



# Generative adversarial networks: a survey on applications and challenges

M. R. Pavan Kumar<sup>1</sup> · Prabhu Jayagopal<sup>2</sup>

Received: 4 August 2020 / Revised: 21 September 2020 / Accepted: 2 October 2020 / Published online: 24 October 2020  
© Springer-Verlag London Ltd., part of Springer Nature 2020

## Abstract

Deep neural networks have attained great success in handling high dimensional data, especially images. However, generating naturalistic images containing ginormous subjects for different tasks like image classification, segmentation, object detection, reconstruction, etc., is continued to be a difficult task. Generative modelling has the potential to learn any kind of data distribution in an unsupervised manner. Variational autoencoder (VAE), autoregressive models, and generative adversarial network (GAN) are the popular generative modelling approaches that generate data distributions. Among these, GANs have gained much attention from the research community in recent years in terms of generating quality images and data augmentation. In this context, we collected research articles that employed GANs for solving various tasks from popular databases and summarized them based on their application. The main objective of this article is to present the nuts and bolts of GANs, state-of-the-art related work and its applications, evaluation metrics, challenges involved in training GANs, and benchmark datasets that would benefit naive and enthusiastic researchers who are interested in working on GANs.

**Keywords** Generative model · Convolutional neural network · segmentation · Object detection · Generative adversarial network

## 1 Introduction

Generating quality images is a challenging task in the field of computer vision and artificial intelligence, having numerous applications and research scope. Supervised machine learning and deep learning models require large and labelled datasets to generalize the decision making process. However, the availability of large and labelled databases is questioned in many domains like medical diagnosis, fault detection, intrusion detection, etc. Hence, the research community heavily depends on unsupervised learning. In unsupervised learning, the model strives to learn the structure and extracts the useful features of the data. Generative modelling is a sub-field of unsupervised learning that work towards the goal of

learning the structure of the data and generate data similar to it. Generative models can be trained with high dimensional probability distributions. They can also be used in reinforcement learning, semi-supervised learning, etc. In general, generative models work on any one of these three principles: inference approximation, maximum likelihood, and Markov chains. Latent Dirichlet distribution [7], restricted Boltzmann machines [36], deep belief networks (DBNs) [37], etc., are the other generative models extensively used in the literature to generate naturalistic data. These models operate on the principle of maximum likelihood. However, these models do not fit the data distributions completely.

Goodfellow et al. [23] introduced GANs, an unsupervised generative model, worked on the principle of maximum likelihood, and used adversarial training. Right from the inception of generative adversarial networks (GANs), they have been the most discussed and most researched domains not only in the field of computer science but also in other domains. GANs have been gained much popularity in generating high-quality realistic data. Thus, GANs also attracted researchers as a data augmentation tool for imbalanced data applications. Generative models, especially GANs, can be deployed in many machine learning tasks where multiple cor-

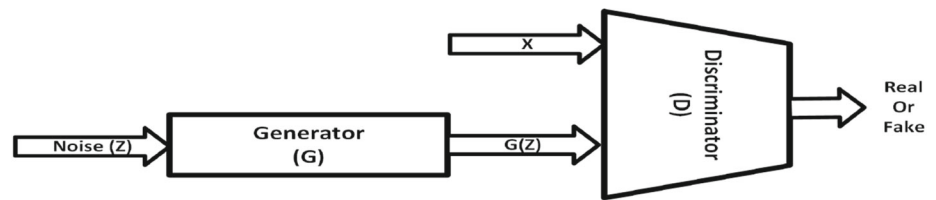
✉ Prabhu Jayagopal  
jprabhuit@gmail.com

M. R. Pavan Kumar  
marabattinasivapavan@gmail.com

<sup>1</sup> School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

<sup>2</sup> School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

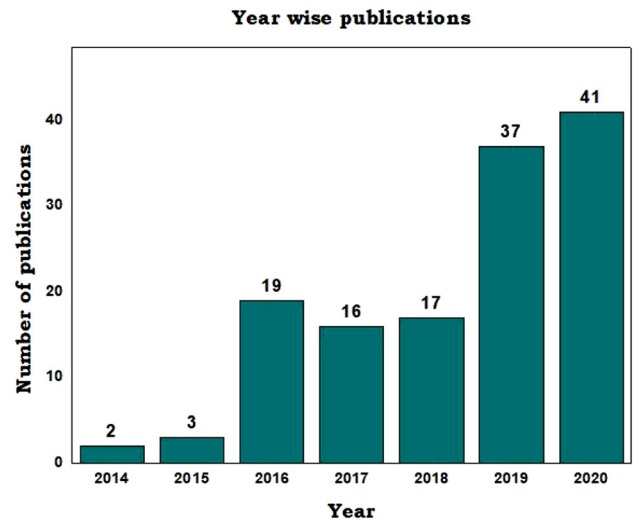
**Fig. 1** The architecture of GAN



rect answers can be inferred from a single input. GANs are also used to attribute more information into the context than it has originally. GANs are used to create models that help researchers in generating artificial naturalistic images. Since then, they have been used in diverse domains and diverse applications. GANs have been leveraged in medical imaging for disease diagnosis, semantic segmentation, image captioning, image attacking to change the classifier decision, image deblurring, image dehazing, synthesizing, face frontalization, high-resolution images from low-resolution images also known as super-resolution (SR), text to image generation or scene generation, steganography, object detection, speech recognition, fault diagnosis, industrial risk analysis, and natural language processing applications like text generation, text summarization, style transfer, etc. However, one should note that GAN is not just an image generation tool, but it retrieves useful information from the training data so that object detection, segmentation, and classification tasks can be performed in various domains. The training data can be any content of multimedia, for example, image, text, audio, video, and animation. The GAN consists of a generator network (G) and a discriminator network (D) as shown in Fig. 1. The generator consumes a noise vector ( $z$ ) as input and generates data distribution similar to real data distribution. The discriminator discriminates between real and artificial data as a binary classifier. During the training, the generator loss and the discriminator loss are computed to compute the overall loss ( $V$ ). The GAN loss is computed using the formula shown in Eq. 1, and the motive of training is to minimize the generator loss and maximize the discriminator loss.  $x \sim P_{data}(x)$  and  $z \sim P_z(Z)$  represents  $x$  is an instance from real distribution and  $z$  is an instance from prior distribution.

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

The motive of this paper is to give a brief introduction about GANs, nuts and bolts of GANs, various derived GANs, and applications of GANs pertaining to different tasks in multiple domains. To this end, we collected 140 research papers to give a detailed summary of generative adversarial networks, especially in terms of applications. Figure 2 shows the graph of the number of publications considered year wise. First, we identified the state-of-the-art works derived relat-



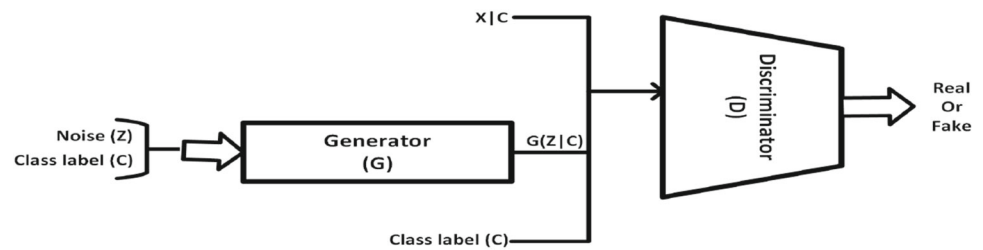
**Fig. 2** Year wise number of applications

ing to different tasks are discussed in Sect. 2. The collected papers were segregated based on their objective and classified into multiple applications, as presented in Sect. 3. Section 4 discusses various evaluation metrics used in the selected papers to evaluate the GAN model. We discussed multiple challenges involved in training a GAN in Sect. 5. Finally, the paper concludes in Sect. 6.

## 2 The beginning

With the advent of GANs that work on the principle of generative and adversarial manner, researchers can synthesize novel and quality images. However, the synthesizing of images has not just started with GANs. There are a few research works to synthesize images using convolutional neural networks (CNN). Abdalmageed et al. [1] discussed a CNN-based pipelined architecture for face recognition task by detecting and correcting multiple poses using deep face representation methods. Masi et al. [69] synthesized more facial images using a deep convolutional model to make the dataset better. The existing facial images are manipulated in three variations pose, shape, and expression. The pose and shape are simulated across three dimensions with closed mouth expression. But, these are subjected to limitations in terms of quality and diversity of images, thereby deterio-

**Fig. 3** The architecture of conditional GAN



rating the performance generalization indirectly. Once the principle of GAN is formulated and achieved success in modelling probability distributions, plenty of other GAN models are derived for diverse applications and produced promising results. Mirza and Osindero [71] introduced class label as an additional input to both generator and discriminator to model conditioned variant of GAN (CGAN). Since the class label is given as other input, the CGAN is capable of generating images specific to the class label. The architecture of CGAN is shown in Fig. 3, and loss function is shown in Eq. 2. Radford et al. [80] devised a GAN model named DCGAN for learning unsupervised representations. Contrary to basic GAN, DCGAN has convolutional layers to upscale the input vector  $z$  of the generator module. Also, it employed regular convolutional layers to classify generated and real images.

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x|c)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z|c)))] \quad (2)$$

Wu et al. [107] modelled the generation of 3D objects by mapping latent space to object space using 3D GAN. They extended the 3D GAN using a variational autoencoder (3D-VAEGAN) that maps a 2D vector space to 3D vector space using the VAE module and then mapped to 3D object space by the GAN model. Liu and Tuzel [66] computed the joint probability distributions of two domains using two GANs. Each GAN learns the distribution of one domain while training and also shares the high-level weights that allow coupled GAN to compute the joint distribution. By deploying multiple generators, kwak and Zhang [53] generated a composite generative system called CGAN. First, every generator separately removes a complex part of the image. These components are then summarized by a blending process to generate a new image. Im et al. [41] demonstrated a novel image generation method based on the recursive adversarial model (GRAN). GRAN incrementally generates high fidelity visual samples. A novel crossover evaluation scheme is also introduced between generator and discriminator networks. Zhu et al. [138] introduced an immersive generative model that allows users to control the visual content naturally and realistically. Perarnau et al. [77] performed image editing operations like expression changing,

hair colour changing, gender-changing, etc., using invertible conditional GAN (IcGAN). An encoder is placed to compress the input vector into a latent and conditional vector. Yoo et al. [118] discussed a framework that transfers one domain to another semantically in pixel-level. The framework contains an encoder-decoder based generator and two discriminators. The two discriminators are trained to learn the semantic relations between the domains. Brock et al. [9] introduced a neural photo editor that affects the semantic changes requested by the user at ease. The intuitive idea behind neural photo editor is that it back propagates the requested changes to compute the change in latent parameters. Given two unlabelled and unrelated domains  $P$  and  $Q$ , and a function  $f$ , a generative function  $G$  can be learned such that a sample from  $P$  can be mapped to  $Q$ , i.e.  $G: P \rightarrow Q$  and  $f(x) \sim G(f(x))$  [93].

Arjovsky et al. [5] introduced a novel training model (WGAN) based on Wasserstein's distance to avoid the mode collapse problem that occurs in the training of traditional GANs. Antipov et al. [3] introduced Age-cGAN that generates high-quality synthetic images in which an aged face synthesized while preserving the person's identity. Kim et al. [50] learned the relations between two different domains using DiscoGAN that consist of two GANs coupled together. Given an image in one domain DiscoGAN, can generate the corresponding image in another domain. Li et al. [60] have developed a GAN for detecting smaller image entities by reducing smaller objects representative margin to larger objects. In [43], the authors have employed a conditional GAN model to translate an image to another analogous image. For example, converting day-photo to night-photo, converting to an aerial image to map design, etc. Karras et al. [46] presented progressive growing GANs that generate quality and high-resolution images. The idea behind progressive growing GAN is that it extends the training process of a normal GAN by adding new layers. A super-resolution GAN (SRGAN) [55] model takes a low-resolution image as input and increases the spatial resolution of the image by an upscaling factor and produces a high-resolution image as output. The applications are ranging from satellite imaging, medical imaging, media content, face recognition in surveillance systems, etc. A StackGAN [126] that automates synthesizing realistic images from human-written descriptions. The StackGAN works in two stages. Stage-1 GAN generates

low-resolution images with initial shape and basic colours of objects. Stage-2 GAN takes low-resolution images, text descriptions as inputs and corrects the errors, completes the details, and generates photo-realistic images with four times better resolution. AnoGAN [87] an unsupervised generative model that can detect diseases from medical image data at early stages. Zhu et al. [139] explored a generative model called Cycle GAN that translates images from one domain to another domain. For example, take an image and creates an image that looks like a painting of the first picture, converting a black-and-white picture to a colour image, etc. Yang et al. [113] integrated the conventional acoustic loss function with the discriminator loss function to model a multi-tasking framework for text to speech synthesis.

Choi et al. [15] developed StarGAN that can translate an image among multiple domains with superior quality. Grover et al. [24] developed Flow-GAN that uses exact likelihood estimation for training and achieved significant improvements in log-likelihood scores. Wang et al. [102] introduced a Residual-in-Residual Dense Block to the SRGAN [55] to model enhanced SRGAN (ESRGAN). The model achieved high-quality images with more natural and realistic textures. Kupyn et al. [52] developed a conditional generative model called DeblurGAN that deblurs a blurred image and detects the objects that are blurred due to the motion. They used content loss function to optimize conditional GAN. A conditional generative adversarial framework is designed for synthesizing  $2048 \times 1024$  high-resolution photo naturalistic images using semantic label maps [101]. Xu et al. [110] proposed an AttnGAN that consists of attention models to generate quality images from text descriptions. The authors have incorporated an attention module as a generator network in which each attention model create sub-regions in the image based on the extracted features from the text. Zhang et al. [128] modelled a self-attention GAN (SAGAN) that generates details using attention-driven and long-term dependency modelling. The authors have also applied spectral normalization to enhance the dynamics of training and achieved significant results. Therefore, researchers have derived plenty of GAN variants like CGAN, WGAN, ProgressiveGAN, image-to-image translation GAN, Cycle GAN, SR GAN, text-to-image GAN, face inpainting GAN, text-to-speech GAN, etc., for various applications. The evolution of a few popular GANs is illustrated in Fig. 4 with the help of the timeline diagram. In the next section, we discuss the applications specific to these variants that were modelled recently in detail.

## 3 Applications

In this section, we discuss diverse applications of GANs like medical diagnosis, text generation, hyperspectral image classification, etc., in detail.

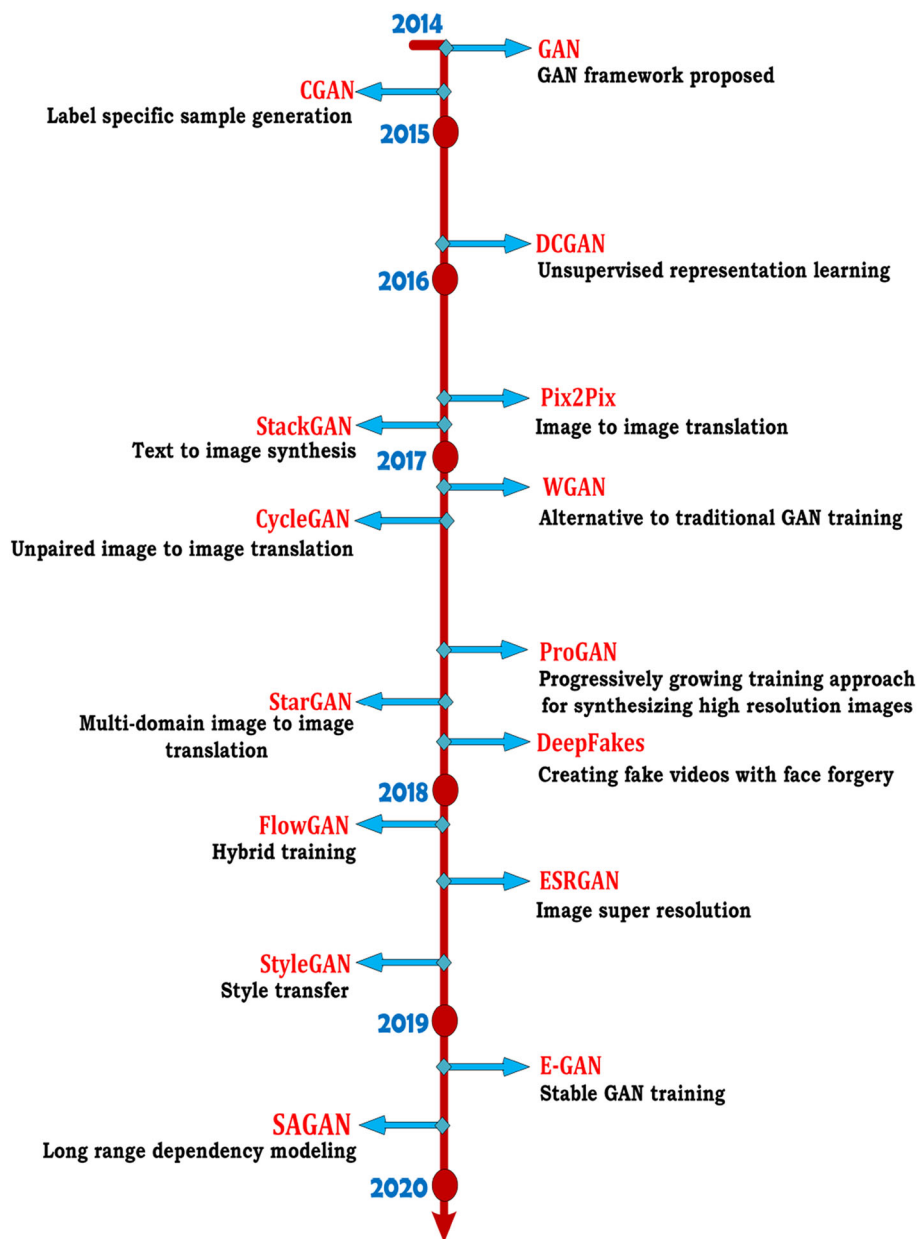
### 3.1 Clinical diagnosis

MRI (magnetic resonance imaging), CT (computed tomography) scan, PET (positron emission tomography), ultrasound imaging (USI), electrocardiogram (ECG), and X-rays are the widely used imaging techniques for clinical diagnosis and to identify the severity of the disease in the medical domain. MRI images the water molecules in the body with the help of a very strong magnetic field. MRI images take the pictures of the soft tissue of the organs and the bones. CT scanners use a pencil-thick beam to take cross-sectional images of the patient's body. The beam rotates around the patient's body. The CT scan slices the patients' body like a loaf of bread and uses radiation to take the images. PET scanner captures the images of minuscule changes in the body's metabolism caused by the growth of abnormal cells. The PET scan can be used in combination with a CT scan that allows physicians to identify the exact location, size, and shape of the diseased tissue or tumour.

In [21], a cycleGAN-based unified framework is discussed to standardize the intensity distribution of MRI images with different parameters coming from multiple groups. The framework consists of two kinds of paths: one forward path using a one GAN and multiple backward paths using multiple GANs. They also employed two jump connections to keep the features safe and to avoid any loss of resolution. The effectiveness of the proposed method is investigated on T2-FLAIR image datasets. Qi et al. [78] developed a model using cascaded conditional GAN (C-cGANs) for automatic bi-ventricle segmentation of magnetic resonance images of the heart. The authors have divided the task of segmentation into two subtasks. For each subtask, they used a specific C-cGAN. In both the C-cGANs they used an encoder module, MSAF module and a decoder module. The first C-cGAN identifies the region of interest using the binary segmentation task. The second C-cGAN implements the bi-ventricle segmentation task.

Analyzing ECG signals enables diagnosing cardiovascular diseases (CVDs) or heart-related diseases in advance and helps to prevent them. Detecting abnormalities in ECG signals is a class imbalance problem due to imbalance distribution among multiple classes. Wang et al. [99] proposed a framework in which a classification model is incorporated in between a GAN model. The generator and discriminator framework is inspired by the ACGAN model [72] to support data augmentation. The classification model is implemented using a residual block and a long short-term model

**Fig. 4** Timeline for a few notable GAN models



(LSTM). The proposed framework is tested on MIT-BIH standard database for single beat detection and competition database for successive beats detection. In [117], the authors have addressed two issues while dealing with clinical data for cardiac disease diagnosis using ECG signals. First, they extracted the global features and then increased the stability of training to extract high-quality diverse samples. The authors have developed a sequential GAN (RPSeqGAN) in which the generator consists of bidirectional gated recurrent units (GRU), and the discriminator is implemented as a ResNet [32]-based ResNet-Pooling block (RPblock) that extracts the global features. Authors have also employed a policy gradient and Monte Carlo search algorithms to gain stability in training. The proposed algorithm achieved qual-

ity images and maximum stability on MIT-BIH arrhythmia dataset. In [88], the authors used a GAN model as a data augmentation tool to generate synthetic data to tackle the imbalanced classification of multi-class ECG data. Arrhythmias MIT-BIH data has 15 ECG classes that can be divided into five (N, S, V, F, Q) categories. Authors have proposed two deep learning models for classification task on data-augmented original data. First, a CNN-based end-to-end approach is used to classify the heartbeat as one among the 15 classes. Second, another CNN-based hierarchical approach has two stages. In the first stage, the model identifies one of the five categories. In the second stage, any one of the five classes is identified under the category identified in the first stage.

Two issues are addressed in [114] while extracting vein features from low-contrast infrared images of fingers. First, the existing CNN-based models increase the processing time when handling low-quality finger vein images. Also, there is a limitation on the size of finger vein images. Second, there is a lack of feature representation about ground truth low-quality finger vein patterns. The authors developed a finger vein GAN (FV-GAN) framework that consists of two generators: an image generator that generates vein images from vein patterns using a UNET architecture and a pattern generator that maps vein images to vein patterns using encoder-decoder network. The discriminator finds the latent space between the correct and wrong vein patterns. The model is evaluated on two publicly available datasets: Tsinghua University Finger Vein and Finger Dorsal Texture Database 2 (THU-FVFD2) and ShanDong University finger vein database (SDU) and achieved significant results. Another useful resource for clinical diagnosis is ultrasound imaging. Ultrasound imaging techniques are profoundly used in the diagnosis of maternal-foetal medicine, abnormality in body parts for example breasts, liver cancer, identification of thyroids, etc. However, capturing ultrasound images requires large infrastructured devices that cannot be easily used in applications like rural medicine, telemedicine, and community medicine applications. Thus, portable ultrasound imaging devices are used too often in the above scenarios to improve global health care. But the low quality of ultrasound images with these portable devices questions the reliability of diagnosis is a limitation. A two-stage GAN structure is devised in [136] to increase the image quality of hand-held or portable ultrasound devices. A U-Net model is placed as a front-end tool for the generator at stage one. It extracts structural features at low frequencies in the reconstructed images. In stage 2, a GAN network is deployed to find the latent space between low-quality images and high-quality images. The generator takes a pair of inputs: a low-quality image and the output image of the U-Net model. The discriminator also takes a pair of inputs: reconstructed generator images and high-quality images. The proposed 2-stage model improved the image-quality of hand-held and portable ultrasound devices.

MRI and PET images are fused in [116] to generate images that have both tissue structure from MRI and functional, metabolic information from PET. The motive behind fusing multiple source images is to get rid of redundant information and to get complementary information in one single image to yield a better clinical diagnosis. The authors have proposed an algorithm based on Wasserstein GAN (MWGAN) to surmount the challenges involved in fusing images from multiple sources. The GAN model consists of one generator and two discriminator networks with a novel loss function. The model can be extended for the fusion of MRI and CT images also. The model is investigated on the publicly available dataset on the Harvard Medical School official page.

Pulmonary nodes in the lungs are examined for the diagnosis of lung cancer at early stages. However, most of the medical domain applications suffer from data scarcity problem. This makes the application of deep learning models on limited data resulting in wrong clinical diagnosis. A GAN-based unsupervised approach is proposed on the principles of anomaly detection for the diagnosis of lung cancer. An encoder module is incorporated along with the GAN model for the training of benign pulmonary nodes. The GAN (MDGAN) consists of a generator network and multiple discriminator networks. The MDGAN computes the feature loss along with image reconstruction loss to assign high scores to malignant nodes and small scores for benign nodes. The performance of the model is evaluated on LIDC-IDRI dataset and proved effective compared to supervised benchmarks [51]. Ghassemi et al. [22] discussed a GAN-based model as a novel data augmentation method for multi-class classification of MR images. First, the GAN is fed with different MR image datasets to generate MR like images as the output of the generator. Later, the data-augmented new dataset is given to the discriminator, which is already trained during data augmentation phase for multi-class classification. The proposed model has achieved significant accuracy rates on MRI dataset compared to state-of-the-arts. Decreasing the dose of radiation during chest imaging adds noise to the generated image, thereby altering clinical diagnosis. Kim et al. [48] devised a conditional GAN (CGAN)-based denoising method that removes the noise in reduced radiation chest images and enhances the image quality for clinical diagnosis. The generator and discriminator of conditioned GAN model are built of convolutional layers. Figure 5 shows the architecture of CGAN and the restored and uncorrupted images. He et al. [34] modelled a label smoothing GAN (LSGAN) for the classification of optical coherence tomography (OCT) images that can help in detecting and avoiding blindness at early stages. The model consists of a generator, discriminator and a classifier. The generator creates synthetic unlabelled OCT images. The discriminator distinguishes between training OCT images and synthetic OCT images while optimizing the performance of the generator to generate high-quality images. A label smoothing strategy is embedded in the classifier that helps in labelling unlabelled OCT images. The LSGAN model is evaluated on UCSD publicly available dataset and a locally developed HUCM dataset and achieved promising results.

From the above discussion, it is observed that image reconstruction, image synthesis (for example conditional image synthesis and cross-modality image synthesis), segmentation, classification, abnormality detection, denoising, data augmentation, etc., are the novel tasks that were solved using GANs.

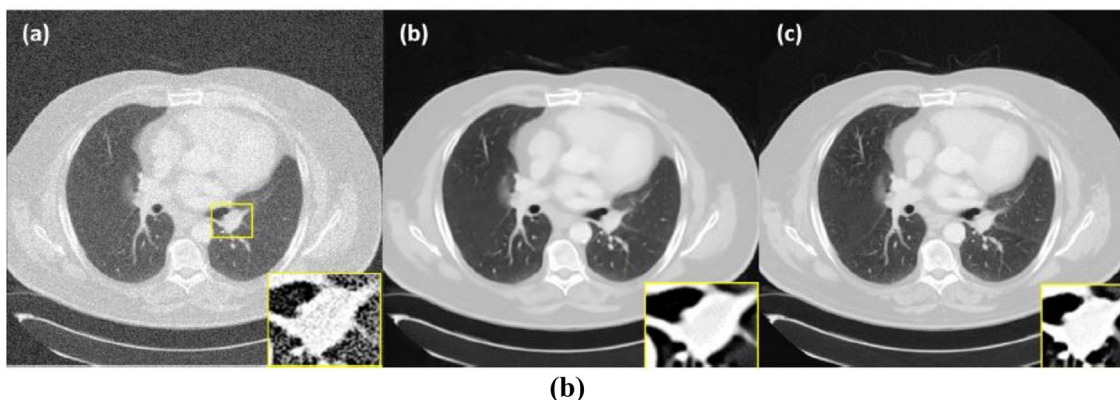
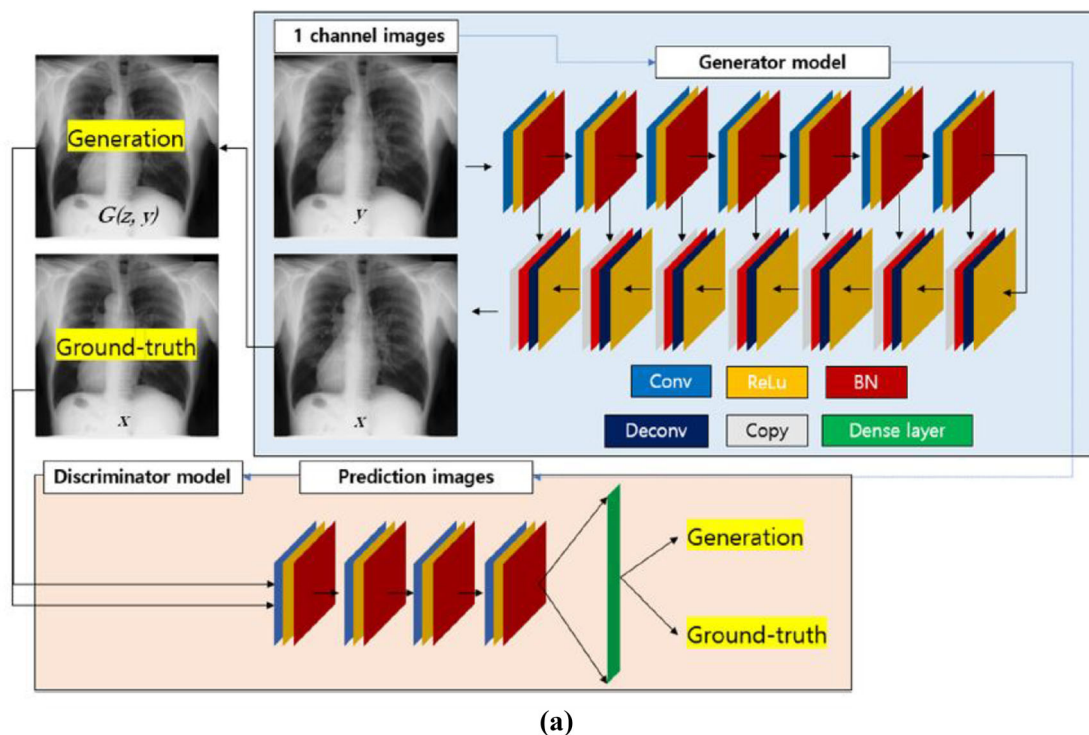


Fig. 5 Images cropped directly from [48] a Architecture of CGAN (b) a Gaussian noise corrupted noise image b CGAN restored image c final uncorrupted image

### 3.2 Intrusion detection

Although the success of machine learning and deep learning in classification, adversarial examples attempt to get, deep learning models to miss-classify the images by inducing small noise patterns. Yuan et al. [124] developed a randomized nonlinear image transformation method to alter and ruin the advanced patterns of attacking noise partly in the adversarial images. They employed a generative cleaning network to retrieve the lost content of the original image during the image transformation phase. The discriminator network is used to defend the classification process and trained not to detect any leftover noise patterns in the images. They evaluated the proposed model using CIFAR-10 and SVHN

datasets. Zhang et al. [130] proposed an extended Monte Carlo tree search (MTCS) algorithm using a GAN model that produces adversarial examples of cross-site scripting (XSS) attack traffic data. They added adversarial examples to an original dataset during the training phase. Also, they assigned a probability value that bypasses the adversarial image from the detector. The model is examined using an intrusion detection (CICIDS2017) dataset that contains up-to-date attacks on real-world data. Huang et al. [40] modelled an imbalance GAN (IGAN) framework to enhance the process of intrusion detection in ad hoc networks. The architecture consists of a feed forward network to extract the features, an IGAN with a filter to synthesize the abnormal class samples and a deep neural network to perform the classification task.

### 3.3 Fault diagnosis

Fault detection is an important task in the field of control engineering to capture the malfunctioning of machine to avoid machine failure and human loss. Shao et al. [89] devised a model for monitoring of machine condition and fault diagnosis using sensor data. The model design is based on ACGAN [72] architecture that consists of 1D convolutional layers. Initially, the model is trained on the limited training data during which it learns hierarchical representations and generates realistic, raw sensor signal data. Later the augmented dataset is used for the classification of a machine fault. They also used a novel quantitative method for the evaluation of generated sensor signal data and used time domain and frequency domain characteristics for assessing the diversity of generated samples. Yan et al. [111] addressed the automated detection and diagnosis (AFDD) of fault training data using an unsupervised framework. However, the number of training instances for normal machine states is higher than faulty machine states. They explored a conditioned version of WGAN deployed to synthesize more training instances of faulty state samples. The authors have deployed a multi-layer perceptron model used as generator and discriminator networks for AFDD. A support vector machine (SVM) is trained as a binary classifier on the augmented dataset. In the detection phase, it identifies the faulty state, and in the diagnosis phase, it classifies the type of fault. Wang et al. [103] showed another GAN-based framework (CVAE-GAN) using the conditional variational autoencoder (CVAE) for imbalanced fault diagnosis in a planetary gear box. The CVAE consists of three modules encoder, decoder, and a sampling network and considered as a generator network. It learns the spectrum distribution features of vibrating signals to generate fault samples at different modes. The discriminator network differentiates true fault sample with generated fault sample and also classifies the variant of fault. Zhang et al. [129] noted a framework that works in two stages for imbalanced fault diagnosis of rotating machines. A GAN model that contains multiple generation modules is placed to generate samples for different fault conditions. A convolution model that ends with fully connected dense layers is placed as a discriminator network. A deep convolutional model is used for the classification task on augmented data. Investigations on CWRU and Bogie datasets proved the effectiveness of the proposed model. In [74], the authors have discussed the semi-supervised and imbalanced fault bearing identification in automation systems of industrial applications. A deconvolutional network is employed as generator and a convolutional network is deployed as discriminator.

### 3.4 Semantic segmentation

Semantic segmentation is one of the tasks from the computer vision domain. Image segmentation divides the image into different sub-parts and classifies each sub-part into a class. In contrast to this, semantic segmentation classifies each pixel of the image to a specific class. Kim et al. [49] proposed a modified generative adversarial model to synthesize the images of jellyfish to avoid jellyfish swarm in fisheries. They employed an auto-encoder in parallel to GAN model. The generator model is used for the synthesis of images. The discriminator takes two inputs: synthesized images from the generator and real images from an autoencoder. The auto-encoder is also used to generate images from the synthesized vectors from the generator. They also estimated the density of jellyfish swarm using full convolutional and regression networks. Wang et al. [100] discussed a model named multi-context GAN (MCGAN) that completes the faces in the images with random missing regions. The model considers the semantic and high frequency features using parallel dilated learning units (DLU). A stack of DLUs is then used to incorporate the fine details using a larger receptive field. The performance of DLUs and the entire model is investigated on CelebA dataset and yielded satisfactory results. In [73] the authors proposed an attentively conditioned GAN (AC-GAN) for semantic segmentation. The generator model is used as a segmentor to generate maps from images. The discriminator model is used to differentiate the segmentor's output from real labels. Also, an attention network is deployed along with segmentor to provide attention probability of each feature map. They investigated the proposed model on PASCAL VOC 2012 and Cam Vid datasets.

The projective imaging nature of X-rays makes it a challenging device to use for clinical diagnosis. The image capturing technique used for X-rays bypasses the 3D spatial information between anatomies. It leads to difficulties in semantic segmentation which in turn deteriorates the clinical diagnosis performance. Also, the large availability of data annotations is not possible in the medical domain. In this context, [131] modelled task-driven GAN (TD-GAN) to perform multi-organ segmentation task. First, synthetic digitally reconstructed radiographs (DRR) are generated from 3D CT images and trained using digital image to image (DI2I) module. Then, the task-driven GAN is deployed that performs image synthesis and segments of multiple organs in an unsupervised manner. Mammography is used extensively in detecting abnormalities in women breasts to diagnose breast cancer in the early stage. Radiologists leverage low energy X-ray signals to find the variance in appearance, location, size, shape, and texture of breasts. Singh et al. [91] modelled a framework (cGAN) that employs a single shot detector [67] to locate the region of interest (ROI) in breast mammograms and surround it by a bounding box. Later, the ROIs are given



as conditioned input to the generator that learns the inherent features like edges, grey-level, gradients, shape, etc., of unhealthy and healthy tissue. It also produces a binary mask (segmentation) based on these features. The discriminator network takes the ground truth and predicted masks as input and indicates the real one. They also used a multi-class CNN network for the classification of irregularities in breast shapes (round, irregular, lobular, and oval). They investigated the model using INbreast, DDSM public datasets and Hospital Sant Joan de Reus private dataset and achieved significant results. Figure 6a shows the workflow of cGAN for breast tumour segmentation and classification. Figure 6b shows the generator network and the discriminator network of cGAN. Figure 6c shows the CNN architecture for classification of type of tumour.

Bisneto et al. [6] employed a conditional GAN model to perform semantic optic disc segmentation for automatic diagnoses of neurodegenerative diseases. The CNN U-net [83] is used as generator and PatchGAN [54] is used as discriminator network. [85] performed segmentation and quantification of tumours simultaneously from CT images for diagnosis of kidney tumours. A residual network that acts as multi-scale feature extractor retrieves the tumour features. A multi-tasking integrated network is used as generator network that performs the semantic segmentation, object detection, and direct quantification. A convolutional model is deployed as a discriminator network to encourage the optimization process. Han et al. [27] presented a GAN-based semi-supervised model for segmentation of lesion in breast ultrasound (BUS) images. It first makes use of annotated images to synthesize more BUS images and thereby enhances the segmentation performance. Delannoy et al. [17] discussed a GAN-based framework (SegSRGAN) that performs super-resolution to increase the image quality and segmentation tasks to segment the region of interest on brain MR images. Lei et al. [57] formulated a new GAN model to differentiate melanoma from a normal skin lesion. The novel GAN contains Unet-SCDC based generator that has skip connections as well as dilated convolutions and produces segmentation masks. It also includes two CNN-based discriminator networks that enhance the generated mask quality. The first CNN takes the concatenation of real input and generated segmented mask as input while the second CNN takes the generated segmented mask alone.

### 3.5 Image to text (I2T) and text to image (T2I) synthesis

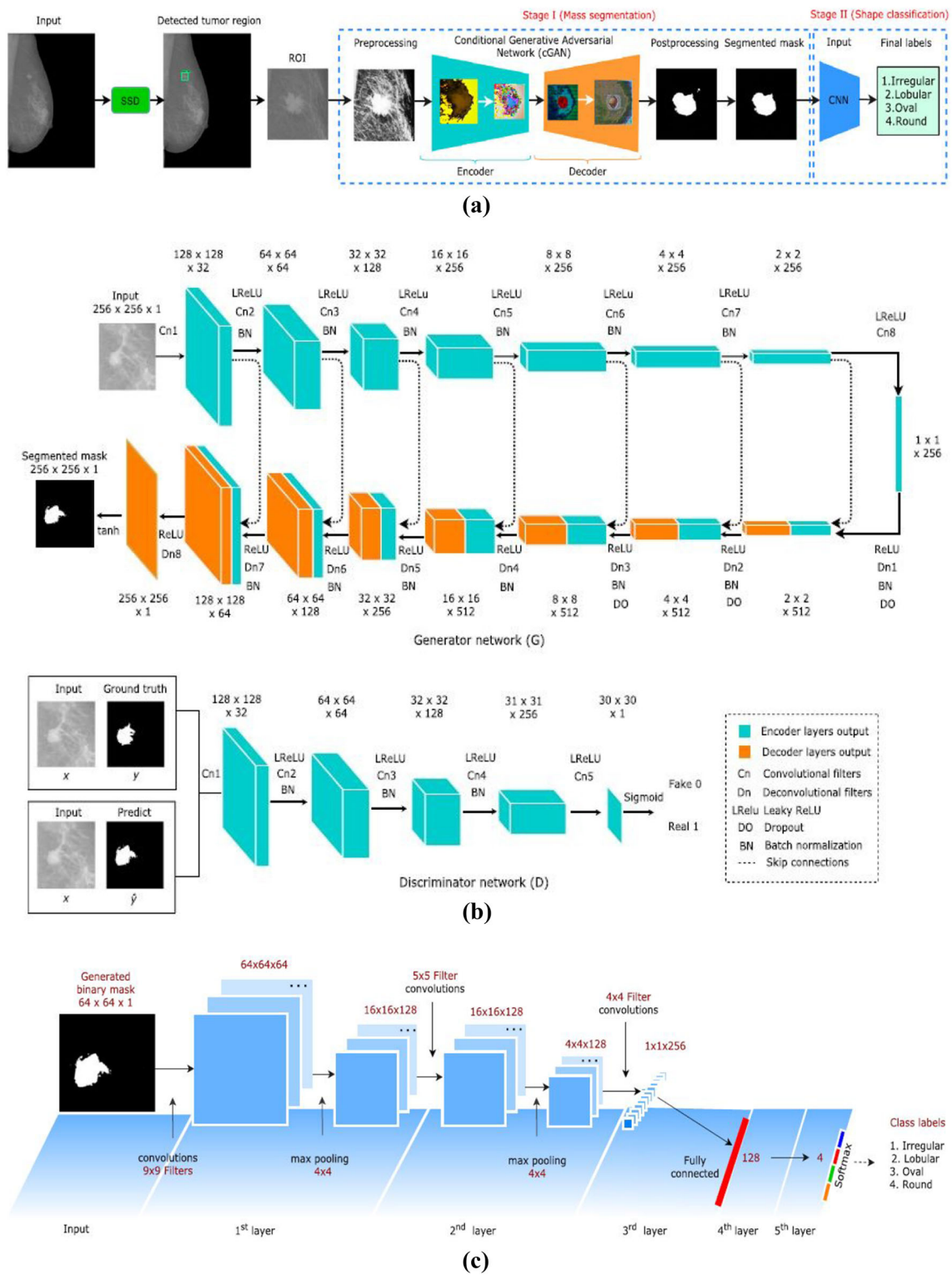
In [127], the authors have proposed two stack models: stack GAN version1 and stack GAN version2 to synthesize images from text. Again stack GAN version1 has two GANs, one in each stage. The stage1 GAN generates low-resolution images from text descriptions. The stage2 GAN generates

high-resolution images from the low-resolution images by considering the missing details of the text and conditioning on the stage1 output. Stack GAN version2 consists of a series of generators and a series of discriminators in a tree-like structure. The stack GAN version2 model is implemented in both conditional and unconditional manner to generate high-resolution naturalistic images. Cai et al. [10] described a Dual attention GAN (DualAttn-GAN) model to generate naturalistic and realistic images from text descriptions. As the name suggests, they incorporated two attention models: textual attention model and visual attention model. The textual attention model is employed to identify the semantics between inputs and outputs. On the other hand, a visual attention model is used to increase the representation power of visual features. They evaluated the model using CUB and Oxford-102 datasets. Contrary to the recognition of general characters as machine-encoded text, extracting text from natural images. It includes challenges from variations in the text shape, colour, size, and patterns. Figure 7 shows images generated by DualAttn-GAN compared to other models on CUB dataset.

It is also difficult to extract text from complex backgrounds with a non-uniform degree of visibility, noise, pollution occlusion, reflections, lightening, and blur. Lei et al. [58] presented a model named defect-restore GAN to extract sequential text from abnormal images of the moving vehicles. The GAN model contains two encoders in the generator, a discriminator, and recurrent neural network (rnn) as an output block. The proposed model is investigated on their proprietary wagon dataset, which has 5000 images and achieved significant results. Yanagi et al. [112] modelled a *Query is GAN* using AttnGAN [110] to extract scenes from the text descriptions. First, three query images are generated using the text description as input to the AttnGAN. Later, the generated query images and a hierarchical structure are used to retrieve the most desired scenes. Ak et al. [2] discussed e-AttnGAN an extension for AttnGAN. The attention module of e-AttnGAN involves contextual features of word and sentences in image generation process. They employed spectral normalization to maintain a stable training process. The e-AttnGAN has proved its effectiveness over the state-of-the-art in image generation.

### 3.6 Natural language processing

Generating text sequences is part of natural language processing tasks. Dialogue systems, machine translation, and writing poetry are also part of the text generation task. Since the inception of GANs, they have been coupled with reinforcement learning to generate text sequences. The output of the discriminator is fed as input to the generator to mimic the reinforcement reward feedback signal. However, this input is a scalar value and cannot maintain the high-level



**Fig. 6** Images cropped directly from [91] **a** Workflow of cGAN **b** cGAN architecture for breast mass segmentation of tumour **c** CNN architecture for shape classification of tumour



**Fig. 7** Image generated by DualAttn-GAN [10] (2nd from right) compared to other models for the given text

semantic information of the text. Also, sampling is performed to complete the text sequences and get a reward signal through the discriminator. The text sequences may contain repeated subjects and missing verbs due to the high randomness of the sampling process [122]. In [115], the authors have addressed these two issues in feature-guiding GAN (FGGAN) to generate text sequences. The reward signal has been replaced by a feature guided vector generated from the features extracted by the discriminator using a feature module. Authors have also created semantic rules to control the next word being generated at each time step preventing words that have low correlation with generated prefix words. Li et al. [61] modelled a dialogue response system using adversarial reinforcement training model. The model is embedded in a reinforcement framework and trained the generator based on the output of the discriminator. The model has generated dialogue sentences that are competitive enough to human-generated sentences. Given the context of the text [63] generated labelled sentences based on category information using category sentence GAN (CS-GAN). To this end, they incorporated RNN to generate sequences, reinforcement learning for predicting next character based on the current state, and GAN for adversarial training and classification. Wang et al. [98] discussed automatic sentimental text generation using SentiGAN framework. The SentiGAN consists of multiple generators generating diverse sentimental texts using a novel penalty based objective function. The discriminator model classifies the high-quality diverse texts to their sentiments. They extended the SentiGAN model, C-SentiGAN to tackle the problem of conditional text generation. The model is evaluated on Movie Reviews, Beer Reviews, customer reviews, and emotional conversations and achieved significant results in terms of the novelty, fluency, intelligence, and diversity of the texts generated.

Rizzo et al. [81] explored the performance of SeqGAN with contextual information encoded in global word embeddings as input. A self-attentive neural network is employed as a discriminator to optimize the SeqGAN performance in embedding knowledge into the generated text. Motivated from the sequence GAN [12,122] discussed a conditional text GAN (CTGAN) that is capable of generating high-quality

diverse text content and variable-length text. An automated method is also proposed to replace keywords that specify the context with the words that are synonymous from the trained text data. CTGAN is conditioned on the emotion label as an auxiliary input to have a control on topic. They used an LSTM model as generator network and a CNN model as discriminator network. The CTGAN model is evaluated on Yelp restaurant reviews, Amazon reviews, and film review data and generated text with high quality of variable length. Automatically generating the text and summarizing it to human-readable and a semantically similar way of the original text is defined as text summarization. Text summarization is categorized into two ways: extractive summarization and abstractive summarization. Extractive summarization extracts the important words, phrases, and sentences and summarizes them. Abstractive summarization generates the text and then summarizes it to reflect as the original. Zhuang et al. [141] proposed an abstractive summarization method using a GAN model which consists of one generator and two discriminators. The generator is responsible for the encoding of long input text sentences into a short text representation. The first discriminator trains the generator to generate the text in human-readable form and the second discriminator trains the generator to keep the prominent features of the original text to convert the generated text semantically similar to the original. The authors have implemented a policy gradient to train the model. The process of rephrasing the sentence from the original style to another style without altering the semantics of the text is defined as style transfer. If it is from source style to target style, then it is called unidirectional style transfer. Alternatively, multi-directional style transfer is also possible but at the cost of multiple trainings. If there are  $k$  attributes, then  $k \times (k - 1)$  training models are required. It has applications in NLP, for example, sentiment transformation, formality modification, etc., and computer vision. Yu et al. [123] discussed a unified GAN (UGAN) model that transfers styles among multiple attributes using one training model. The original text and target attributes are given as input to the generator that generates text based on the given attributes. The discriminator takes this text and real text as input and generates as output

a rank and classification output. The proposed model significantly reduced the training time for multi-directional style transfer.

### 3.7 Image deblurring and dehazing

Adverse weather conditions like fog, rain, haze, and pollution deteriorate the quality of images. Increasing the contrast, colour, and texture of images to improve the quality of images is referred to as image dehazing. In general, image dehazing techniques are categorized into image enhancement and model-based dehazing approaches. Pang et al. [75] introduced a model based method named haze removal GAN (HRGAN) that uses mathematical inversion techniques to reconstruct the haze-free images. The generator network consists of three modules: a transmission map module, atmospheric module, and a processing module that generates a reconstructed haze-free image. The CNN-based discriminator classifies between real haze-free images and reconstructed haze-free images. It employs a significant loss that consists of pixel-wise loss, perceptual loss, and an adversarial loss to train the HRGAN model. The HRGAN model achieved significant results in terms of removing haze and the quality of image compared to other benchmarks on NYU2, synthetic, Middlebury, and SOTS datasets. Zhao et al. [132] developed a pyramid GAN (PGAN) in which the authors have placed three GAN models in a pyramid shape. The first GAN block captures the non-local features of images at multiple levels. The second GAN block captures the local features of images at multiple levels. At this stage, the PGAN combines and balances the local and non-local features of images. The final GAN block identifies the sharp edges of the reconstructed image. The performance of the model is evaluated on GOPRO dataset and MS COCO dataset. Rain is the prime factor that affects the quality of images captured by surveillance systems in terms of blurring, raindrop obstacles, and deformation. Xiang et al. [109] discussed a feature supervised GAN (FS-GAN) that removes the rain streaks from a single image and enhances the image quality. It introduced a feature supervised guidance at the last layer of the generator network and achieved fair results. Jin et al. [44] discussed an asynchronous interactive GAN (AI-GAN) that deals with feature-wise extrication and finds the mutuality between feature-wise coupled components. Later, this interdependency is leveraged to achieve the deraining effect successively. The AI-GAN is capable of decomposing all the diverse features involved in a single image using a two-branch structure. Zhao et al. [134] discussed a double discriminator GAN (DD-GAN) that has two generators and leveraged two discriminators against each generator. The main motive behind employing two discriminators is to maintain a stable training process with limited training. They also used a weight clipping algorithm to increase the convergence

speed, to tackle unstable training, and mode collapse problems of GANs. The model has achieved promising results on RESIDE dataset and O-Haze dataset.

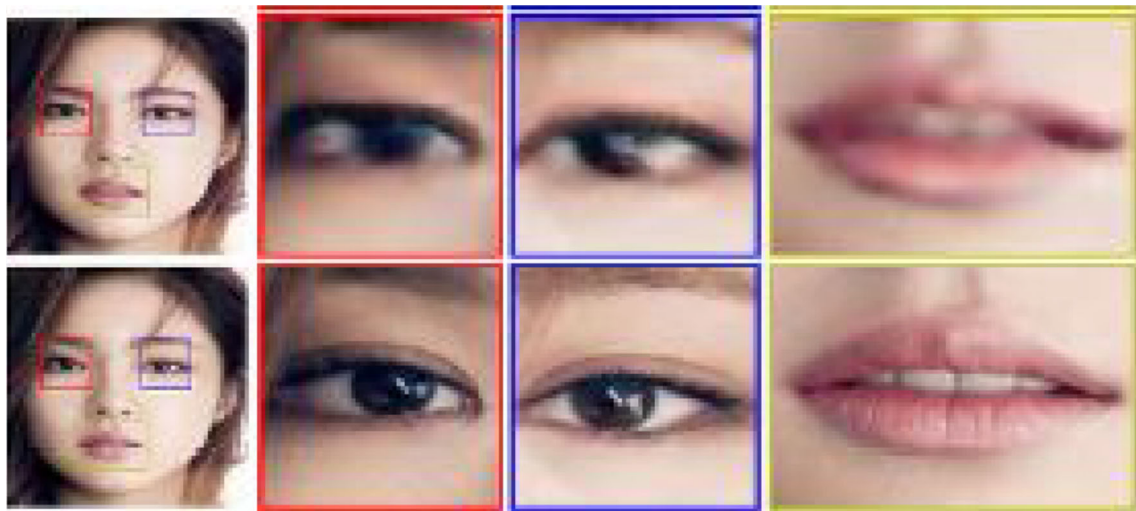
Li et al. [59] discussed an improved-SAGAN model to generate high-quality dairy goat images. He used a self-attention based normalized feature map method to compute the correlation between features. They also replaced the one-hot label for class labels with multi-class labels to improve the quality of images. They investigated the model on a collection of goat images and CelebA datasets and got significant improvements in results. [76] proposed an outdoor image dehazing technique that consists of two GANs: cycleGAN and cGAN with different properties. First, the cycleGAN is trained on outdoor images and to generate haze-free coloured images. On the other hand, cGAN is trained to keep the texture details like light, contrast, etc., of hazed images. Finally, a convolutional neural network is fused to generate haze-free images. Figure 8 shows the dehazed image of the hazed image using CycleGAN.

### 3.8 Face image synthesis

Facial image synthesis and super-resolution, is also known as face hallucination, are the two most discussed topics in the field of image processing and computer vision research. Face hallucination is the process of upscaling the low-resolution images to high-resolution images. Preserving the identity of the person is a challenge while performing face hallucination. Hsu et al. [38] discussed a Siamese GAN model (SiGAN) to reconstruct the faces in the process of face hallucination. The SiGAN consists of two generators and a discriminator. The two generators receive a low-resolution paired image as input and reconstruct a high-resolution paired image. This high-resolution image is given as input to the discriminator. They employed SiGAN loss which is a combination of GAN loss and contrastive loss, and reconstruction loss is used to train the SiGAN model. Experimental results on CASIA, LFW, and CelebA datasets proved the effectiveness of SiGAN model. [65] reconstructed a high-resolution facial image using a component semantic prior GAN (CSPGAN) from a low-resolution facial image. They introduce a gradient loss along with perceptual loss in computing the content loss of the generator. The discriminator network in the GAN is capable of predicting multiple task semantic category. The proposed model effectiveness is investigated on labelled faces in the wild (LFW) and facial HR images online (FHRO) datasets. Figure 9 shows the ground truth textures in the first row, and the second row shows the realistic textures captures by CSPGAN with multi-tasking capable discriminator.

Given a photo, synthesizing a pencil sketch is referred to as photo-sketch synthesis and has applications in the fields of digital entertainment and suspects identification in

**Fig. 8** Dehazed image (right) of hazed image (left) using CycleGAN [76]



**Fig. 9** Realistic textures captured by CSPGAN (2nd row) with multi-tasking discriminator [65]

law enforcement. This task suffers from loss of content, colour inconsistency, distorted faces, lack of clarity, and missing texture. [135] discussed a GAN model (EGGAN) which is guided by a feature encoder. The feature encoder is trained particularly to search for effective face photo-sketch domain latent space. This model can perform photosynthesis and sketch synthesis simultaneously. The model is validated on two publicly available benchmark datasets: CUFS and CUFSF. Han et al. [28] discussed another face frontalization GAN model named face merged GAN (FM-GAN) that has two generators and one discriminator. The first generator extracts the local face features from upper and lower parts of profile face using an encoder network. Then, a decoder network merges these features to synthesize a frontal face view. The decoder of the second generator takes the encoded features of profile face and the merged frontal face view as inputs and extracts the global features and generates a high dimensional frontal face view. Later, the discriminator trained on real and synthesized data and produced promising results on

Karolinska directed emotional faces (KDEF) dataset. Iranmanesh et al. [42] devised a coupled GAN (CpGAN) for face recognition task across diverse spectrums. It consists of two GAN-based sub-networks: Visible GAN and non-visible GAN paired by a contrastive loss function and performs nonlinear transformations. Both the generators are formed by encoder-decoder network, and discriminators are CNNs. The proposed model is evaluated on six different databases: CasiaHFB, Casia NIR-VIS, NightVision (NVESD), Notre Dame X1 (UNDX1), Polarimetric Thermal, and Wright State (WSRI) and achieved significant results.

In [31], He et al. introduced a super-resolution GAN model that synthesizes high-resolution facial images which are scaled four times of low facial resolution images at different resolutions. The Bicubic interpolation method is used to resize the low-resolution blurred images. These images along with the ground truth images from CelebA dataset are given as input to the stacked GAN that has three generators and three discriminators. They incorporated residual learn-

ing for upsampling of images. Experimental results proved that the proposed super resolution model outperformed other methods in terms of SR performance and generated realistic images. Sun et al. [92] considered the problem of short-term facial age synthesis along with long-term facial age synthesis over various age spans. They employed a GAN network guided by age label distribution (IdGAN), especially for short-term facial age synthesis. The label distribution consists of various age groups. The proposed model is experimented on Audience, CACD, FG-NET, MORPH, and UTK Face facial age databases and yielded remarkable results. The task of identifying the frontal face images from the profile face images is referred to face frontalization. It has applications in face recognition systems. Recently, GANs have proved their effectiveness in synthesizing frontal face images from profile face images with small face poses. To address this issue, [82] developed face frontalization method feature improving GAN (FIGAN) that achieved improved results with large face poses. The authors have employed a feature mapping block (FMB) that identifies the variance between the frontal face poses and profile face poses. The discriminator network is modelled with a feature discriminator that improves the latent features generated by FMB block in the generator. The model is investigated on celebrities in frontal faces (CFP), labelled faces in the wild (LFW), and MultiPIE databases.

### 3.9 Geoscience and remote sensing

Spectral sensors have been used for capturing hyperspectral images of the object from long distances. It captures both spatial information and spectral information of the target object. Classification of such information is much useful in the applications of land change monitoring, resource management, remote sensing of ground water resources, remote sensing of agriculture and vegetation, distant observing of forestry, urban development, scene interpretation in law enforcement, etc. Deep learning model requires a large number of samples for a successful classification process. However, the remote sensing community suffers from the problem of limited samples. Due to this, the training process end up with over-fitting problem, i.e. data perform well during training and fails to generalize. In [90] Shi et al. automatically generated building footprints from the satellite images using a conditional GAN. Instead of using the generic cost function, authors have deployed a Wasserstein distance to update the parameters. A gradient penalty term is also used along with Wasserstein distance. The generator functionality is implemented with U-Net architecture, and the discriminator functionality is implemented with PATCHGAN architecture. Zhu et al. [140] presented two schemes for the classification of 1D and 3D hyperspectral images. First, a spectral classifier is modelled using a 1D GAN. Second, a spectral-spatial classifier is

designed using a 3D GAN. The authors have used the GAN model as a regularization unit to alleviate the effect of over-fitting due to limited samples. The performance of the model is evaluated on three publicly available datasets: Salina, Indiana Pines, and Kennedy space centre. Cloud obstruction is a conventional problem in remote sensing object detection field. The cloud obstruction makes the remote sensing images of sea surface temperatures (SST) unclear and hazy. To overcome this problem, [19] proposed a deep convolutional-based GAN model with a novel inpainting loss function. The loss function consists of a supervision term that removes the unclearness and identifies the nearest encodings in the low-dimensional images.

Feng et al. [20] discussed a spatial-spectral GAN model that performs a multi-class classification of hyperspectral images. This model addresses two issues of the classification process. First, it addresses the inability of the discriminator in multi-class classification and Second, consideration of spatial and spectral information in the classification of hyperspectral images. Wang et al. [96] proposed a variational GAN using a semi-supervised method to classify hyperspectral images with limited labels. The semi-supervised context is incorporated using an encoder-decoder network, and a collaborative optimization framework is used to find the latent space between classification and sample generation tasks. The effectiveness of variational GAN is validated on four benchmark datasets: University of Pavia, Pavia centre, DCMall, and Jiamusi. Zhu et al. [137] devised a multi-branch conditional GAN (MCGAN) model to increase data for objection in remote sensing images. The MCGAN architecture consists of one generator, three discriminators, and a classifier that build using deep CNNs. The data augmentation process is carried out on NWPU VHR - 10 dataset with an alternative of DOTA dataset for severely low numbered instance group in NWPU VHR - 10 dataset. Later, NWPU VHR - 10 and DOTA datasets were merged to train the MCGAN. Experimental results proved the effectiveness of the model on the quality of objects detected from the generated images.

### 3.10 Video generation

Given a context, the process of forecasting the next sequence of frames is known as video prediction. It has a wide range of applications like autonomous driving, object tracking, robotic planning, etc. An underlying uncertainty associated with the dynamics of the real-world challenges the predictions. Wen et al. [106] generated a sequence of video frames  $y_{i+1}, y_{i+2}, \dots, y_k$  given two key input frames  $x_i$  and  $x_{k+1}$ . They used two generators  $G_1, G_2$  and two discriminators  $D_1, D_2$ . The generators are placed in a sequential manner, where the output of the first generator is fed into the second generator.  $G_1$  learns motions from real videos during

training and  $G_2$  adds more details to the output of  $G_1$ .  $D_1$  and  $D_2$  optimize the performance of generators through adversarial training. Investigations proved that the generated video frames are clear and smooth. In [39], Hu et al. introduced a novel stochastic video prediction GAN (VPGAN) that is trained based on the cycle-consistent loss to predict the next sequence of actions in a video. An image segmentation model is also incorporated using two generators to extract the features. The proposed model is investigated on four datasets: Moving Mnist, KTH, BAIR, and UCF101 and achieved significant improvement in the quality of predicted future frames. [13] devised a model bottom-up GAN (BoGAN) to generate video frames from text descriptions. The model has an attention model that computes the region loss to fill the sub-regions of the video frame conditioned by words. The discriminator employs a frame-level loss to keep the semantic matching between successive frames. Finally, another discriminator that maintains the global level semantics between the sequence of frames in the final video. The model produced promising results compared to benchmarks.

### 3.11 Animation creation

Anime character and animation creation is a challenging task in the domain of multimedia applications. Image-to-Image translation and image super-resolution are the two major tasks involved in anime character synthesis. [45] modelled a cartoonGAN that transforms real-world images into cartoon style images a challenging task in computer graphics. The generator network of cartoonGAN consists of convolution, deconvolution, and residual blocks. The discriminator comprises convolutional layers. The model takes a set of real-world images and another set of cartoon images for training. The model also employs two loss functions during training to find the latent space between the two sets of images. A semantic content loss that manages the variations between real images and cartoon images, and an edge enhancing adversarial loss that maintains the sharp edges. Experimental results proved that the generated cartoon images are of high quality and surmounts the state-of-the-art style transforming methods. [26] imposed structural conditions at each scale of image generation during progressive training of progressive structure-conditional GAN (PSGAN). PSGAN generates anime images at  $1024 \times 1024$  resolution with full-body structure. A landmark assisted CycleGAN [108] is modelled to generate high-quality cartoon faces from real faces. The unpaired real faces and cartoon faces are used to train the model. The model employs a regressor to detect the landmarks in the generated cartoon faces. A novel landmark consistency loss is used during training to capture the important features of real faces. Landmark consistency, along with the local discriminators, alleviates the structural variance

between the real faces and cartoon faces. Figure 10 shows the cartoon faces generated by landmark assisted CycleGAN.

[119] proposed an image reconstruction method (PI-REC) that takes the flat colour domain and binary sparse edge as input to produce high quality reconstructed images. The authors have incorporated a GAN model that consists of three generators, and three discriminators in parallel, and each GAN model works in a phase and refines the reconstructed image details progressively. The sparse and interpretable inputs ensure the control over style and content of images being generated. Finally, the method also produced significant results on image to image translation task, provided the domains should be similar. [120] proposed a GAN model in which the discriminator generates two kinds of pseudo-labels using the self-supervised approach. Later, the discrete pseudo-labels are mapped to latent variables during training and eventually mapped to animation features to generate diverse animation clips. The continuous pseudo-labels are used to create diverse frames in one animation clip. They also discussed a novel metric to investigate the quality of animations.

### 3.12 Other application domains

When training data and testing data do not agree with each other, they pose a challenge for speech recognition in noisy environments. Qian et al. [79] discussed a GAN model for data augmentation and to improve the task of speech recognition, especially under noise conditions. A basic GAN is employed for data generation process based on FBANK feature map and generated frame by frame feature map. Since the generated data do not have labels, later, an unsupervised learning framework is deployed for the speech recognition task. The authors have conditioned one GAN on acoustic state and the other GAN on clean speech for better data generation. The collection of hard labels and soft labels achieved promising performance using conditional GAN on Aroua-4 and AMI-SDM datasets. Industries heavily depend on failure data to mitigate the occurrence of hazardous events and loss of human life. Thus, a risk warning system is an essential tool for identifying and avoiding such rare events. However, these rare events suffer from the problem of data scarcity for risk analysis. In [33] He et al. constructed a semi-supervised real-time risk management system by integrating fuzzy HAZOP risk analysis with a distributed control system (DCS). They also employed a GAN model that augments labelled process data which enhances the assessment of the type of risk classification. The framework is evaluated using a case study on the processing of polyolefin using a multizone circulating reactor (MZCR). Domain adaptation is an important area of research in the field of computer vision. Given two distributions: labelled and unlabelled relating to target data shifting domain from labelled to unlabelled is defined as domain

**Fig. 10** Cartoon faces generated by [108] on the last column for the given real faces on the first column compared to others in the middle column



adaptation. In this context, [14] modelled an unsupervised framework that contains a feature extractor, attention module embedded GAN (GAACN), and a classifier. The attention module is placed between the generator and discriminator to shift the transferable regions among different domains. They also used a label classifier module to keep the class consistency in discriminator network. The feature extractor is forced to learn the joint feature distribution by the GAN module. The feature extractor and classifier module are used in the testing phase to label the unlabelled target data. The GAN model and the attention module are built of convolutional layers. The experimental results on i. Digits dataset: MNIST, USPS, SVHN ii. ImageCLEF-DA dataset: Caltech-256, ILSVRC 2012, Pascal VOC 2012 iii. Office 31 dataset: Amazon, Webcam, DSLR iv. Office-Home dataset: Artistic domain, Clip Art, Product domain, Real-World domain v. VisDA 2017 dataset: synthetic domain, the real domain has produced significant results compared to other conventional models.

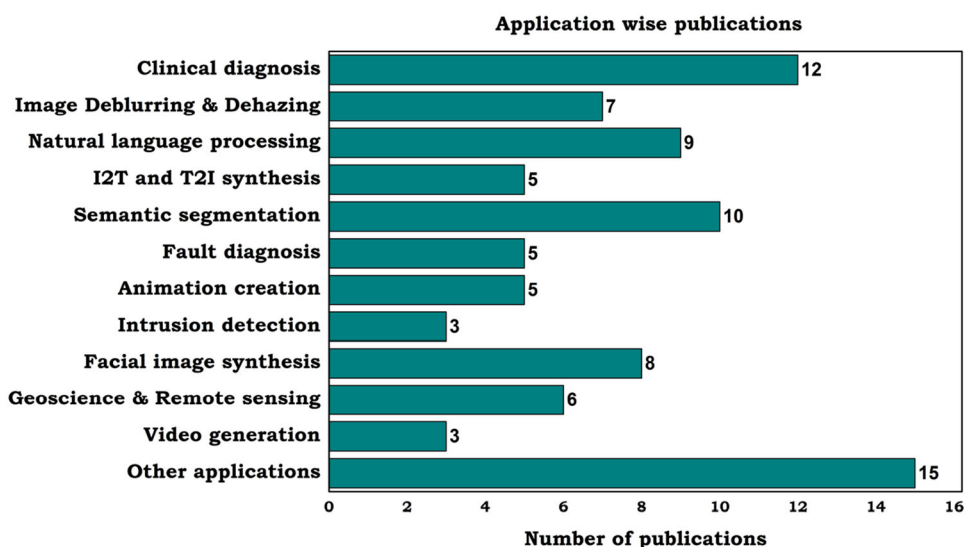
Wang et al. [97] discussed a new deep learning-based model named adaptive balancing GAN (AdaBalGAN) model to identify the defective types in imbalanced wafer maps data. They used a conditioned GAN model to generate wafer maps of a specific type, and a generative controller is used to change the sample distribution of the wafer maps concurring to the various defective patterns. The proposed model is evaluated on real-world fabricated WM-811K wafer maps. [29] proposed a conditional generation method which generates time-series data that belongs to multiple classes. The authors have employed a canonical correlation analysis to exemplify the characteristics between the input and generated data. They also deployed the LSTM model in both generator and discriminator. [56] designed a controllable GAN (Control

GAN) to reduce the overfitting problem occurred by auxiliary classifier in the discriminator of ACGAN [72]. ResNet [32]-based generator network and discriminator network are used in ControlGAN. Authors have used a ResNet-based independent classifier to evaluate the generated samples. The proposed model is evaluated using CIFAR-10, CelebA, and LSUN datasets. Mandal et al. [68] developed a deep CNN-based semi-supervised GAN (SSGAN) for the food recognition task. Food recognition is a thought provoking task due to huge interclass variation in food images. Experimental results proved the effectiveness of the proposed semi-supervised model on ETH Food-101 Dataset and Indian Food Dataset. In [64] Lin et al. designed a defect enhancement GAN (DEGAN) based on deep convolutional GAN (DCGAN) [80] and energy-based GAN (EBGAN) [133] to generate microcrack defective samples. It incorporates a defect enhancement algorithm in the forward path and after the discriminator also. The generator model consists of convolutional layers, and the discriminator is implemented using an encoder and decode network. [62] notified a similarity constraint GAN (SCGAN) that identifies the entangled feature and represents it in disentangled representation in an unsupervised manner. The proposed model is investigated on MNIST, FASHION MNIST, SVHN, CIFAR-10, and CelebA datasets and gained significant improvements in results.

Reference [95] designed an evolutionary algorithm based GAN (EGAN) framework in which they stabilized the GAN training. The evolutionary algorithm optimizes the generator's objective using multiple training objectives. It consists of three phases: evaluation, variation, and selection. The proposed framework is evaluated using three datasets: CIFAR-10, LSUN bedroom, and CelebA and obtained significant improvement in results. Kasem et al. [47] introduced



**Fig. 11** Application wise number of publications



robust super-resolution GAN (RSR-GAN) that addresses two issues while improving the quality of subjects in the images. First, it regains the texture details with extreme upscaling factors. Second, it alleviates the noise generated due to geometric transformations. The RSR-GAN has a transformer module in discriminator that enhances the discrimination capacity. The generator loss term has an additional DCT loss term that finds the right mapping between generated and real images. The authors have used Berkeley segmentation data set for training and BDS100, MANGA109, SET5, SET14, and URBAN100 datasets for testing and obtained significant results. [30] proposed a style consistent GAN (GlyphGAN) that generates novel font types. The generated font types are style consistent, legible, and diverse overall characters. [94] designed a compressive privacy GAN (CPGAN) to defend attacks while sharing data using machine learning as a service (MLaaS) in cloud platforms. [121] devised a long short-term memory based conditional GAN (LSTM-GAN) to identify the taxi hotspots in both dimensions: spatial and temporal. [16] generated realistic user behaviour data related to the products that have not been released yet using a conditional GAN framework. Figure 11 shows the graph of application wise number of publications considered.

## 4 Evaluation metrics

There has been an extensive usage of GANs in diverse applications in the late years of this decade. Generative modelling aims at mimicking the trained data with generated data. Hence, it obvious to measure the distance between the real data and artificial data. In general, a distance function that computes the distance between a real distribution and generated distribution are used as loss functions. However, no standardized metrics are devised to evaluate how good GANs

are in mimicking trained data. Thus, in this section, we present a few metrics that are used in the literature extensively to evaluate the GAN model.

### 4.1 1-Nearest neighbour classifier (1-NN)

It is a version of classifier two sample tests (C2ST) and is not an evaluation metric. It checks the similarity between real data distribution  $P_w(\cdot)$  and generated data distribution  $P(\cdot)$ . It computes the leave one out (LOO) cross-validation accuracy of classification, where all the data points except one point are used to estimate accuracy and left out point is used for prediction.

### 4.2 Inception scores (IS)

It is a metric derived by [86] to evaluate the quality and diversity of synthesized images by generative models. First, they computed the conditional probability of an instance belonging to a class. Later, these conditional probabilities are used to compute the inception score on a pre-trained inception network. If the conditional label distribution has low entropy, then the generated images are of good quality. To produce a variety of images, the network should have a low marginal conditional probability distribution. The inception score ranges between 1 and total classes. The limitation of this metric is that it does not consider the statistics (mean, variance, and standard deviation) of the original data distribution to compare with generated samples distribution.

$$IS(P_g) = e^{E_{x \sim p_g} [\mathcal{K}\mathcal{L}(p(y|x) \| p(y))]} \quad (3)$$

### 4.3 Mode score (MS)

This metric overcomes the limitation faced by IS metric and considers the statistics of prior distribution to evaluate the quality of images and diversity of images [11]. The mode score can be computed using the Eq. 4.  $p(y^*)$  represents the distribution of ground truth labels computed using original data distribution.

$$MS(P_g) = e^{\left(\mathbb{E}_{\mathbf{x} \sim \mathbf{g}}[\mathcal{KL}(p(y|\mathbf{x})\|p(y^*))] - \mathcal{KL}(p(y)\|p(y^*))\right)} \quad (4)$$

### 4.4 Frechet inception distance (FID)

It is a distance metric between the feature vectors of real data and generated data. It measures the quality of generated images and finds the occurrence of intra-class mode collapse. However, this metric considers the mean ( $m$ ) and covariance ( $C$ ) of two Gaussians under study. The Frechet distance [35]  $d(m, C)$  between the real data distribution  $P_w(m_w, C_w)$  and the synthetic data distribution  $P_s(m_s, C_s)$  is defined as follows:

$$d^2((m_s, C_s), (m_w, C_w)) = \|m_s - m_w\|_2^2 + \text{Tr}\left(C_s + C_w - 2(C_s C_w)^{1/2}\right) \quad (5)$$

where  $Tr$  represents the trace computation.

### 4.5 Maximum mean discrepancy (MMD)

The MMD metric is used to compute the dissimilarity between the real data distribution  $P_r$  and the generated data distribution  $P_g$ . If we employ a fixed Gaussian kernel  $k(x, x') = e^{-\|x-x'\|^2}$ , then the kernel MMD [25] is computed as shown in Eq. 6. A lower MMD indicates  $P_g$  is more similar to  $P_r$ .

$$KMMD(P_r, P_g) = E_{x, x' \sim P_r} [k(\mathbf{x}, \mathbf{x}')] - 2E_{x \sim P_r, y \sim P_g} [k(\mathbf{x}, \mathbf{y})] + E_{y, y' \sim P_g} [k(\mathbf{y}, \mathbf{y}')] \quad (6)$$

### 4.6 Multi-scale structural similarity for image quality

Wang, et al. [105] used structural information  $s(x, y)$ , luminance information  $l(x, y)$ , and contrast information  $c(x, y)$  to derive the structural similarity index (SSIM) as shown in

7. It assesses the similarity index between two images.

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (7)$$

$\mu_x, \mu_y, \sigma_x, \sigma_y$  denote the mean and standard deviations of image signal  $x$  and image signal  $y$ , respectively.  $C_i$  is a constant. [104] extended this metric to multi-scale to assess the quality of images by integrating different image resolutions. The multi-scale SSIM is computed as follows:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = [l_M(\mathbf{x}, \mathbf{y})]^{\alpha M} \cdot \prod_{j=1}^M [c_j(\mathbf{x}, \mathbf{y})]^{\beta j} [s_j(\mathbf{x}, \mathbf{y})]^{\gamma j} \quad (8)$$

The exponents  $\alpha, \beta$ , and  $\gamma$  are included to alter the relative importance of various components.

### 4.7 Wasserstein critic

It estimates the Wasserstein distance between the real data distribution  $P_r$  and the generated data distribution  $P_g$ . This metric estimates lower values for generated instances and higher values for real instances. In case of discrete distribution transformations, it is also known as Earth Mover's distance (EMD). The Wasserstein critic between  $P_g$  and  $P_r$  is estimated as shown in Eq. 9.

$$W(P_r, P_g) \propto \max_f \mathbb{E}_{\mathbf{x} \sim P_r} [f(\mathbf{x})] - \mathbb{E}_{\mathbf{x} \sim P_g} [f(\mathbf{x})] \quad (9)$$

where  $f: R^D \rightarrow R$  denotes the Lipschitz continuous function [8].

## 5 Challenges

Despite the success and extensive usage of GANs, often they face some common challenges during the training. The three most important challenges faced by GANs are as follows:

- If the generator is not as good as the discriminator, then the discriminator always differentiates between real and artificial data. Hence, the gradients of the generator will be vanished, leading to failure of the generator. [5] succeeded in eliminating the vanishing gradient problem of generators by introducing Wasserstein loss 4.7 to compute the distance between real and artificial data. However, it is not guaranteed that replacing the min–max loss with Wasserstein loss can eradicate the vanishing gradient problem as it also depends on other factors like

**Table 1** A summary of the tasks, applications, and the datasets used

Task/Reference	Applications	Datasets
Image Dehazing [134]	Traffic monitoring	RESIDE
	Security monitoring	O-Haze
Face photo - sketch synthesis [135]	Digital entertainment	CUFS
	Law enforcement	CUFSF
Face frontalization [82]	Face recognition systems	MultiPIE
		LFW
		CFP
Cross-modal face recognition [42]	Nighttime face recognition	WSRI
		UND X1
		NVESD
		Casia NIR-VIS
		Casia HFB
		Polarimetric Thermal
Person re-identification [125]	Surveillance systems	VPIeR
		CUHK03
		Market-1501
Text generation [115]	Dialogue systems	COCO
	Machine translation	Chinese poetry
	Writing poetry	
	BOT applications	
Abstractive text summarization [141]	News letters	CNN-Daily Mail
	Social media marketing	
Text style transfer [123]	Sentiment transformation	YELP
	Formality direction	AMAZON
		CAPTION
Hyper-spectral image classification [140]	Land change monitoring	Salinas
	Resource management	Indiana Pines
	Remote sensing of agriculture	Kennedy Space Center
Video prediction [39]	Autonomous driving	Moving Mnist
	Robotic planning	BAIR
	Object tracking	KTH
		UCF101
Semantic segmentation [73]	Autonomous driving	PASCAL VOC 2012
	Medical diagnostics	Cam Vid
	Robotic systems	CelebA
Fault diagnosis [89]	Industrial machine monitoring	Induction motor signal vibration
	Risk analysis	
	Clinical diagnosis	
ECG Analysis [99]	Cardiovascular diseases	MIT-BIH Database

available data, hyperparameter settings, model structure, etc.

- The generator tries to over-optimize the discriminator in each epoch of the GAN training. If the discriminator caught in the local minimum trap and always rejecting every instance of its input, then the generator keeps on generates the same set of instances. This is popularly

known as mode collapse problem of GANs. Wasserstein loss [5] does not let discriminator struck at local optimum, and hence, generator produces a new set of outputs. [70] modelled generators objective in coherence with optimal discriminator to alleviate mode collapse problem. [25] devised an empirical model that automatically detects the problem of mode anomaly.

**Table 1** continued

Task/Reference	Applications	Datasets
MRI, PET, and CT Analysis [21]	Cancer detection Functional abnormalities in body parts	T2-Flair LIDC-IDRI
Mammogram segmentation [131]	Detection of breast cancer	INbreast DDSM Hospital Sant Joan de Reus
Data augmentation [79]	Speech recognition Audio denoising	Aroua-4 AMI-SDM
DeepFakes [18]	Educational applications Enlightening mankind	DFDC CelebDF FaceForensics++DF

- Often GANs fail to converge due to irregularities in the structure of the model, hyperparameter tuning, and training strategies. [4] added noise to the inputs of a discriminator for stable training and GAN convergence. [84] introduced a novel regularization method to eliminate the problem of convergence.
- The above-said challenges are discussed in the perspective of algorithmic and training issues of GANs. GANs are highly successful in generating high-quality naturalistic images. However, the performance of GANs is questioned in creating fake videos, also called deepfakes. Creating fake videos using deep learning techniques to swap the identity of a person with another person is called deepfakes. However, the deepfakes resembles realistic, it is difficult to create a deepfake that mimics eye blinking, since nobody likes to take a picture with eyes closing. Also, while creating deepfakes, we need images that have persons with similar skin tone, the orientation of faces, etc. Otherwise, the output deepfake would not be optimal. Deepfakes created using pairwise deepfake auto-encoder (DFAE) models are higher in quality compared to deepfakes created using GAN-based methods on deepfake detection challenge (DFDC) dataset [18].

## 6 Conclusion

This paper presents the ins and outs of GANs, derived GANs, their application areas, evaluation metrics, and challenges involved in training GANs. A total of 88 publications are summarized based on their objective with an ease of understanding terminology to a naive researcher. At this point, it may be noted that the main objective of some publications related to clinical diagnosis is segmentation. So, the said publications are discussed in detail in the respective sect. 3.4. It is obvious that image super-resolution is an application of GAN worth discussing. But it is covered as part

of section 3.8 to limit the size of the paper. We also noted that all supervised, semi-supervised, and unsupervised algorithms are discussed. However, mostly semi-supervised and unsupervised algorithms are used in case of data insufficient problems. The application of GANs are spread over diverse domains, and they are not limited to the ones discussed in this paper. Often, GANs face a challenge with the increase in the number of distributions in the real data. With sophisticated techniques for training GANs, especially when dealing with a large number of distributions, the applications of GANs can be widespread. It also provides a summary on various applications, tasks achieved, and the datasets in Table 1 for quick referencing.

## References

1. AbdAlmageed W, Wu Y, Rawls S, Harel S, Hassner T, Masi I, Choi J, Lekust J, Kim J, Natarajan P, et al. (2016) Face recognition using deep multi-pose representations. In: 2016 IEEE winter conference on applications of computer vision (WACV), IEEE, pp 1–9
2. Ak KE, Lim JH, Tham JY, Kassim AA (2020) Semantically consistent text to fashion image synthesis with an enhanced attentional generative adversarial network. *Pattern Recognit Lett* 135:22–29
3. Antipov G, Baccouche M, Dugelay JL (2017) Face aging with conditional generative adversarial networks. In: 2017 IEEE international conference on image processing (ICIP), IEEE, pp 2089–2093
4. Arjovsky M, Bottou L (2017) Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862*
5. Arjovsky M, Chintala S, Bottou L (2017) Wasserstein gan. *arXiv preprint arXiv:1701.07875*
6. Bisneto TRV, de Carvalho Filho AO, Magalhães DMV (2020) Generative adversarial network and texture features applied to automatic glaucoma detection. *Appl Soft Comput* 90:106165
7. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *J Mach Learn Res* 3:993–1022
8. Borji A (2019) Pros and cons of gan evaluation measures. *Comput Vis Image Underst* 179:41–65

9. Brock A, Lim T, Ritchie JM, Weston N (2016) Neural photo editing with introspective adversarial networks. arXiv preprint [arXiv:1609.07093](https://arxiv.org/abs/1609.07093)
10. Cai Y, Wang X, Yu Z, Li F, Xu P, Li Y, Li L (2019) Dualattn-gan: text to image synthesis with dual attentional generative adversarial network. *IEEE Access* 7:183706–183716
11. Che T, Li Y, Jacob AP, Bengio Y, Li W (2016) Mode regularized generative adversarial networks. arXiv preprint [arXiv:1612.02136](https://arxiv.org/abs/1612.02136)
12. Chen J, Wu Y, Jia C, Zheng H, Huang G (2019) Customizable text generation via conditional text generative adversarial network. *Neurocomputing*. <https://doi.org/10.1016/j.neucom.2018.12.092>
13. Chen Q, Wu Q, Chen J, Wu Q, van den Hengel A, Tan M (2020) Scripted video generation with a bottom-up generative adversarial network. *IEEE Trans Image Process* 29:7454–7467
14. Chen W, Hu H (2020) Generative attention adversarial classification network for unsupervised domain adaptation. *Pattern Recognit* 107:107440
15. 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
16. Chonwiharnphan P, Thienprapasith P, Chuangsuwanich E (2020) Generating realistic users using generative adversarial network with recommendation-based embedding. *IEEE Access* 8:41384–41393
17. Delannoy Q, Pham CH, Cazorla C, Tor-Díez C, Dollé G, Meunier H, Bednarek N, Fablet R, Passat N, Rousseau F (2020) Segsrgan: super-resolution and segmentation using generative adversarial networks-application to neonatal brain mri. *Comput Biol Med* 120:103755
18. Dolhansky B, Bitton J, Pflaum B, Lu J, Howes R, Wang M, Ferrer CC (2020) The deepfake detection challenge dataset. arXiv preprint [arXiv:2006.07397](https://arxiv.org/abs/2006.07397)
19. Dong J, Yin R, Sun X, Li Q, Yang Y, Qin X (2018) Inpainting of remote sensing sst images with deep convolutional generative adversarial network. *IEEE Geosci Remote Sens Lett* 16(2):173–177
20. Feng J, Yu H, Wang L, Cao X, Zhang X, Jiao L (2019) Classification of hyperspectral images based on multiclass spatial-spectral generative adversarial networks. *IEEE Trans Geosci Remote Sens* 57(8):5329–5343
21. 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(9):2059–2069
22. Ghassemi N, Shoeibi A, Rouhani M (2020) Deep neural network with generative adversarial networks pre-training for brain tumor classification based on mr images. *Biomed Signal Process Control* 57:101678
23. Goodfellow I, 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*, pp 2672–2680
24. Grover A, Dhar M, Ermon S (2017) Flow-gan: Combining maximum likelihood and adversarial learning in generative models. arXiv preprint [arXiv:1705.08868](https://arxiv.org/abs/1705.08868)
25. Guo C, Huang D, Zhang J, Xu J, Bai G, Dong N (2020) Early prediction for mode anomaly in generative adversarial network training: an empirical study. *Inf Sci* 534:117–138
26. Hamada K, Tachibana K, Li T, Honda H, Uchida Y (2018) Full-body high-resolution anime generation with progressive structure-conditional generative adversarial networks. In: *Proceedings of the European conference on computer vision (ECCV)*
27. Han L, Huang Y, Dou H, Wang S, Ahamad S, Luo H, Liu Q, Fan J, Zhang J (2020) Semi-supervised segmentation of lesion from breast ultrasound images with attentional generative adversarial network. *Comput Methods Programs Biomed* 189:105275
28. Han Z, Huang H, Huang T, Cao J (2019) Face merged generative adversarial network with tripartite adversaries. *Neurocomputing* 368:188–196
29. Harada S, Hayashi H, Uchida S (2019) Biosignal generation and latent variable analysis with recurrent generative adversarial networks. *IEEE Access* 7:144292–144302
30. Hayashi H, Abe K, Uchida S (2019) Glyphgan: style-consistent font generation based on generative adversarial networks. *Knowl-Based Syst* 186:104927
31. He J, Zheng J, Shen Y, Guo Y, Zhou H (2020a) Facial image synthesis and super-resolution with stacked generative adversarial network. *Neurocomputing* 402:359–365
32. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 770–778
33. He R, Li X, Chen G, Chen G, Liu Y (2020b) Generative adversarial network-based semi-supervised learning for real-time risk warning of process industries. *Expert Syst Appl* 150:113244
34. He X, Fang L, Rabbani H, Chen X, Liu Z (2020c) Retinal optical coherence tomography image classification with label smoothing generative adversarial network. *Neurocomputing* 405:37–47
35. 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
36. Hinton GE (2012) A practical guide to training restricted Boltzmann machines. In: *Montavon G, Orr GB, Muller KR (eds) Neural networks: tricks of the trade*. Springer, Berlin, pp 599–619
37. Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554
38. Hsu CC, Lin CW, Su WT, Cheung G (2019) Sigan: siamese generative adversarial network for identity-preserving face hallucination. *IEEE Trans Image Process* 28(12):6225–6236
39. Hu Z, Turki T, Wang JT (2020) Generative adversarial networks for stochastic video prediction with action control. *IEEE Access* 8:63336–63348
40. Huang S, Lei K (2020) Igan-ids: an imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks. *Ad Hoc Netw* 105:102177
41. Im DJ, Kim CD, Jiang H, Memisevic R (2016) Generating images with recurrent adversarial networks. arXiv preprint [arXiv:1602.05110](https://arxiv.org/abs/1602.05110)
42. Iranmanesh SM, Riggan B, Hu S, Nasrabadi NM (2020) Coupled generative adversarial network for heterogeneous face recognition. *Image Vis Comput* 94:103861
43. 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
44. Jin X, Chen Z, Li W (2020) Ai-gan: asynchronous interactive generative adversarial network for single image rain removal. *Pattern Recognit* 100:107143
45. Jin Y, Zhang J, Li M, Tian Y, Zhu H, Fang Z (2017) Towards the automatic anime characters creation with generative adversarial networks. arXiv preprint [arXiv:1708.05509](https://arxiv.org/abs/1708.05509)
46. Karras T, Aila T, Laine S, Lehtinen J (2017) Progressive growing of gans for improved quality, stability, and variation. arXiv preprint [arXiv:1710.10196](https://arxiv.org/abs/1710.10196)
47. Kasem HM, Hung KW, Jiang J (2019) Spatial transformer generative adversarial network for robust image super-resolution. *IEEE Access* 7:182993–183009
48. Kim HJ, Lee D (2020) Image denoising with conditional generative adversarial networks (cgan) in low dose chest images. *Nucl Instrum Methods Phys Res, Sect A* 954:161914

49. Kim K, Myung H (2018) Autoencoder-combined generative adversarial networks for synthetic image data generation and detection of jellyfish swarm. *IEEE Access* 6:54207–54214
50. 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, JMLR. org, pp 1857–1865
51. Kuang Y, Lan T, Peng X, Selasi GE, Liu Q, Zhang J (2020) Unsupervised multi-discriminator generative adversarial network for lung nodule malignancy classification. *IEEE Access* 8:77725–77734
52. Kupyn O, Budzan V, Mykhailych M, Mishkin D, Matas J (2018) Deblurgan: blind motion deblurring using conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 8183–8192
53. Kwak H, Zhang BT (2016) Generating images part by part with composite generative adversarial networks. *arXiv preprint arXiv:1607.05387*
54. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444
55. 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
56. Lee M, Seok J (2019) Controllable generative adversarial network. *IEEE Access* 7:28158–28169
57. Lei B, Xia Z, Jiang F, Jiang X, Ge Z, Xu Y, Qin J, Chen S, Wang T, Wang S (2020) Skin lesion segmentation via generative adversarial networks with dual discriminators. *Med Image Anal* 64:101716
58. Lei M, Zhou Y, Zhou L, Zheng J, Li M, Zou L (2019) Noise-robust wagon text extraction based on defect-restore generative adversarial network. *IEEE Access* 7:168236–168246
59. Li H, Tang J (2020) Dairy goat image generation based on improved-self-attention generative adversarial networks. *IEEE Access* 8:62448–62457
60. Li J, Liang X, Wei Y, Xu T, Feng J, Yan S (2017a) Perceptual generative adversarial networks for small object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1222–1230
61. Li J, Monroe W, Shi T, Jean S, Ritter A, Jurafsky D (2017b) Adversarial learning for neural dialogue generation. *arXiv preprint arXiv:1701.06547*
62. Li X, Chen L, Wang L, Wu P, Tong W (2018a) Scgan: disentangled representation learning by adding similarity constraint on generative adversarial nets. *IEEE Access* 7:147928–147938
63. Li Y, Pan Q, Wang S, Yang T, Cambria E (2018b) A generative model for category text generation. *Inf Sci* 450:301–315
64. Lin S, He Z, Sun L (2019) Defect enhancement generative adversarial network for enlarging data set of microcrack defect. *IEEE Access* 7:148413–148423
65. Liu L, Wang S, Wan L (2019) Component semantic prior guided generative adversarial network for face super-resolution. *IEEE Access* 7:77027–77036
66. Liu MY, Tuzel O (2016) Coupled generative adversarial networks. In: *Advances in neural information processing systems*, pp 469–477
67. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) Ssd: Single shot multibox detector. In: *European conference on computer vision*, Springer, pp 21–37
68. Mandal B, Puhon NB, Verma A (2018) Deep convolutional generative adversarial network-based food recognition using partially labeled data. *IEEE Sens Lett* 3(2):1–4
69. Masi I, Trn AT, Hassner T, Leksut JT, Medioni G (2016) Do we really need to collect millions of faces for effective face recognition? In: *European conference on computer vision*, Springer, pp 579–596
70. Metz L, Poole B, Pfau D, Sohl-Dickstein J (2016) Unrolled generative adversarial networks. *arXiv preprint arXiv:1611.02163*
71. Mirza M, Osindero S (2014) Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*
72. Odena A, Olah C, Shlens J (2016) Conditional image synthesis with auxiliary classifier gans *arxiv e-prints*. (oct. *arXiv preprint arXiv:1610.09585*)
73. Oluwasanmi A, Aftab MU, Shokanbi A, Jackson J, Kumeda B, Qin Z (2020) Attentively conditioned generative adversarial network for semantic segmentation. *IEEE Access* 8:31733–31741
74. Pan T, Chen J, Xie J, Chang Y, Zhou Z (2020) Intelligent fault identification for industrial automation system via multi-scale convolutional generative adversarial network with partially labeled samples. *ISA Trans* 101:379–389
75. Pang Y, Xie J, Li X (2018) Visual haze removal by a unified generative adversarial network. *IEEE Trans Circuits Syst Video Technol* 29(11):3211–3221
76. Park J, Han DK, Ko H (2020) Fusion of heterogeneous adversarial networks for single image dehazing. *IEEE Trans Image Process* 29:4721–4732
77. Perarnau G, Van De Weijer J, Raducanu B, Álvarez JM (2016) Invertible conditional gans for image editing. *arXiv preprint arXiv:1611.06355*
78. Qi L, Zhang H, Tan W, Qi S, Xu L, Yao Y, Qian W (2019) Cascaded conditional generative adversarial networks with multi-scale attention fusion for automated bi-ventricle segmentation in cardiac mri. *IEEE Access* 7:172305–172320
79. Qian Y, Hu H, Tan T (2019) Data augmentation using generative adversarial networks for robust speech recognition. *Speech Commun* 114:1–9
80. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*
81. Rizzo G, Van THM (2020) Adversarial text generation with context adapted global knowledge and a self-attentive discriminator. *Inf Process Manag* 102217
82. Rong C, Zhang X, Lin Y (2020) Feature-improving generative adversarial network for face frontalization. *IEEE Access* 8:68842–68851
83. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional network for biomedical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*, Springer, pp 234–241
84. Roth K, Lucchi A, Nowozin S, Hofmann T (2017) Stabilizing training of generative adversarial networks through regularization. In: *Advances in neural information processing systems*, pp 2018–2028
85. Ruan Y, Li D, Marshall H, Miao T, Cossetto T, Chan I, Daher O, Accorsi F, Goela A, Li S (2020) Mb-fsgan: joint segmentation and quantification of kidney tumor on ct by the multi-branch feature sharing generative adversarial network. *Med Image Anal* 64:101721
86. 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
87. Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G (2017) Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: *International conference on information processing in medical imaging*, Springer, pp 146–157
88. Shaker AM, Tantawi M, Shedeed HA, Tolba MF (2020) Generalization of convolutional neural networks for ecg classification

- using generative adversarial networks. *IEEE Access* 8:35592–35605
89. Shao S, Wang P, Yan R (2019) Generative adversarial networks for data augmentation in machine fault diagnosis. *Comput Ind* 106:85–93
  90. Shi Y, Li Q, Zhu XX (2018) Building footprint generation using improved generative adversarial networks. *IEEE Geosci Remote Sens Lett* 16(4):603–607
  91. Singh VK, Rashwan HA, Romani S, Akram F, Pandey N, Sarker MMK, Saleh A, Arenas M, Arquez M, Puig D et al (2020) Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. *Expert Syst Appl* 139:112855
  92. Sun Y, Tang J, Shu X, Sun Z, Tistarelli M (2020) Facial age synthesis with label distribution-guided generative adversarial network. *IEEE Trans Inf Forensics Secur* 15:2679–2691
  93. Taigman Y, Polyak A, Wolf L (2016) Unsupervised cross-domain image generation. *arXiv preprint arXiv:1611.02200*
  94. Tseng BW, Wu PY (2020) Compressive privacy generative adversarial network. *IEEE Trans Inf Forensics Secur* 15:2499–2513
  95. Wang C, Xu C, Yao X, Tao D (2019a) Evolutionary generative adversarial networks. *IEEE Trans Evol Comput* 23(6):921–934
  96. Wang H, Tao C, Qi J, Li H, Tang Y (2019b) Semi-supervised variational generative adversarial networks for hyperspectral image classification. In: *IGARSS 2019-2019 IEEE International geoscience and remote sensing symposium*, IEEE, pp 9792–9794
  97. Wang J, Yang Z, Zhang J, Zhang Q, Chien WTK (2019c) Adabalgan: an improved generative adversarial network with imbalanced learning for wafer defective pattern recognition. *IEEE Trans Semicond Manuf* 32(3):310–319
  98. Wang K, Wan X (2019) Automatic generation of sentimental texts via mixture adversarial networks. *Artif Intell* 275:540–558
  99. Wang P, Hou B, Shao S, Yan R (2019d) Ecg arrhythmias detection using auxiliary classifier generative adversarial network and residual network. *IEEE Access* 7:100910–100922
  100. Wang Q, Fan H, Zhu L, Tang Y (2018a) Deeply supervised face completion with multi-context generative adversarial network. *IEEE Signal Process Lett* 26(3):400–404
  101. Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B (2018b) High-resolution image synthesis and semantic manipulation with conditional gans. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 8798–8807
  102. Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y, Change Loy C (2018c) Esrgan: Enhanced super-resolution generative adversarial networks. In: *Proceedings of the European conference on computer vision (ECCV)*
  103. Yr Wang, Sun Gd, Jin Q (2020) Imbalanced sample fault diagnosis of rotating machinery using conditional variational auto-encoder generative adversarial network. *Appl Soft Comput* 92:106333
  104. Wang Z, Simoncelli EP, Bovik AC (2003) Multiscale structural similarity for image quality assessment. In: *The Thrity-Seventh Asilomar conference on signals, systems & computers*, 2003, IEEE, vol 2, pp 1398–1402
  105. 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
  106. 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* 29(8):2337–2348
  107. Wu J, Zhang C, Xue T, Freeman B, Tenenbaum J (2016) Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In: *Advances in neural information processing systems*, pp 82–90
  108. Wu R, Gu X, Tao X, Shen X, Tai YW, Jia J (2019) Landmark assisted cyclegan for cartoon face generation. *arXiv preprint arXiv:1907.01424*
  109. Xiang P, Wang L, Wu F, Cheng J, Zhou M (2019) Single-image deraining with feature-supervised generative adversarial network. *IEEE Signal Process Lett* 26(5):650–654
  110. Xu T, Zhang P, Huang Q, Zhang H, Gan Z, Huang X, He X (2018) Attngan: fine-grained text to image generation with attentional generative adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1316–1324
  111. Yan K, Chong A, Mo Y (2020) Generative adversarial network for fault detection diagnosis of chillers. *Build Environ* 172:106698
  112. Yanagi R, Togo R, Ogawa T, Haseyama M (2019) Query is gan: Scene retrieval with attentional text-to-image generative adversarial network. *IEEE Access* 7:153183–153193
  113. Yang S, Xie L, Chen X, Lou X, Zhu X, Huang D, Li H (2017) Statistical parametric speech synthesis using generative adversarial networks under a multi-task learning framework. In: *2017 IEEE Automatic speech recognition and understanding workshop (ASRU)*, IEEE, pp 685–691
  114. Yang W, Hui C, Chen Z, Xue JH, Liao Q (2019a) Fv-gan: finger vein representation using generative adversarial networks. *IEEE Trans Inf Forensics Secur* 14(9):2512–2524
  115. Yang Y, Dan X, Qiu X, Gao Z (2020) Fggan: feature-guiding generative adversarial networks for text generation. *IEEE Access* 8:105217–105225
  116. Yang Z, Chen Y, Le Z, Fan F, Pan E (2019b) Multi-source medical image fusion based on wasserstein generative adversarial networks. *IEEE Access* 7:175947–175958
  117. Ye F, Zhu F, Fu Y, Shen B (2019) Ecg generation with sequence generative adversarial nets optimized by policy gradient. *IEEE Access* 7:159369–159378
  118. Yoo D, Kim N, Park S, Paek AS, Kweon IS (2016) Pixel-level domain transfer. In: *European conference on computer vision*, Springer, pp 517–532
  119. You S, You N, Pan M (2019) Pi-rec: progressive image reconstruction network with edge and color domain. *arXiv preprint arXiv:1903.10146*
  120. Yu C, Wang W, Yan J (2020a) Self-supervised animation synthesis through adversarial training. *IEEE Access* 8:128140–128151
  121. Yu H, Li Z, Zhang G, Liu P, Wang J (2020b) Extracting and predicting taxi hotspots in spatiotemporal dimensions using conditional generative adversarial neural networks. *IEEE Trans Veh Technol* 69(4):3680–3692
  122. Yu L, Zhang W, Wang J, SeqGAN YY (2016) Sequence generative adversarial nets with policy gradient. *arxiv e-prints*, page. *arXiv preprint arXiv:1609.05473*
  123. Yu W, Chang T, Guo X, Wang X, Liu B, He Y (2020c) Ugan: unified generative adversarial networks for multidirectional text style transfer. *IEEE Access* 8:55170–55180
  124. Yuan J, He Z (2020) Adversarial dual network learning with randomized image transform for restoring attacked images. *IEEE Access* 8:22617–22624
  125. Zhang C, Wu L, Wang Y (2019a) Crossing generative adversarial networks for cross-view person re-identification. *Neurocomputing* 340:259–269
  126. 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
  127. Zhang H, Xu T, Li H, Zhang S, Wang X, Huang X, Metaxas DN (2018) Stackgan++: realistic image synthesis with stacked generative adversarial networks. *IEEE Trans Pattern Anal Mach Intell* 41(8):1947–1962
  128. Zhang H, Goodfellow I, Metaxas D, Odena A (2019b) Self-attention generative adversarial networks. In: *International conference on machine learning*, PMLR, pp 7354–7363

129. Zhang W, Li X, Jia XD, Ma H, Luo Z, Li X (2020a) Machinery fault diagnosis with imbalanced data using deep generative adversarial networks. *Measurement* 152:107377
130. Zhang X, Zhou Y, Pei S, Zhuge J, Chen J (2020b) Adversarial examples detection for xss attacks based on generative adversarial networks. *IEEE Access* 8:10989–10996
131. Zhang Y, Miao S, Mansi T, Liao R (2020c) Unsupervised x-ray image segmentation with task driven generative adversarial networks. *Med Image Anal* 62:101664
132. Zhao B, Li W, Gong W (2019a) Deep pyramid generative adversarial network with local and nonlocal similarity features for natural motion image deblurring. *IEEE Access* 7:185893–185907
133. Zhao J, Mathieu M, LeCun Y (2016) Energy-based generative adversarial network. arXiv preprint [arXiv:1609.03126](https://arxiv.org/abs/1609.03126)
134. Zhao J, Zhang J, Li Z, Hwang JN, Gao Y, Fang Z, Jiang X, Huang B (2019b) Dd-cycleGAN: unpaired image dehazing via double-discriminator cycle-consistent generative adversarial network. *Eng Appl Artif Intell* 82:263–271
135. Zheng J, Song W, Wu Y, Xu R, Liu F (2019) Feature encoder guided generative adversarial network for face photo-sketch synthesis. *IEEE Access* 7:154971–154985
136. Zhou Z, Wang Y, Guo Y, Qi Y, Yu J (2019) Image quality improvement of hand-held ultrasound devices with a two-stage generative adversarial network. *IEEE Trans Biomed Eng* 67(1):298–311
137. Zhu D, Xia S, Zhao J, Zhou Y, Jian M, Niu Q, Yao R, Chen Y (2020) Diverse sample generation with multi-branch conditional generative adversarial network for remote sensing objects detection. *Neurocomputing* 381:40–51
138. Zhu JY, Krähenbühl P, Shechtman E, Efros AA (2016) Generative visual manipulation on the natural image manifold. In: *European conference on computer vision*, Springer, pp 597–613
139. 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
140. 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
141. Zhuang H, Zhang W (2019) Generating semantically similar and human-readable summaries with generative adversarial networks. *IEEE Access* 7:169426–169433

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