAN	Resolution Improvement	CelebA, LSUN, CIFAR10	sliced Wasserstein distance (SWD), Multi-scale SSIM MS-SSIM	Resolution improvement using progressive growing of GANs
Wang et al. [129]	ESRGAN	Super Resolution of a single image	DIV2K, Flickr2K, OST, Set5, Set14, BSD100, Urban100, PIRM-SR	PSNR, SSIM	Enhanced SRGAN to improve hallucinated details and artifacts produced by SRGAN, use of RRDB, RaGAN, feature addition before activation

Table 6 Advantages, Disadvantages and features of typical GANs [17]

GAN Model	Advantages	Disadvantages	Scenarios
DCGAN	Stability, easy convergence, generate diverse sample	Model collapse	Useful for most of the scenarios
CGAN	The label is added to make convergence faster	A large dataset is required	Supervised and semi supervised-learning,
WGAN	Stable training solved gradient disappearance problem	Pruning of gradients is not accurate	For other models of GANs, do not converge
WGAN GP	Generator and discriminator are balanced, stable training, direct process	Convergence is slow	For uncertain model parameters

resolution images. Figure 6 shows the simulation results of face super resolution for low resolution faces in the real-world.

Realistic pedestrian images in real scenes were generated using the GAN model in 2018 [102] and then train the CNN-based pedestrian detector using this augmented data. Generative Adversarial Network (GAN) of multiple discriminators was used, with the purpose to generate realistic pedestrians and learn the background simultaneously. To solve the issue of different sizes of the pedestrians, a Spatial Pyramid Pooling layer was used in the discriminator. Simulations on cross-dataset were also performed i.e., training the model on one dataset and testing on another dataset. It was concluded that PSGAN is also able to produce well with improvement in the performance of CNN-based detectors. Figure 7 shows the PS-GAN model used to generate realistic pedestrian images.

WaterGAN [73] is a water generative adversarial network used to generate underwater realistic images from the in-air image and along this color of monocular underwater images was also corrected using the unsupervised pipeline as shown in Fig. 8. High-resolution images can be captured using autonomous cameras and remotely operated vehicles to map the seafloor. With WaterGAN, a large training dataset was generated for raw underwater as well for true color in-air. It served as input to a novel network for the color correction of underwater images.

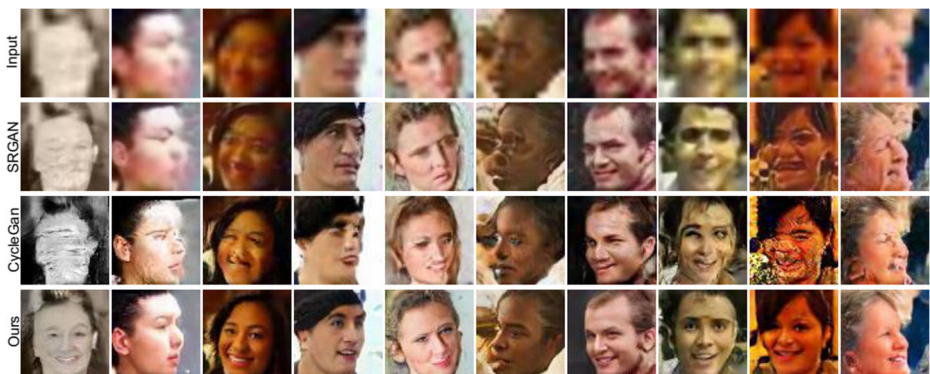


Fig. 6 Super-resolution results produced by Bulat et al. [15] on real-world low-resolution faces and compared to SRGAN [67] and CycleGan [165]

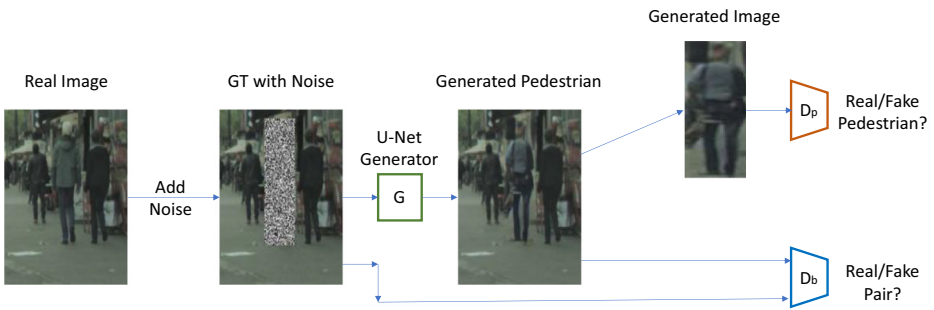


Fig. 7 PS-GAN model using multiple discriminators (Dp and Db) to generate realistic pedestrian images [102]

3.2 Video processing

GANs have also been utilized in video processing. Some of its major applications are discussed in the upcoming sections. The applications are also summarized and important publications are listed in Table 8.

3.2.1 Frame generation and prediction

Intermediate frame generation is an important task. Previous methods that were deployed to perform this task using interpolation yielded videos of less quality with extreme blurriness. Keyframe generation is an important research area as it is used in slow motion filming, compressing videos, forensic analysis etc. Wen et al. [134] used two concatenated GANs for frame generation. One GAN captured the motions while the other one generated the frame details. The novel technique used three losses, namely, adversarial loss, normalized product correlation loss and gradient difference.

For both video generation and recognition tasks, a frame transform model is needed. However, due to the availability of a large number of ways for frames and object change, the model of dynamics is a challenging task. Vondrick et al. [124] presented a GAN video for separating the scene’s foreground and background video using spatiotemporal convolutional architecture. Simulation results prove that this model generates full frame rate tiny video. It has the utility to predict

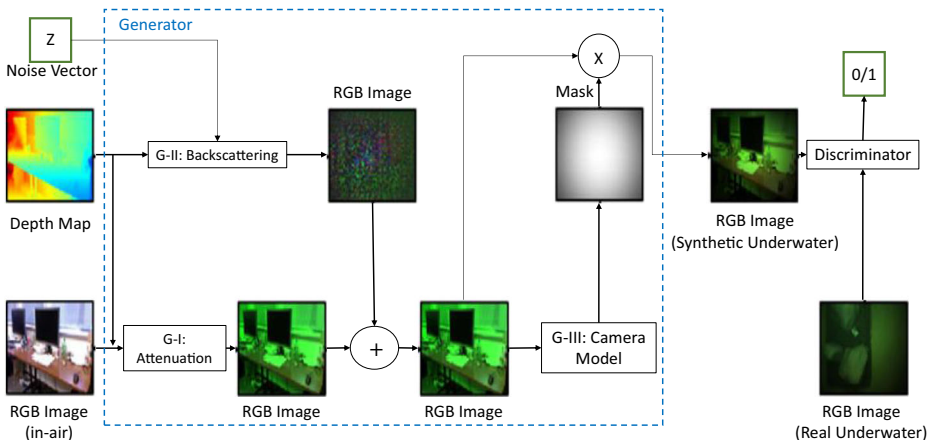


Fig. 8 WaterGAN model used for generating realistic underwater images [73]

plausible future of still images. Moreover, visualizations and experiments also show that the model learns suitable features with minimum supervision. Mathieu et al. [92] trained CNN to produce future frames when an input sequence is given. Three feature learning strategies were also given to solve the problem of blur prediction observed from Mean Squared Error (MSE). UCF101 dataset was used in this work to compare the predictions. Tulyakov et al. [121] suggested MoCoGAN, Motion and content decomposed Generative Adversarial Network for the generation of video. It will generate a video by mapping random vectors series to video frames series. A random vector has content and motion parts. The content part remains fixed and the motion is kept as a variable process. Image and video discriminators use adversarial learning scheme. Experimental results on the various dataset with performance comparable to the existing approaches, verify the efficacy of the framework.

3.2.2 Video De-blurring

Video de-blurring is a difficult task as it requires modeling along with the spatial domain which comprises image planes and temporal domain which comprises of the neighboring frames from the video. Another challenge is the retrieval of sharp images along the lines of pixel-wise error. To combat these issues, Zhang et al. [160] suggested that 3D convolution can be used in the spatial-temporal domains using DBLRNet for video de-blurring. To address the problem of the generation of sharp images, the DBLRNet was used as the generator in the GAN architecture. In addition to the regular adversarial loss, a content loss was also used. This GAN was named de-blurring GAN and was tested on a benchmark dataset, achieving better results.

Shen et al. [115, 116] presented an encoder and decoder triple-branch architecture. For sharpening foreground (FG) and background (BG) details; two branches were learned, and the last branch produces harmonious. This model was further endowed with a supervised, human-aware awareness mechanism up to end fashion. To encode FG human information, the soft mask is used and it used the FG or BG decoder branches to give attention to domains. To further get advantage from Human-aware Deblurring, the HIDE dataset, having blurry and sharp image pairs was also proposed. HIDE covered a large number of scenes, motion patterns, and background scenes.

3.2.3 Haze removal

Pang et al. [104] proposed the Haze Removal GAN or the HRGAN for removing Haze from videos in order to provide better visual quality in surveillance videos and their analysis. It contains a specialized generator and discriminator network. The generator works on an estimation of transmission maps, haze-free images and atmospheric light all at once. Changes were also proposed to cover adversarial loss and pixel-wise loss all created by the discriminator. The authors demonstrated the superiority of HRGAN against state-of-the-art de-hazing algorithms.

3.2.4 De-identification

xGANs have also been used to hide information in videos. Brkic et al. [14] used GANs for person de-identification which means that removal of non-biometric features and replacing them with the ones generated via GANs. The non-biometric features include clothing colors, hairstyles etc. In some cases, even the faces can be replaced. Such tools are very helpful when a person's information is to be concealed as simple methods like video de-blurring do not offer much safety. The de-identified videos are also immune to re-identification attacks.

3.2.5 Video super-resolution

Lucas et al. [83] made the first of its kind attempt to use GAN for video super resolution. They made a new GAN model called the VSRResFeatGAN model where the generator is named VSRResNet. The generator is pre-trained with the mean square loss which in turn leads to a better performance quantitatively as compared to other VSR models. This model of GAN used a new evaluation metric called the PercepDist metric to accurately examine the perceptual quality of videos as opposed to formerly used SSIM metrics.

Chen et al. in 2017 [21] suggested a video super resolution framework using GAN with temporal information fusion. It was observed that the temporal information of consecutive video frames contributes to the task of video super-resolution. The main contribution of this work was twofold: a new generator architecture was suggested with various mechanisms of temporal fusion of information, with early fusion, slow fusion as well as 3D convolution. Also compared with others as shown in Fig. 9. Then a discriminator architecture based on SRGAN with an adaptive training routine was applied to train the GAN and to keep the burden of hand-tuning easy for the large number of model parameters. Figure 7 shows the video frames of the using temporal fusion, bicubic interpolation and per-frame based SRGAN.

Several GANs are used nowadays for video resolution according to the requirement of the application. But to differentiate between them, PSNR is selected as a common parameter given in Table 7.

3.2.6 Real-world video applications

To spread out from image processing to video is quite a difficult task due to three dimensions of videos, limitation of memory and training stability introduces challenges. In 2019 [53], video enhancement using divide-and-conquer approach using adversarial learning was suggested, which divide and merge on perception-based, frequency-based as well as dimension-based problems. Mainly photo enhancement process was decomposed into multiple problems, which was recovered back using bottom to up. At the top-level, to learn additive and multiplicative components; a perception-based division was presented. It was required to convert a low-quality image/video into high-quality. However, at the intermediate level, frequency-based division using a generative adversarial network was used to supervise photo enhancement. The lower level used a dimension-based division to enable the GAN model for better estimation of the distribution distance at multiple data and to train the GAN model.



Fig. 9 Comparison of Foreman and football frames. Left to right: Ground truth frame, bicubic interpolation, SRGAN, early fusion, slow fusion and 3D convolution [21]

Table 7 Difference in super-resolution Video GANs based on evaluation parameters

Dataset	GAN	PSNR	SSIM
VIDSET4	VSRResFeatGAN [81]	30.90	0.94
	ERSGAN	22.98	0.63
	SRGAN	23.58	0.705
MYANMAR	VSRResFeatGAN [84]	40.58	–
	VSRResAffNet [82]	40.69	0.98

Video generation was divided into frame and sequence generation using GANs [53], and the task became easy to solve as well. For this, a two-step training scheme was used: a generator was trained with static frames as shown in Fig. 10. Afterward, to generate natural look scenes, a recurrent model was trained from the previously trained frames generator. Both training steps used the adversarial steps. However, to avoid training instabilities while using GANs, an approach of multiple discriminators was used.

Chu et al. [24] proposed a temporally self-supervised algorithm and adversarial learning to get temporally coherent solutions without loss of spatial details. Ping-Pong loss was suggested for better long-term temporal consistency. It efficiently avoids artifacts in the recurrent networks without depressing features. Moreover, the progressive growth of Generative Adversarial Networks is shown in Fig. 11 for higher resolution video generation in 2018 [1]. In particular, videos of low resolution and short-duration were generated, and then gradually increase in resolution and duration by adding new spatio-temporal convolutional layers was performed. The progressive model learns spatio-temporal information and generates higher

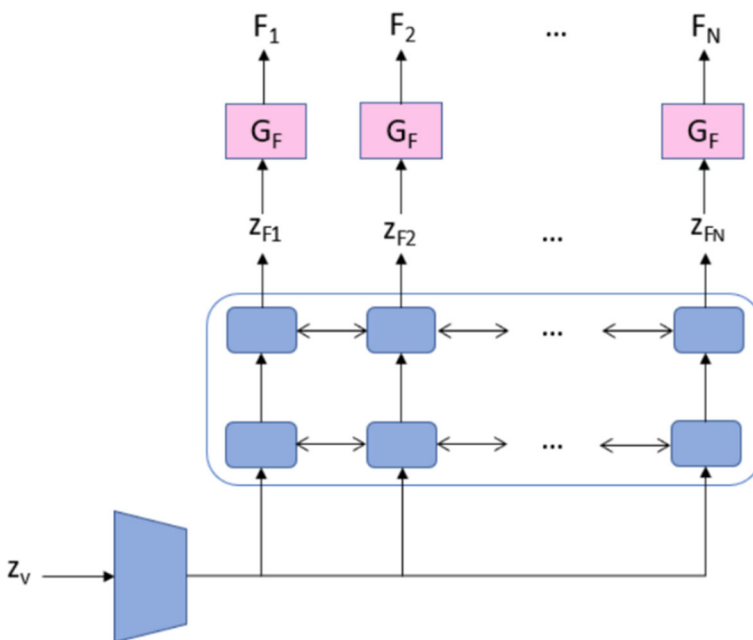


Fig. 10 Video generator using GAN, Pink blocks are pre-trained frame (F_1, F_2, \dots, F_N) generator with Z_F noise vectors 1 to N [53]

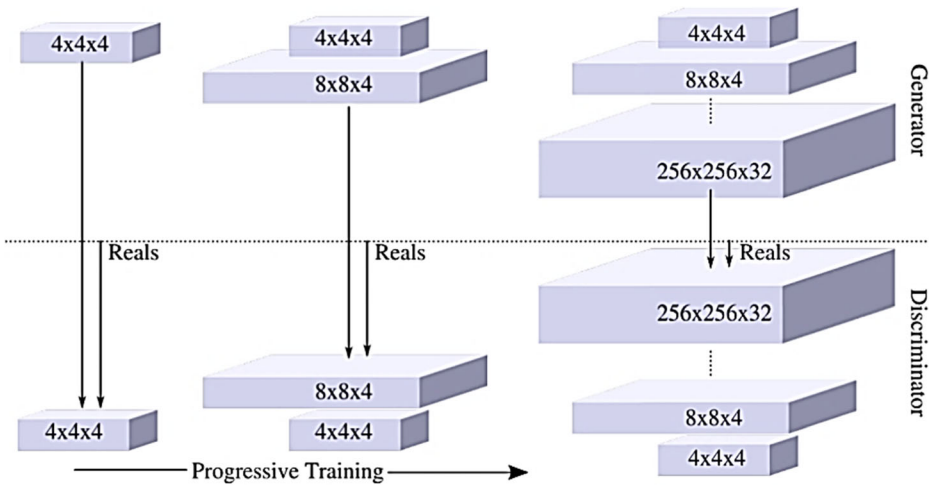


Fig. 11 Video generation using the Progressive model [1]

resolution video. Table 8 presents major Publications of Generative Adversarial Networks in various fields.

4 Performance evaluation

With the availability of a plethora of GAN models for different applications, it has been observed that a suitable metric GAN's evaluation is still a challenge, for comparison between generative models. However, still there exists no globally accepted benchmark metrics for evaluating the performance of a GAN architecture in all aspects. But, there are some essential and desirable characteristics of GAN evaluation metrics as defined below. An effective GAN evaluation metric should possess the following characteristics.

- It should favor models that create highly distinguishable generated samples from real ones.
- It should be a sensitive over-fitting of the model.
- It should have well-defined boundary values.
- It must be sensitive to image distortions and transformations.
- It should be able to match human perceptual judgments and rankings of models.
- It must have low computational complexity. GANs performance is measured using qualitatively and quantitative measures.

Qualitative evaluation is visual check by humans which includes naturalness, shapes, perspective, and structure as well. It is the common and most used way to evaluate GANs as given in Table 9.

While qualitative measures help to inspect models, it has some drawbacks also. Firstly, image quality evaluation with human vision is cumbersome and biased [76]. It is difficult to replicate and does not reflect the capacity of the model. Secondly, the high variance in human inspectors makes it compulsory to be normal over a huge number of subjects [29]. For instance, it also fails to observe whether a model drops modes. Actually, mode dropping

Table 8 Major Publications of Generative Adversarial Networks in various fields

Field of Study	References
Audio and Speech Processing	2018- IEEE Access -Yuan et al. [151], 2018- IEEE T Audio Speech-Saito et al. [111], 2019- IEEE T Audio Speech-Tuan et al. [120]
Industrial Electronics	2018-IEEE T Ind Electron-Xuan et al. [141], 2019-IEEE T Ind Electron-Li et al. [74], 2019- IEEE T Ind Electron - Chen et al. [22], 2018 - IEEE T Ind Electron - Li et al. [71]
Medical Science	2019-IEEE T Med Imaging-Mardani et al. [91], 2018-IEEE T Med Imaging-Yang et al. [143], 2017-IEEE T Med Imaging-Wolterink et al. [135], 2018-IEEE T Med Imaging-Costa et al. [26], 2019-IEEE T Med Imaging-Gao et al. [37], 2018-IEEE T Med Imaging-Quan et al. [106], 2019-IEEE T Bio-Med Eng-Nie et al. [95], 2018-IEEE T Med Imaging-Shan et al. [114], 2019-IEEE T Med Imaging-Yu et al. [150], 2017-arXiv- Hu et al. [52], 2018-IEEE T Med Imaging-Bentaieb et al. [8], 2018-IEEE T Med Imaging-Yang et al. [144], 2018-IEEE T Med Imaging-Huo et al. [54], 2018-IEEE T Med Imaging-Wang et al. [130], 2018-IEEE T Med Imaging-Huo et al. [131]
Geoscience and Remote Sensing	2018- IEEE Geo Sci Remote S-Shi et al. [117], 2018- IEEE Geo Sci Remote S-Ghamisi et al. [39], 2017- IEEE Geo Sci Remote S-Gong et al. [40], 2019- IEEE Geo Sci Remote S-Dong et al. [33], 2018- IEEE Geo Sci Remote S-Zhang et al. [157], 2018- IEEE Geo Sci Remote S-Zhan et al. [153], 2019- IEEE Geo Sci Remote S-Niu et al. [97], 2019- IEEE Geo Sci Remote S-Ma et al. [88], 2018-IEEE Geo Sci Remote S-Zhu et al. [166], 2018- IEEE Geo Sci Remote S-Oliveira et al. [100], 2019- IEEE Geo Sci Remote S-Lin et al. [79], 2019- IEEE Geo Sci Remote S-Bermudez [10]
Image Segmentation	2018-IET Comput Vis-Chang et al. [18], 2018-IEEE Access-Li et al. [70], 2018-IEEE Access-Wu et al. [136], 2018-IEEE Access-Sakkos et al. [112]
Image Inpainting	2018-IEEE Conference-Wang et al. [128], 2018- arXiv-Demir et al. [28], 2018-arXiv-Yu et al. [148], 2018-IEEE Conference-Yu et al. [149], 2018-IEEE Access-Li et al. [72]
Translation	2018-IEEE Access-Ge et al. [38], 2018-IEEE Signal Proc Let-Wang et al. [132], 2017-IEEE Conference-Yi et al. [146], 2017-IEEE Conference-Isola et al. [56], 2018-IEEE Conference-Choi et al. [23], 2017- arXiv-Royer et al. [110], 2018-Adv Neural Inf Sys-Zhu et al. [163], 2017- arXiv-Zhu et al. [164], 2018-IEEE Conference-Ma et al. [87], 2017-IEEE Conference-Zhang et al. [155], 2017-arXiv-Zhang et al. [156]
Security and forensics	2019- IEEE T Inf Foren Sec-Xu et al. [140], 2019- IEEE T Inf Foren Sec-He et al. [49], 2019- IEEE T Inf Foren Sec-Cao et al. [16], 2019- IEEE T Inf Foren Sec-Li et al. [75], 2018- IEEE Signal Proc Let-Kim et al. [61]
Video Processing	2018- IEEE T Circ Syst Vid -Pang et al. [104], 2019-IEEE T Circ Syst Vid-Lee et al. [68], 2019-IEEE T Image Process-Zhang et al. [160], 2018-IEEE T Image Process-Yang et al. [145], 2019-IEEE T Image Process-Lucas et al. [83], 2018-IEEE T Circ Syst Vid -Wen et al. [134], 2017-IET Signal Process-Brkic et al. [14], 2017-IEEE Conference-Liang et al. [76], 2018-IEEE Conference-Xiong et al. [139], 2018-IEEE Conference-Tulyakov et al. [121], 2017- Adv Neural Inf Sys-Bhattacharjee et al. [12], 2017-IEEE Conference-Walker et al. [125], 2018-IEEE Conference-Bansal et al. [7], 2018-ACM Conference-Pan et al. [103], 2016- Adv Neural Inf Sys-Vondrick et al. [124], 2018-AAAI Conference-Ohnishi et al. [99]

helps in visual sample quality. Hence quantitative measures have an eminent part in GANs evaluation as summarized in Table 10 [13]. It will be important to note that most evaluation schemes do not examine the disentanglement in the latent zones. The table also represents relative ratings in terms of high, moderate and low. “-” means the value is missing or unknown and requires further research. “*” shows that there are multiple scores available in that group measure.

Table 9 Summary of Qualitative Performance Evaluation Metrics used in image GANs [13, 31]

Measure	Description
Nearest Neighbours	Nearest neighbors are calculated using Euclidean distance to detect overfitting.
Rapid scene categorization	It is expected that human is able to report particular characteristics of scenes in a little glance
Rating and Preference Judgment	Fidelity of the models in calculated to rank the models.
Mode drop & Collapse	modes are calculated by measuring the distances between generated data to mode centers
Network Internals	Exploring the interior representation & dynamics of models and visualizing learned.

4.1 Pros and cons

Based on the above analysis, the advantages and inherent limitations of the most significant evaluation metrics can be summarized, and the conditions under which they produce meaningful results. Some metrics enable us to study the problem of over-fitting, perform model selection on GAN models and compare GAN models without resorting to human evaluation based on selected samples. There is no particular standard to select the best score. Basically, several scores evaluate aspects of the image generation process, rather it is very difficult that a single score may be used to measure. However, some measure is more considered than others. In the over-fitting case, nearest neighbor visualization and rapid categorization quality measures are mostly used. Overall, it is an open problem to find a measure for diversity and visual fidelity evaluation simultaneously. Diversity denotes that all modes can be covered whereas visual fidelity implies that there should be a high likelihood for generated samples. Parzen window estimation is an example of likelihood favours trivial. But, it almost fails to estimate the true likelihood for high dimensional spaces. IS and FID are two mostly used scores, that depend on deep networks (pre-trained) to equate original and generated samples. IS calculates correlation and diversity of generated images with quality. However, it can evaluate Pg only as an image generation model in comparison to its similarity to Pr. It may encourage GAN models to learn sharp images; in spite of Pr. Some evaluation methods like MS-SSIM try to evaluate the diversity of generated images. It can detect mode collapse; as well as how can a generator detects the true data distribution. MMD is always able to categorize generative/noise images from real images. Both have low computational complexity. Given these advantages, even though MMD is biased, still it is recommended. Figure 12 shows the comparison of typical GANs for performance parameters F1, Precision and Recall. where the best F1 score for a fixed model is selected and vary the budget. It is observed that even for this simple task, GAN models struggle to achieve a high F1 score. Analogous plots for precision or recall for various thresholds have been given also.

5 Future challenges

GANs have become extremely popular in a very short duration of time, making the world of GANs a magnum opus in itself. These are robust algorithms that can be manipulated and molded to suit according to requirements. Owing to this ability, numerous application-specific

Table 10 Summary of Quantitative Performance Evaluation Metrics used in GANs [13, 31, 36]

Measure	Discriminability	Over-fitting Detection	Well Defined Bounds	Perceptual Judgements	Sensitivity to Distortions
Classification Performance [36]	High	Low	[0,1]	High	–
Frechet Inception Distance [35, 50]	High	Moderate	[0, ∞]	High	High
Accuracy of real and synthetic shadow scene [101]	High	Moderate	[0,1]	High	Moderate
Maximum Mean Discrepancy [31]	Low	High	[0,1]	High	Moderate
Inception Score [113]	High	Moderate	[1, ∞]	High	Moderate
Wasserstein Critic [4]	High	Moderate	[0, ∞]	–	–
Precision, recall, F1 Score [85]	Low	High	[0,1]	–	–
Image Quality Measures [58, 118, 126]	Low	Moderate	*	High	High
Adversarial Accuracy and Divergence [142]	High	Low	[0,1][0, ∞]	–	–
Low level Image Statistics [59, 152]	Low	Low	*	Low	Low
Reconstruction error [137]	Low	Low	[0, ∞]	–	Moderate
NRDS [159]	High	Low	[0,1]	–	–
Generative Adversarial Metric (GAM) [55]	High	Low	*	–	–
Image Retrieval Performance [127]	Moderate	Low	*	Low	–
Classification Performance NDB [109]	High	Low	[0,1]	Low	–
Two Sample Classifier [69]	Low	High	[0, ∞]	–	Low
Birthday paradox test [5]	High	Low	[0,1]	–	–
AM Score [162]	Low	High	[1, ∞]	Low	Low
Mode Score [19]	High	Moderate	[0, ∞]	High	Moderate
Modified IS (m-IS) [48]	High	Moderate	[0, ∞]	High	Moderate
Average Log-Likelihood [42, 119]	High	Moderate	[1, ∞]	High	Moderate
Maximum Mean Discrepancy [43]	Low	Low	[−∞,∞]	Low	Low
Maximum Mean Discrepancy [43]	High	Low	[0, ∞]	–	–

GANs have been developed in diverse fields. In the future, more of such applications can be developed and amalgamated with newer models and mechanisms like attention and relativism.

- In the future, GANs can be used in the following potential resolutions:
- Creating infographics from text
- Generating website designs:
- Compressing data
- Drug discovery and development
- Generating music
- Researches can find a more stable distance metric for GAN loss function.
- As GAN lack any universal performance metric, there is an urgent need to invent one as soon as possible.

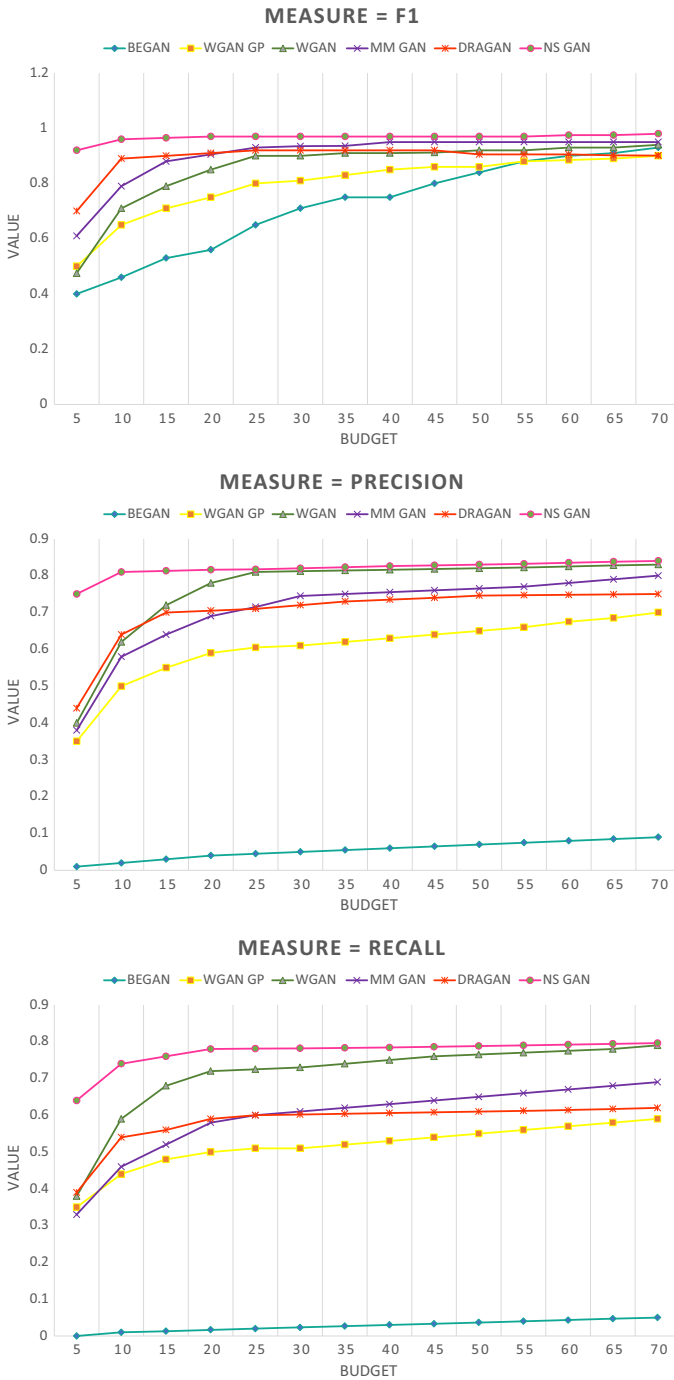


Fig. 12 Comparison of F1, Precision, Recall vs. computational budget for Boundary equilibrium generative adversarial networks (BEGAN), Deep Regret Analytic Generative Adversarial Networks (DRAGAN), Mean Minimax GAN (MM GAN), Non-Saturating GAN (NS GAN), Wasserstein GAN (WGAN), Wasserstein GAN gradient penalty (WGAN GP). [86]

- The newest deep learning mechanisms are the attention mechanism [154] and relativism. It will be great to see an amalgam of both these new technologies in the future along with reinforcement learning [6].
- Best aspect of GANs is that the general models can be easily molded to form new variants or new application specific variants in miscellaneous fields of audio processing, biological sciences, medical science, astronomy etc.
- Mode Collapse: An open-ended futuristic research area is the tendency of the system to collapse, also called mode collapse. The system is a combination of two interacting entities. Feedback and input from each are pivotal to the survival of the other. In this situation, they might go on forever or failure of one might cause overall failure.
- Equilibrium: From the system's perspective, the system is designed to minimize the cost function of each generator and the discriminator but it isn't designed to reach Nash Equilibrium which causes the system to keep going without reaching anywhere. Hence, the convergence of the system is not guaranteed. Hence another open-ended research area is algorithms that help a system reach equilibrium.
- Repetition: GANs often fall into a trap wherein when one image is accepted by the discriminator i.e. manages to fool the discriminator, they either create the same image again and again or give minor variations of the image as output. This limits the diversity in the output and introduces inflexibility in the system.
- Evaluation metrics: As of now, there is a lack of proper evaluation metrics for GANs. The ones being used are nearest neighbor algorithm for measuring the distance between the produced image/sample and the training set.
- Training instability—Saddle points become indefinite in GANs. So the optimal solution is to find a saddle point than its local minimum.

6 Conclusion

In this paper, image/video processing GANs have been thoroughly reviewed. Publication trends, development, various new variants of GAN and applications have been discussed in-depth in an attempt to bring all relevant information under one umbrella. The recent rush among the scientific community to utilize GANs in core as well as interdisciplinary fields is proof that GANs have a very promising future. Their ability to deal with missing data and unlabelled data is also very peculiar. They have obvious advantages over other generative models. However, they have also introduced some new problems. Their training instability, mode collapse, and non-convergence have led to the development of various new models. But in a paper by Google brain, after a thorough examination, authors reached the conclusion that it cannot be established that any of the new variants have any consistent upper hand over the initially proposed vanilla GAN algorithm. Conclusions can be summarized as below:

- It can be very easily concluded that GANs have a plethora of applications; however, there is still a need for an algorithm that provides stability in GAN training. Much of the doubts remain due to unavailability of mathematical proofs in support of the GAN mechanism. Some authors have tried to investigate the underlying cause of training failures and mode collapse, but it is more of speculation rather than a solution. In terms of stable training, the loss function plays an important role and various loss functions have been given to deal

with this viewpoint. Having reviewed different types of loss functions, it is observed that spectral normalization has good generalization i.e., it is able to be applied to every GAN, is easy to be implemented and has a very low computational cost.

- Despite numerous old and new performance evaluation metrics, GANs lack a universal metric that keeps all the parameters in view. This paper has given the strengths and limitations of quantitative and qualitative performance measures used for evaluating GANs. The current favorites among researchers, IS and FID, both fail to represent all the information required to represent while analyzing GANs. Ultimately some directions for future research have been concluded for evaluation measures. There should be comparative empirical and analytical studies for evaluation measures in the future. A code repository for evaluation measures must be available on authenticated websites for researchers.

Finally, it is concluded that there are many prospects for future research and applications in many fields in particular.

References

1. Acharya D (2018) Towards High Resolution Video Generation with Progressive Growing of Sliced Wasserstein GANs. arXiv,1–22
2. Agnese J, Herrera J, Tao H, Zhu X (2020) A survey and taxonomy of adversarial neural networks for text-to-image synthesis. *Data Min Knowl Disc.* <https://doi.org/10.1002/widm.1345>
3. Alqahtani H, Kavakli-Thome M, Kumar G (2019) Applications of generative adversarial networks (GANs): an updated review. *Archives Computat Methods Eng.* <https://doi.org/10.1007/s11831-019-09388-y.1-28>
4. Arjovsky M, Chintala S, Bottou L (2017) Wasserstein GAN. arXiv preprint arXiv:1701.07875
5. Arora S, Zhang Y (2017) Do GANs actually learn the distribution? An empirical study. arXiv preprint arXiv:1706.08224
6. Azar MG, Munos R, Kappen HJ (2013) Minimax PAC bounds on the sample complexity of reinforcement learning with a generative model. *Mach Learn* 91(3):325–349
7. Bansal A, Ma S, Ramanan D, Sheikh Y (2018) Recycle-GAN: Unsupervised video retargeting. In *Proceedings of the European Conference on Computer Vision*, pp 119–135
8. BenTaieb A, Hamarneh G (2018) Adversarial stain transfer for histopathology image analysis. *IEEE Trans Med Imaging* 37(3):792–802
9. Berg T, Liu J, Woo Lee S, Alexander ML, Jacobs DW, Belhumeur PN (2014) Birdsnap: Large-scale fine-grained visual categorization of birds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 2011–2018
10. Bermudez JD, Happ PN, Feitosa RQ, Oliveira DA (2019) Synthesis of multispectral optical images from SAR/optical multitemporal data using conditional generative adversarial networks. *IEEE Geosci Remote Sens Lett* 16:1220–1224
11. Berthelot D, Schumm T, Metz L (2017) BcGAN: boundary equilibrium generative adversarial networks. arXiv preprint arXiv:1703.10717
12. Bhattacharjee P, Das S (2017) Temporal coherency based criteria for predicting video frames using deep multi-stage generative adversarial networks. In *advances in neural information processing systems*, pp 4268–4277
13. Borji A (2018) Pros and cons of GAN evaluation measures. *Comput Vis Image Underst*
14. Brkić K, Hrkać T, Kalafatić Z, Sikirić I (2017) Face hairstyle and clothing colour de-identification in video sequences. *IET Signal Process* 11(9):1062–1068
15. Bulat A, Yang J, Tzimiropoulos G (2018) To learn image superresolution, use a GAN to learn how to do image degradation first. *ECCV*, 185–200
16. Cao J, Hu Y, Yu B, He R, Sun Z (2019) 3D aided duet GANs for multi-view face image synthesis. *IEEE Trans Inform Forensics Secur* 14:2028–2042
17. Cao Y, Jia LL, Chen YX, Lin N, Yang C, Zhang B, Liu Z, Li X, Dai H (2019) Recent advances of generative adversarial networks in computer vision. *IEEE Access*, 14985–15006, 7

18. Chang W, Yang G, Yu J, Liang Z (2018) Real-time segmentation of various insulators using generative adversarial networks. *IET Comput Vis* 12(5):596–602
19. Che T, Li Y, Jacob AP, Bengio Y, Li W (2016) Mode regularized generative adversarial networks. *arXiv preprint arXiv:1612.02136*
20. 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 2172–2180
21. Chen K Y, Lu C Y, Xing Y (2017) Video super-resolution using temporal fusion generative adversarial network
22. Chen X, Yu J, Kong S, Wu Z, Fang X, Wen L (2019) Towards real-time advancement of underwater visual quality with GAN. *IEEE Trans Ind Electron*
23. Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J (2018) StarGAN: Unified Generative Adversarial Networks for multi-domain image-to-image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 8789–8797
24. Chu M, Xie Y, Mayer J, Leal-Taixe L, Thurey N (2018) Learning Temporal Coherence via Self-supervision for GAN-based Video generation. *arXiv*,1–22
25. Coates A, Ng A, Lee H (2011) An analysis of single-layer Networks in unsupervised feature learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp 215–223
26. Costa P, Galdran A, Meyer MI, Niemeijer M, Abràmoff M, Mendonça AM, Campilho A (2018) End-to-end Adversarial retinal image synthesis *IEEE transactions on medical imaging* 37(3):781–791
27. Cousins S, Shawe-Taylor J (2017) High-probability minimax probability machines. *Mach Learn* 106(6): 863–886
28. Demir U, Unal G (2018) Patch-based image inpainting with generative adversarial networks. *arXiv preprint arXiv:1803.07422*
29. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database
30. Denton EL, Chintala, Fergus R (2015) Deep generative image models using a laplacian pyramid of adversarial networks. In *Advances in Neural Information Processing Systems*, pp 1486–1494
31. Dinh L, Sohl-Dickstein J, Bengio S (2016) Density estimation using real nvp. *arXiv preprint arXiv: 1605.08803*
32. Donahue J, Krähenbühl P, Darrell T (2016) Adversarial feature learning. *arXiv preprint arXiv:1605.09782*
33. Dong J, Yin R, Sun X, Li Q, Yang Y, Qin X (2019) Inpainting of remote sensing SST images with deep convolutional generative adversarial network. *IEEE Geosci Remote Sens Lett* 16(2):173–177
34. Fahlman SE, Hinton GE, Sejnowski TJ (1983) Massively parallel architectures for AI: NETL thistle and Boltzmann machines. In *National Conference on Artificial Intelligence*
35. Frey BJ, Hinton GE, Dayan P (1996) Does the wake-sleep algorithm produce good density estimators? In *Advances in neural information processing systems*, pp 661–667
36. Frey BJ, Brendan JF, Frey BJ (1998) *Graphical models for machine learning and digital communication*. MIT press
37. Gao Y, Liu Y, Wang Y, Shi Z, Yu J (2019) A universal intensity standardization method based on a many-to-one weak-paired cycle generative adversarial network for magnetic resonance images. *IEEE Trans Med Imaging* 38:2059–2069
38. Ge H, Yao Y, Chen Z, Sun L (2018) Unsupervised transformation network based on GANs for target-domain oriented image translation. *IEEE Access* 6:61342–61350
39. Ghamisi P, Yokoya N (2018) *Img2dsm: height simulation from single imagery using conditional generative adversarial net*. *IEEE Geosci Remote Sens Lett* 15(5):794–798
40. Gong M, Niu X, Zhang P, Li Z (2017) Generative Adversarial Networks for change detection in multispectral imagery. *IEEE Geosci Remote Sens Lett* 14(12):2310–2314
41. Goodfellow IJ, Bulatov Y, Ibarz J, Amoud S, Shet V (2013) Multi-digit number recognition from street view imagery using deep convolutional neural networks. *arXiv preprint arXiv:1312.6082*
42. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S and Bengio Y (2014) Generative Adversarial nets. In *Advances in neural information processing systems*, pp 2672–2680
43. Gretton A, Borgwardt KM, Rasch MJ, Schölkopf B, Smola A (2012) A kernel two-sample test. *J Mach Learn Res* 13:723–773
44. Griffin G, Holub A, Perona P (2007) Caltech-101 object category dataset
45. Griffin G, Holub A, Perona P (2007) Caltech-256 object category dataset
46. Grzegorzczuk M (2016) A non-homogeneous dynamic Bayesian network with a hidden Markov model dependency structure among the temporal data points. *Mach Learn* 102(2):155–207
47. 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

48. Gurumurthy S, Kiran Sarvadevabhatla R, Venkatesh Babu R (2017) DeligAN: Generative Adversarial Networks for diverse and limited data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 166–174
49. He Y, Zhang J, Shan H, Wang L (2019) Multi-task GANs for view-specific feature learning in gait recognition. *IEEE Trans Inform Forensics Secur* 14(1):102–113
50. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S (2017) GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In advances in neural information processing systems, pp 6626–6637
51. Hinton GE, Sejnowski TJ (1986) Learning and relearning in Boltzmann machines parallel distributed processing: explorations in the microstructure of cognition 1(282-317) 2
52. Hu B, Tang Y, Chang EI, Fan Y, Lai M, Xu Y (2017) Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks. arXiv preprint arXiv: 1711.11317
53. Huang Z, Paudel DP, Li G, Wu J, Timofte R, Gool LV (2019) Divide-and-Conquer Adversarial Learning for High-resolution Image and Video enhancement, arXiv,1–17
54. Huo Y, Xu Z, Bao S, Bermudez C, Moon H, Parvathaneni P, Landman BA (2018) Splenomegaly segmentation on multi-modal MRI using deep convolutional networks. *IEEE Trans Med Imaging*
55. Im DJ, Kim CD, Jiang H, Memisevic R (2016) Generating images with recurrent adversarial networks. arXiv preprint arXiv:1602.05110
56. Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional Adversarial Networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1125–1134
57. Jolicoeur-Martineau A (2018) The relativistic discriminator: a key element missing from standard GAN. arXiv preprint arXiv:1807.00734
58. Juefei-Xu F, Boddeti VN, Savvides M (2017) GANg of GANs: generative adversarial networks with maximum margin ranking. arXiv preprint arXiv:1704.04865
59. Karras T, Aila T, Laine S, Lehtinen J (2017) Progressive growing of GANs for improved quality and stability and variation. arXiv preprint arXiv:1710.10196
60. Kim T, Cha M, Kim H, Lee JK, Kim J (2017) Learning to discover cross-domain relations with Generative Adversarial Networks. In Proceedings of the 34th International Conference on Machine Learning-Volume 70, pp 1857–1865. *JMLRorg*
61. Kim D, Jang HU, Mun SM, Choi S, Lee HK (2018) Median filtered image restoration and anti-forensics using adversarial networks. *IEEE Signal Process Lett* 25(2):278–282
62. Kingma DP, Salimans T, Jozefowicz R, Chen X, Sutskever I, Welling M (2016) Improving variational inference with inverse autoregressive flow. In Advances in Neural Information Processing Systems. arXiv: 1606.04934
63. Kodali N, Abernethy J, Hays J, Kira Z (2017) On convergence and stability of GANs. arXiv preprint arXiv:1705.07215
64. Krizhevsky A, Hinton G (2010) Convolutional deep belief networks on cifar-10. Unpublished manuscript 40(7)
65. LeCun Y (1998) The MNIST database of handwritten digits. <https://yann.lecun.com/exdb/mnist/>
66. Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, Shi W (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
67. Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, (2017) Photo-realistic single image superresolution using a generative adversarial network. *CVPR*
68. Lee HJ, Kim ST, Lee H, Ro YM (2019) Lightweight and effective facial landmark detection using adversarial learning with face geometric map generative network. *IEEE Trans Circuits Syst Video Technol*
69. Lehmann EL, Romano JP (2006) Testing statistical hypotheses. Springer, Science and Business Media
70. Li Y, Shen L (2018) cC-GAN: a robust transfer-learning framework for HEP-2 specimen image segmentation. *IEEE Access* 6:14048–14058
71. Li J, Liu S, He H, Li L (2018) A novel framework for gear safety factor prediction. *IEEE Trans Industrial Informatics*
72. Li H, Li G, Lin L, Yu H, Yu Y (2018) Context-aware semantic inpainting. *IEEE transactions on cybernetics*
73. Li J, Skinner KA, Eustice RM, Matthew JR (2018) WaterGAN: unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics and Automation Letters* 3(1): 387–394
74. Li J, He H, Li L, Chen G (2019) A novel generative model with bounded-GAN for reliability classification of gear safety. *IEEE Trans Ind Electron* 66(11):8772–8781

75. Li P, Prieto L, Mery D, Flynn PJ (2019) On low-resolution face recognition in the wild: comparisons and new techniques. *IEEE Trans Inform Forensics Secur* 14:2000–2012
76. Liang X, Lee L, Dai W, Xing EP (2017) Dual motion GAN for future-flow embedded video prediction. In proceedings of the IEEE international conference on computer vision, pp 1744–1752
77. Liao K, Lin C, Zhao Y, Gabbouj M (2020) DR-GAN: automatic radial distortion rectification using conditional GAN in real-time. *IEEE Trans Circuits Syst Video Technol* 30(3):725–733
78. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Zitnick CL (2014) Microsoft coco: common objects in context. In: European conference on computer vision. Springer, Cham, pp 740–755
79. Lin D, Fu K, Wang Y, Xu G, Sun X (2017) MARTA GANs: unsupervised representation learning for remote sensing image classification. *IEEE Geosci Remote Sens Lett* 14(11):2092–2096
80. Liu Z, Luo P, Wang X, Tang X (2018) Large-scale celebfaces attributes (celeba) dataset
81. Lopez-Tapia S, Lucas A, Molina R, Katsaggelos A K (2018) A single video super-resolution GAN for multiple Downsampling operators based on pseudo-inverse image formation models
82. Lopez-Tapia S, Lucas A, Molina R, Katsaggelos A K (2019) GAN-based video super-resolution with direct regularized inversion of the low-resolution formation model. ICIP conference
83. Lucas A, Lopez-Tapiad S, Molinae R, Katsaggelos AK (2019) Generative adversarial networks and perceptual losses for video super-resolution. *IEEE Trans Image Process* 28:3312–3327
84. Lucas A, Lopez-Tapia S, Molina R, Katsaggelos AK (2019) Generative adversarial networks and perceptual losses for video super-resolution. *IEEE Trans Image Process* 28(7):3312–3327
85. Lucic M, Kurach K, Michalski M, Gelly S, Bousquet O (2018) Are GANs created equal? A large-scale study. In *Advances in neural information processing systems*, pp 698–707
86. Lucic M, Kurach K, Michalski M, Bousquet O, Gelly S (2018) Are GANs created equal? A large-scale study. *NeurIPS* 1–10
87. Ma S, Fu J, Wen Chen C, Mei T (2018) DA-GAN: Instance-level image translation by deep attention Generative Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 5657–5666
88. Ma D, Tang P, Zhao L (2019) SiftingGAN: generating and sifting labeled samples to improve the remote sensing image scene classification baseline in vitro. *IEEE Geosci Remote Sens Lett* 16:1046–1050
89. 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
90. Mao X, Li Q, Xie H, KLau RY, Wang Z, Smolley SP (2018) On the effectiveness of least squares generative adversarial networks. *IEEE Trans Pattern Anal Mach Intell* 41:2947–2960. <https://doi.org/10.1109/TPAMI.2018.2872043>
91. Mardani M, Gong E, Cheng JY, Vasanawala SS, Zaharchuk G, Xing L, Pauly JM (2019) Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Trans Med Imaging* 38(1):167–179
92. Mathieu M, Couprie C, Le Y (2016) Deep multi-scale video prediction beyond mean square error. *ICLR*: 1–14
93. Mirza M, Osindero S (2014) Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*
94. Miyato T, Kataoka T, Koyama Y, Yoshida Y (2018) Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*
95. Nie D, Trullo R, Lian J, Wang L, Petitjean C, Ruan S, Shen D (2018) Medical image synthesis with deep convolutional adversarial networks. *IEEE Trans Biomed Eng* 65(12):2720–2730
96. Nilsback ME and Zisserman A (2008). Automated flower classification over a large number of classes. In 2008 Sixth Indian Conference on Computer Vision Graphics and Image Processing, pp 722–729, IEEE
97. Niu X, Gong M, Zhan T, Yang Y (2019) A conditional adversarial network for change detection in heterogeneous images. *IEEE Geosci Remote Sens Lett* 16(1):45–49
98. Odena A, Olah C, Shlens J (2017) Conditional image synthesis with auxiliary classifier GANs. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp 2642–2651, JMLRorg
99. Ohnishi K, Yamamoto S, Ushiku Y, Harada T (2018) Hierarchical video generation from orthogonal information: optical flow and texture. In *Thirty-Second AAAI Conference on Artificial Intelligence*
100. Oliveira DA, Ferreira RS, Silva R, Brazil EV (2018) Interpolating seismic data with conditional generative adversarial networks. *IEEE Geosci Remote Sens Lett* 99:1–5
101. Oord AVD, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kavukcuoglu K (2016) Wavenet: a generative model for raw audio. *arXiv preprint arXiv:1609.03499*
102. Ouyang X, Cheng Y, Jiang Y, Li C L, Zhou P (2018) Pedestrian-synthesis-GAN: generating pedestrian data in real scene and beyond. *arXiv* 1–22
103. Pan Y, Qiu Z, Yao T, Li H, Mei T (2017) To create what you tell: generating videos from captions. In proceedings of the 25th ACM international conference on multimedia, pp 1789–1798. ACM

104. Pang Y, Xie J, Li X (2018) Visual haze removal by a unified generative adversarial network. *IEEE Trans Circuits Syst Video Technol*
105. Pascual S, Bonafonte A, Serrà J (2017) SEGAN: speech enhancement generative adversarial network. *arXiv preprint arXiv:1703.09452*
106. Quan TM, Nguyen-Duc T, Jeong WK (2018) Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans Med Imaging* 37(6):1488–1497
107. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*
108. Rezende DJ, Mohamed S (2015) Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*
109. Richardson E, Weiss Y (2018) On GANs and gmms. In *Advances in Neural Information Processing Systems*, pp 5852–5863
110. Royer A, Bousmalis K, Gouws S, Bertsch F, Mosseri I, Cole F, Murphy K (2017) XGAN: unsupervised image-to-image translation for many-to-many mappings. *arXiv preprint arXiv:1711.05139*
111. Saito Y, Takamichi S, Saruwatari H (2018) Statistical parametric speech synthesis incorporating generative adversarial networks. *IEEE/ACM Trans Audio Speech Language Process* 26(1):84–96
112. Sakkos D, Ho ES, Shum HP (2019) Illumination-aware multi-task GANs for foreground segmentation. *IEEE Access, Illumination-Aware Multi-Task GANs for Foreground Segmentation*
113. 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 2234–2242
114. Shan H, Zhang Y, Yang Q, Kruger U, Kalra MK, Sun L, Wang G (2018) 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network. *IEEE Trans Med Imaging* 37(6):1522–1534
115. Shen Z, Sheng W, Xu L T, Kautz J, Yang M H (2018) Deep Semantic Face Deblurring, *CVPR*
116. Shen Z, Wang W, Lu X, Shen J, Ling H, Xu T, Shao L (2019) Human-Aware Motion Deblurring, *ICCV*
117. Shi Y, Li Q, Zhu XX (2018) Building footprint generation using improved generative adversarial networks. *IEEE Geosci Remote Sens Lett*
118. Snell J, Ridgeway K, Liao R, Roads BD, Mozer MC, Zemel RS (2017) Learning to generate images with perceptual similarity metrics. In *2017 IEEE International Conference on Image Processing*, pp 4277–4281. *IEEE*
119. Theis L, Oord AVD, Bethge M (2015) A note on the evaluation of generative models. *arXiv preprint arXiv:1511.01844*
120. Tuan YL, Lee HY (2019) Improving conditional sequence generative adversarial networks by stepwise evaluation. *IEEE/ACM Trans Audio Speech Language Process* 27(4):788–798
121. Tulyakov S, Liu MY, 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
122. Van Horn G, Branson S, Farrell R, Haber S, Bary J, Ipeirotis P, Belongie S (2015) Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 595–604
123. Van Horn G, Mac Aodha O, Song Y, Shepard A, Adam H, Perona P, Belongie S (2017) The inaturalist challenge 2017 dataset. *arXiv preprint arXiv:1707.06642* 1(2)
124. Vondrick C, Pirsaviash H, Torralba A (2016) Generating videos with scene dynamics. In *Advances In Neural Information Processing Systems*, pp 613–621
125. Walker J, Marino K, Gupta A, Hebert M (2017) The pose knows: video forecasting by generating pose futures. In *proceedings of the IEEE international conference on computer vision*, pp 3332–3341
126. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612
127. Wang Y, Zhang L, van de Weijer J (2016) Ensembles of generative adversarial networks. *arXiv preprint arXiv:1612.00991*
128. Wang W, Huang Q, You S, Yang C, Neumann U (2017) Shape inpainting using 3d generative adversarial network and recurrent convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pp 2298–2306
129. Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Loy CC (2018) EsrGAN: enhanced super-resolution generative adversarial networks. In: *European Conference on Computer Vision*. Springer, Cham, pp 63–79
130. Wang Y, Zhou L, Yu B, Wang L, Zu C, Lalush DS, Shen D (2018) 3D auto-context-based locality adaptive multi-modality GANs for PET synthesis. *IEEE Trans Med Imaging*
131. Wang Y, Zhou L, Yu B, Wang L, Zu C, Lalush DS, Shen D (2018) 3D auto-context-based locality adaptive multi-modality GANs for PET synthesis. *IEEE Trans Med Imaging*

132. Wang Z, Chen Z, Wu F (2018) Thermal to visible facial image translation using generative adversarial networks. *IEEE Signal Process Lett* 25(8):1161–1165
133. Welinder P, Branson S, Mita T, Wah C, Schroff F, Belongie S, Perona P (2010) Caltech-UCSD birds 200
134. Wen S, Liu W, Yang Y, Huang T, Zeng Z (2018) Generating realistic videos from keyframes with concatenated GANs. *IEEE Trans Circuits Syst Video Technol*
135. Wolterink JM, Leiner T, Vieregger MA, Išgum I (2017) Generative Adversarial Networks for noise reduction in low-dose CT. *IEEE Transactions on Medical Imaging* 36(12):2536–2545
136. Wu W, Qi H, Rong Z, Liu L, Su H (2018) Scribble-supervised segmentation of aerial building footprints using adversarial learning. *IEEE Access* 6:58898–58911
137. Xiang S, Li H (2017) On the effects of batch and weight normalization in generative adversarial networks. *arXiv preprint arXiv:1704.03971*
138. Xiao H, Rasul K, Vollgraf R (2017) Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*
139. Xiong W, Luo W, Ma L, Liu W, Luo J (2018) Learning to generate time-lapse videos using multi-stage dynamic Generative Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 2364–2373
140. Xu C, Ren J, Zhang D, Zhang Y, Qin Z, Ren K (2019) GANobfuscator: mitigating information leakage under GAN via differential privacy. *IEEE Trans Inform Forensics Secur* 14:2358–2371
141. Xuan Q, Chen Z, Liu Y, Huang H, Bao G, Zhang D (2018) Multi-view generative adversarial network and its application in pearl classification. *IEEE Trans Ind Electron* 66(10):8244–8252
142. Yang J, Kannan A, Batra D, Parikh D (2017) Lr-GAN: layered recursive generative adversarial networks for image generation. *arXiv preprint arXiv:1703.01560*
143. Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, Firmin D (2018) DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging* 37(6):1310–1321
144. Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Wang G (2018) Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans Med Imaging* 37(6):1348–1357
145. Yang Y, Zhou J, Ai J, Bin Y, Hanjalic A, Shen HT, Ji Y (2018) Video captioning by adversarial lstm. *IEEE Trans Image Process* 27(11):5600–5611
146. Yi Z, Zhang H, Tan P, Gong M (2017) DualGAN: unsupervised dual learning for image-to-image translation. In *proceedings of the IEEE international conference on computer vision*, pp 2849–2857
147. Yu F, Seff A, Zhang Y, Song S, Funkhouser T, Xiao J (2015) LSUN: construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*
148. Yu J, Lin Z, Yang J, Shen X, Lu X, Huang TS (2018) Free-form image inpainting with gated convolution. *arXiv preprint arXiv:1806.03589*
149. 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
150. Yu B, Zhou L, Wang L, Shi Y, Fripp J, Bourgeat P (2019) Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis. *IEEE Transactions on Medical Imaging*
151. Yuan Y, Tian C, Lu X (2018) Auxiliary Loss Multimodal GRU Model in Audio-Visual Speech Recognition. *IEEE Access* 6:5573–5583
152. Zeng Y, Lu H, Borji A (2017) Statistics of deep generated images. *arXiv preprint arXiv:1708.02688*
153. Zhan Y, Hu D, Wang Y, Yu X (2018) Semisupervised hyperspectral image classification based on generative adversarial networks. *IEEE Geosci Remote Sens Lett* 15(2):212–216
154. Zhang L, Liu P, Gulla (2019) *J Artificial Intell Mach Learning* <https://doi.org/10.1007/s10994-018-05777-9>
155. Zhang H, Xu T, Li H, Zhang S, Wang X, Huang X, Metaxas DN (2017) StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks. In *proceedings of the IEEE international conference on computer vision*, pp 5907–5915
156. Zhang H, Xu T, Li H, Zhang S, Wang X, Huang X, Metaxas D (2017) StackGAN++: realistic image synthesis with stacked generative adversarial networks. *arXiv preprint arXiv:1710.10916*
157. Zhang M, Gong M, Mao Y, Li J, Wu Y (2018) Unsupervised feature extraction in Hyperspectral images based on Wasserstein generative adversarial network. *IEEE Trans Geosci Remote Sens*
158. Zhang H, Goodfellow I, Metaxas D, Odena A (2018) Self-attention generative adversarial networks. *arXiv preprint arXiv:1805.08318*
159. Zhang Z, Song Y, Qi H (2018) Decoupled learning for conditional Adversarial Networks. In *2018 IEEE Winter Conference on Applications of Computer Vision*, pp 700–708. IEEE

160. Zhang K, Luo W, Zhong Y, Ma L, Liu W, Li H (2019) Adversarial spatio-temporal learning for video deblurring. *IEEE Trans Image Process* 28(1):291–301
161. Zhao J, Mathieu M, LeCun Y (2016) Energy-based generative adversarial network. arXiv preprint arXiv:1609.03126
162. Zhou Z, Cai H, Rong S, Song Y, Ren K, Zhang W, Wang J (2017) Activation maximization generative adversarial nets. arXiv preprint arXiv:1703.02000
163. Zhu JY, Zhang R, Pathak D, Darrell T, Efros AA, Wang O, Shechtman E (2017) Toward multimodal image-to-image translation. In *Advances in Neural Information Processing Systems*, pp 465–476
164. Zhu JY, 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
165. Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. *ICCV*
166. Zhu L, Chen Y, Ghamisi P, Benediktsson JA (2018) Generative adversarial networks for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 56(9):5046–5063
167. Zhu X, Zhang L, Zhang L, Liu X, Shen Y, Zhao S (2020) GAN-based image super-resolution with a novel quality loss. *Math Probl Eng* 2020:1–12. <https://doi.org/10.1155/2020/5217429>

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