Generative Adversarial Networks: A Primer for Radiologists

Jelmer M. Wolterink, PhD
Anirban Mukhopadhyay, PhD
Tim Leiner, MD, PhD
Thomas J. Vogl, MD, PhD
Andreas M. Bucher, MD
Ivana Išgum, PhD

Artificial intelligence techniques involving the use of artificial neural networks—that is, deep learning techniques—are expected to have a major effect on radiology. Some of the most exciting applications of deep learning in radiology make use of generative adversarial networks (GANs). GANs consist of two artificial neural networks that are jointly optimized but with opposing goals. One neural network, the generator, aims to synthesize images that cannot be distinguished from real images. The second neural network, the discriminator, aims to distinguish these synthetic images from real images. These deep learning models allow, among other applications, the synthesis of new images, acceleration of image acquisitions, reduction of imaging artifacts, efficient and accurate conversion between medical images acquired with different modalities, and identification of abnormalities depicted on images. The authors provide an introduction to GANs and adversarial deep learning methods. In addition, the different ways in which GANs can be used for image synthesis and image-to-image translation tasks, as well as the principles underlying conditional GANs and cycle-consistent GANs, are described. Illustrated examples of GAN applications in radiologic image analysis for different imaging modalities and different tasks are provided. The clinical potential of GANs, future clinical GAN applications, and potential pitfalls and caveats that radiologists should be aware of also are discussed in this review.

The online slide presentation from the RSNA Annual Meeting is available for this article.

Abbreviations: AI = artificial intelligence, CNN = convolutional neural network, GAN = generative adversarial network

RadioGraphics 2021; 41:840–857

https://doi.org/10.1148/rg.2021200151

Content Codes: AI IN

From the Department of Applied Mathematics, Faculty of Electrical Engineering, Mathematics and Computer Science, Technical Medical Centre, University of Twente, Zilverling, PO Box 217, 7500 AE Enschede, the Netherlands (J.M.W.); Department of Biomedical Engineering and Physics (J.M.W., I.I.) and Department of Radiology and Nuclear Medicine (I.I.), Amsterdam University Medical Center, Amsterdam, the Netherlands; Department of Informatics, Technische Universität Darmstadt, Darmstadt, Germany (A.M.); Department of Radiology, Utrecht University Medical Center, Utrecht, the Netherlands (T.L.); and Institute of Diagnostic and Interventional Radiology, Universitätsklinikum Frankfurt, Frankfurt, Germany (T.J.V., A.M.B.). Recipient of a Certificate of Merit award for an education exhibit at the 2019 RSNA Annual Meeting. Received June 8, 2020; revision requested September 22 and received October 20; accepted January 15, 2021. For this journal-based SA-CME activity, the authors A.M., T.L., A.M.B., and I.I. have provided disclosures (see end of article); all other authors, the editor, and the reviewers have disclosed no relevant relationships. **Address correspondence to** J.M.W. (e-mail: *j.m.wolterink@utwente.nl*).

SA-CME LEARNING OBJECTIVES

After completing this journal-based SA-CME activity, participants will be able to:

- Describe how GANs work and can be used for image synthesis.
- Recognize potential applications of GANs in the radiologic workflow.
- Discuss the pitfalls and caveats of using GANs for image-to-image translation with unpaired training data.

See rsna.org/learning-center-rg.

Introduction

In recent years, there has been a tremendous increase in the number of publications describing the application of artificial intelligence (AI) to problems in radiology. This has resulted in the formation of new conferences, journals, and companies that are focused on bringing AI into the clinical workflow, and the first randomized clinical trials demonstrating the potential of such methods (1). These developments have been largely driven by the use of deep learning techniques that involve the application of artificial neural networks. Among the most successful of such networks are convolutional neural networks (CNNs), which are applied to medical images.

Successful applications of CNNs include the detection, segmentation, and quantification of pathologic conditions (2). These tasks require discriminative models that can distinguish among different classes or categories of images. Recently, AI-based applications have started to include those involving the use of generative models. These are models that can be used to synthesize new data. The most widely used generative models are generative adversarial networks (GANs) (Fig 1) (3). A lot of interest in GANs has been generated owing to the highly realistic images that they can synthesize. Similar to the

TEACHING POINTS

- The key idea behind a GAN is that there are two CNNs that are adversaries in a game. The first CNN is called the generator, and its goal is to synthesize images that resemble real images. The second CNN is called the discriminator, and its goal is to differentiate the synthetic images from the real images.
- A GAN is an example of a generative model. Such a model learns a distribution of images, thereby allowing a user to generate new images from this distribution.
- Conditional GANs allow the synthesis of images based on a condition, which could be the disease that should be visible on the images or outlines of structures that should be shown.
- One limitation of GANs is the concern about introducing false disease.
- Future applications of GANs in imaging are likely to include cross-modality image synthesis, improved detection of abnormalities, and synthesis of newly obtained images for training radiologists.

application of CNNs, the application of GANs to medical images has much potential for operations in radiology. For example, GANs have led to visually convincing results in applications such as imaging artifact reduction (4) and cross-modality image synthesis (5).

Future clinical AI applications will likely involve the use of GANs. Hence, it is important that end users—specifically, radiologists and other medical experts who rely on images—have an understanding of the underlying mechanisms and potential pitfalls and caveats of these models. Herein, we provide a primer on GANs for radiologists. The aim of this article is not to provide a full technical introduction (6) or review (7–9) of GANs in medical image analysis but rather to provide a practical introduction to this topic for radiologists. We assume that the reader is familiar with several key concepts related to deep learning; most important among these concepts are CNNs and the training of these networks.

The uninitiated reader is referred to primers on deep learning aimed at radiologists (10,11) or more seminal works on deep learning (12,13). The structure of this review is as follows: In the next section, the concept of adversarial training in GANs and how GANs can be used in image analysis is introduced. A section in which we review the most important groups of clinical applications in which GANs can play a role follows. The final section summarizes the key concepts of this review and addresses future applications as well as some of the pitfalls and caveats of GANs.

Generative Adversarial Networks

The key idea behind a GAN is that there are two CNNs that are adversaries in a game (3). The first CNN is called the generator, and its goal is to synthesize images that resemble real images. The second CNN is called the discriminator, and its goal is to differentiate the synthetic images from the real images. These two networks are trained together so that the generator CNN learns to synthesize visually convincing images.

A GAN is an example of a generative model. Such a model learns a distribution of images, thereby allowing a user to generate new images from this distribution (Fig 2). In the case of radiologic images, this means that the generative model can replicate the images that it is trained with, but it can also synthesize new images that share characteristics with the images in the training dataset. While the actual image synthesis happens in the generator, we consider the GAN as a whole—that is, the combination of generator and discriminator—to be a generative model. The discriminator is an inseparable part of this generative model, as the generator would not be able to learn the data distribution without it.

Being a generative model sets GANs apart from purely discriminative models, which are much more commonly used in radiologic applications. Discriminative models are trained to distinguish between two or more classes, such as radiographs showing pneumonia and radiographs without signs of pneumonia. A trained discriminative model can be used to classify new and unseen images, but a discriminative model cannot synthesize new images.

The use of GANs has led to tremendous advances in image synthesis. Figure 3 shows the basic setup of a GAN for image synthesis, in this case for the synthesis of natural images. We indicate the generator CNN with G and the discriminator CNN with D. Both of these CNNs are an essential part of the GAN, but the actual synthesis happens in the generator. To synthesize an image, the generator requires some input. This input, z, is a vector that is randomly drawn from a probability distribution that describes the so-called latent space. The example in Figure 3 uses a two-dimensional normal distribution as the latent space; however, in practice, a latent space can be of much higher dimensionality. Nevertheless, the dimensionality of the latent space is always significantly lower than the number of pixels in the output image of the generator, which we indicate by $G(z)$. Hence, the generator requires a way to introduce new information. A typical CNN architecture progressively reduces the image size by using downsampling and convolution layers, but a generator CNN architecture does the opposite: it increases the image size. To do this, the generator CNN architecture includes upsampling layers (14). A typical upsampling layer doubles the image size along each axis. There are multiple ways to achieve this. With the most basic approach, one

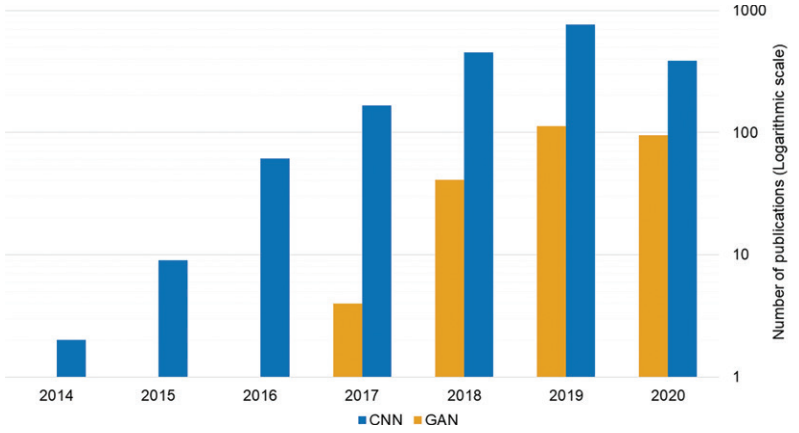

Figure 1. Graph shows the number of articles with "CNN" or "GAN" in the title, as indexed in PubMed and updated in April 2020. The number of articles with titles featuring "GAN" follows the number featuring "CNN," indicating the increasing relevance of GANs in radiologic applications.

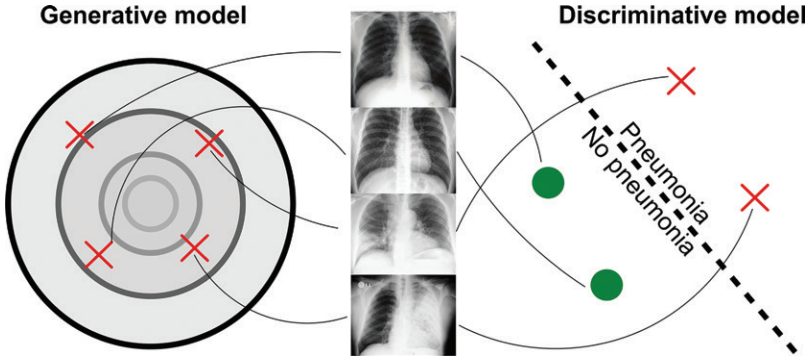

Figure 2. Diagram illustrates the conceptual difference between discriminative and generative models in deep learning. The two models use training data (the four frontal chest radiographs) in a different way. Discriminative models are trained to distinguish between two or more image classes. In this example, this means finding a decision boundary (dashed line) between chest radiographs showing pneumonia and radiographs not showing pneumonia. Generative models learn the distribution of images from each class. In this example, this would mean the distribution of chest radiographs. After training, a generative model can be used to synthesize new images from these distributions.

can simply replace each pixel with 4 (2 × 2) pixels with the same value. With a more complex but also more powerful approach, learned upsampling is used with so-called transposed convolutions (Fig 4). To synthesize a new image, the generator applies a sequence of such upsampling and convolution layers to its input vector, z.

The discriminator CNN, D, is also a CNN but one that performs classification. As input, the discriminator takes a real or synthetic image. The architecture of this CNN is more similar to that used in discriminative models that classify, for example, whether a radiograph shows or does not show pneumonia. That is, this architecture contains a sequence of downsampling and convolution layers. The output of the discriminator is a value that describes the estimated probability that the input image is real. This probability value is between 0 and 1. We write the output for a synthetic image, $G(z)$, as $D[G(z)]$ and the output for a real image, x, as $D(x)$. A discriminator that can perfectly distinguish synthetic from real images predicts that $D[G(z)] = 0$ and $D(x) = 1$.

Training a GAN

The stages in GAN development and use are identical to those in other deep learning models. First, there is a training stage in which a training dataset is used to optimize the parameters of the model. Then, there is a testing stage, in which the trained model is validated and eventually deployed. In the training stage, the generator and discriminator are jointly trained by using a single training set consisting of real images and the probability distribution from which input vectors to the generator are drawn. This means that the thousands or millions of parameters in both CNNs are modified at the same time to optimize an objective function. This is done iteratively, in thousands of small steps or iterations. To understand the training process of a

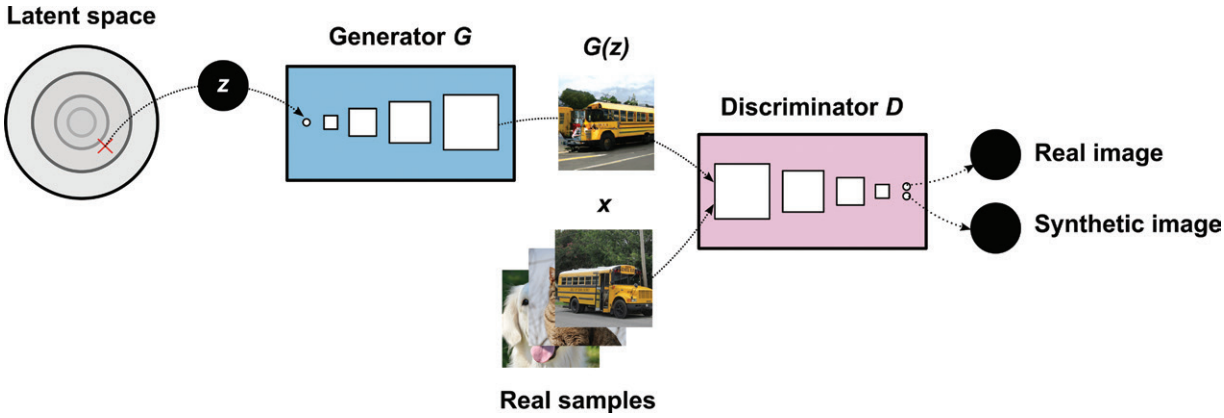

Figure 3. Diagram illustrates a GAN. A vector, *z,* is randomly sampled from a latent space, which in this case is a two-dimensional normal distribution. This vector is used as input to a generator CNN, *G,* which transforms the vector into an image, *G(z),* by using a sequence of upsampling and convolution layers (rectangular blocks). The discriminator CNN, *D,* uses a sequence of downsampling and convolution layers (rectangular blocks) to classify each image as either real or synthetic on the basis of a training data set of real images to which it has access and from which a real image, *x,* can be drawn. These networks are optimized jointly so that the generator CNN learns to synthesize increasingly realistic images.

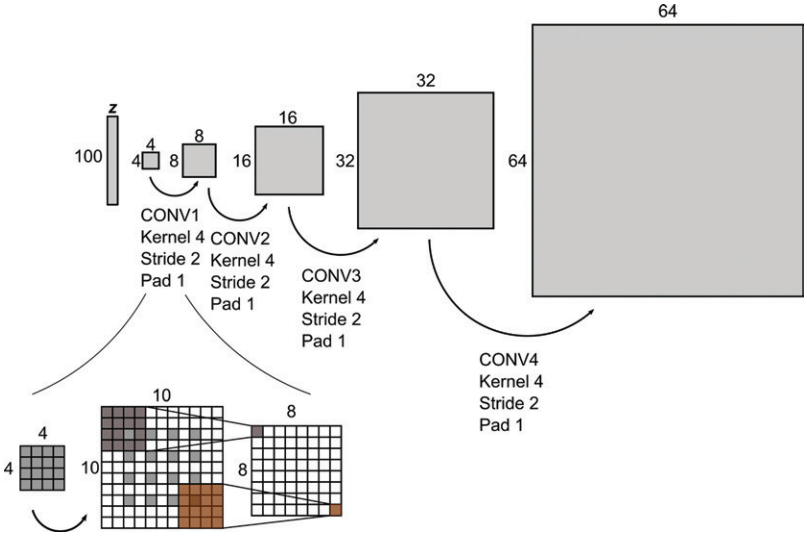

Figure 4. Diagram illustrates a generator CNN architecture that uses transposed convolutions to synthesize a 64 × 64-pixel image on the basis of the popular deep convolutional GAN (DCGAN) architecture (14). The input to the generator is a vector, *z,* with 100 elements drawn from a latent space. This vector is projected into a 4 × 4-pixel image, and four transposed convolutional layers (*CONV1–CONV4*), each of which doubles the size of the image, are used. Each kernel has a stride of 2 pixels and a kernel size of 4 pixels. Inset (bottom left) shows the details for transposed convolutional layer 1. The layer first adds spacings between the elements and spacing around the image, resulting in a 10 × 10-pixel image. Consequently, a 4 × 4-pixel kernel is moved over the image, resulting in an 8 × 8-pixel image. In practice, these operations are highly optimized.

GAN, it is important to consider how the training objectives of the generator CNN and discriminator CNN are intertwined.

We first consider the objective of the discriminator CNN, *D*. This CNN tries to make a distinction between images by minimizing a loss function. For this, the value of the probability that the CNN outputs for each input image should be similar to the target label of that image. Recall that the target for a real image, *x,* is 1. For such an image, the discriminator maximizes $\log D(x)$. By predicting a probability value of 1 for a real image, *x*, the value of this term becomes 0: $\log D(x) = \log(1) = 0$. Conversely, a synthetic image, $G(z)$, has a target label of 0, and for such an image, the discriminator maximizes $\log\{1 - D[G(z)]\}$. By predicting a probability value of 0 for a synthetic image, $G(z)$, the value of this term also becomes 0:

$$\log\{1 - D[G(z)]\} = \log(1 - 0) = \log 1 = 0. \quad (1)$$

Hence, a discriminator that predicts a probability value of 0 for all synthetic images and a probability value of 1 for all real images incurs a loss of 0, which is optimal in this case.

At the same time, the generator tries to make sure that for synthetic images, the discriminator predicts a value higher than 0. Hence, it aims to minimize $\log\{1 - D[G(z)]\}$. The higher the value for the probability that the discriminator outputs for a synthetic image, the lower the value of $1 - D[G(z)]$. To achieve this, the generator should synthesize realistic images. A more realistic synthetic image, $G(z)$, is more difficult for the discriminator to distinguish, and, thus the discriminator is less likely to predict a probability value of 0 for such an image.

Putting both these objectives into a single mathematical equation gives the standard objective

function of a GAN (3). This shared objective function $V(D, G)$ is written as follows:

$$\min_{G}\max_{D} V(D,G) = E_{x\sim Pdata}\log[D(x)] + E_{z\sim Pz}\log\{1-D[G(z)]\}. \quad (2)$$

The min and max terms at the beginning of the equation indicate that the generator, *G*, aims to minimize this objective function, while the discriminator, *D*, aims to maximize it. The notation *E* indicates the expected value, which means that this objective is not based on a single data point but rather on an average of many sampled data points. On the right-hand side of the equation—that is, after the equal sign—there are two terms indicating such expected values. The first term, $E_{x\sim Pdata}$, corresponds to images drawn from the real image distribution, P_{data}. The second term, $E_{z\sim Pz}$, corresponds to vectors drawn from the latent space distribution, P_z. This objective function forms the basis from which all GAN models are derived.

GANs are trained iteratively. In each training iteration, a minibatch of points in the latent space is randomly sampled. Each of these points is transformed into a synthetic image by the generator. The same number of real images is then randomly selected from the dataset of real images. The discriminator is shown the combined minibatch of synthetic and real images and makes a prediction for each image—for example, that it is or is not a real image. The output values of the discriminator are compared with the target labels of the images, and thus the loss for the discriminator is computed on the basis of the full minibatch of real and synthetic images. According to this loss, small changes are made to the parameters of the discriminator to take a step toward minimizing its loss function.

Similarly, we use the predictions of the discriminator on only the synthetic images to determine the loss for the generator. On the basis of this loss, small changes are made to the parameters of the generator. In the next iteration, we repeat this process, but this time with a freshly picked set of random points in the latent space and newly sampled real images. This process is repeated for hundreds or thousands of iterations.

To get a better understanding of this training process, it is important to realize which information each network has access to. The discriminator sees the synthetic images as well as the real training images, while the generator does not have access to the training images. Hence, the only information that the generator receives is the decision of the discriminator regarding the synthetic data that it generates.

We can visualize this process by looking at the loss for the discriminator and generator for a synthetic output image of the generator. In Figure 5,

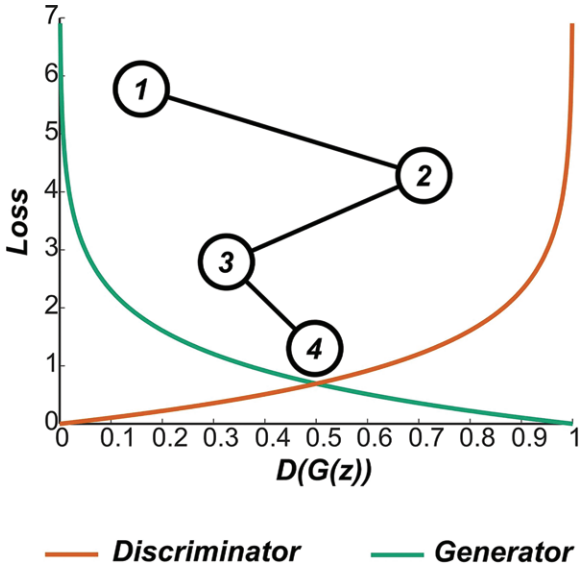

Figure 5. The loss for the discriminator, *D*, and generator, *G*, based on the discriminator output, *D[G(z)]*, for a synthetic image shows how the two networks are adversaries. *1*, If *D* predicts a low-output value, its loss is low but that of *G* is high. *2*, If *D* predicts a high-output value, its loss is high but that of *G* is low. *3, 4*, During training, both the discriminator and the generator iteratively get better at their respective tasks *(3)* until an equilibrium at which the discriminator cannot distinguish real from synthetic images is reached, thus predicting *D[G(z)]* = 0.5 for synthetic images *(4)*.

we can see how the decision of the discriminator network on a synthetic image affects both the generator and the discriminator. During training, the generator initially synthesizes images that do not resemble the real data and that the discriminator can easily classify as synthetic (1). However, the generator gets feedback from the discriminator about the images that it synthesizes and can adjust its weights accordingly so that new images are harder to distinguish (2). As the images synthesized by the generator become more realistic, the discriminator finds it more difficult to distinguish synthetic from real images and has to get better at making this distinction (3). Consequently, the generator has to synthesize better images, and this process repeats until, ideally, an equilibrium is reached in which the discriminator cannot differentiate between real and synthetic images (4).

During the training of a GAN, the image that the generator CNN synthesizes for a fixed point, *z*, in the latent space could change in each iteration. However, once the training of a GAN has finished, all CNN parameters are frozen. The generator has learned a fixed mapping between the latent space and images, meaning that providing the same input vector, *z*, to the generator, *G*, will always result in the same image: *G(z)*.

Different input vectors ideally correspond to different output images. Points that are close to

Figure 6. State-of-the-art GANs can generate high-resolution photograph-realistic portraits of people who do not exist. The five realistic photographs of nonexistent persons shown here were generated by a StyleGAN model (17).

each other in the latent space are more likely to lead to images that share certain characteristics (14). However, in some cases, many points in the latent space could correspond to the same image. This problem is called mode collapse and affects the diversity of the images that the generator can synthesize.

Alternative Objectives

Joint training of two CNNs in a GAN is, by definition, more difficult than training each CNN individually. Thus, in practice, GANs can be challenging to optimize, with problems like mode collapse occurring. To address this issue, variants of the objective function that provide a better metric to describe the difference between real and synthetic images have been proposed. Among the most popular criteria are those based on the Wasserstein distance. The Wasserstein distance describes the optimal transport solution between the distribution of real images and the distribution of synthetic images. As an analogy for this optimal transport solution, one could consider each distribution as a pile of dirt. The Wasserstein distance quantifies the amount of work that would be required to change one pile of dirt into the other pile, which is a product of the amount of dirt that would need to be displaced and the distance that this dirt would have to be displaced. Consequently, this distance is also referred to as the earth mover's distance. GANs using a Wasserstein distance are called Wasserstein GANs and are the current de facto standard for GAN training (15,16).

Evaluation

State-of-the-art models for image synthesis are based on a combination of advanced CNN architectures, smart training schemes, and discriminator objectives. Figure 6 shows an example of portraits synthesized with a state-of-the-art style-based GAN (StyleGAN) model (17). This GAN is trained by using a large dataset of portraits of living people and thus learns the distribution of human faces. The trained model can be used to synthesize new portraits of people who do not exist. All portraits shown in Figure 6 are purely synthetic and do not correspond to any living person. Note how the faces look realistic; yet, on closer inspection, there are details on the images that reveal that these are artificial portraits. For example, the clothing on the fourth photograph (man wearing glasses) contains a small artifact. As a human observer, it may not be too difficult to pick up these small mistakes. However, quantifying how realistic the images generated by a GAN are is not a trivial task. One way to estimate this quantification is to use a third CNN, in addition to the generator and discriminator, that has been trained to categorize images by using a large dataset of images. The assumption is that if this CNN can confidently classify the generated image, it is more likely to resemble a realistic image. If the CNN cannot do this, then the image is less likely to look like a real image. This metric is called the inception score, named after the inception CNN architecture that is often used for the third CNN (18).

Alternatively, one can look at the features that are extracted in the intermediate layers of this pretrained inception CNN. If the distribution of features extracted for the synthetic images resembles that of the features extracted for the real images, then the generated images are considered to be more realistic. This metric is called the Frèchet inception distance (19).

Conditional GANs

The standard definition of a GAN does not infer much control over the outputs. For example, it is not possible to ask a GAN for an image that shows a particular pathologic condition if the training set contains many different diseases. The GAN has only learned what a real image looks like and has no understanding of the different pathologic conditions shown on the images. To overcome this issue, both the generator and the discriminator can be conditioned, or trained, to include additional information (20). A GAN that follows this approach is called a conditional GAN. The training set of a conditional GAN not only includes real images but also has a label for each of these images—for example, a label describing the disease depicted on an image. The

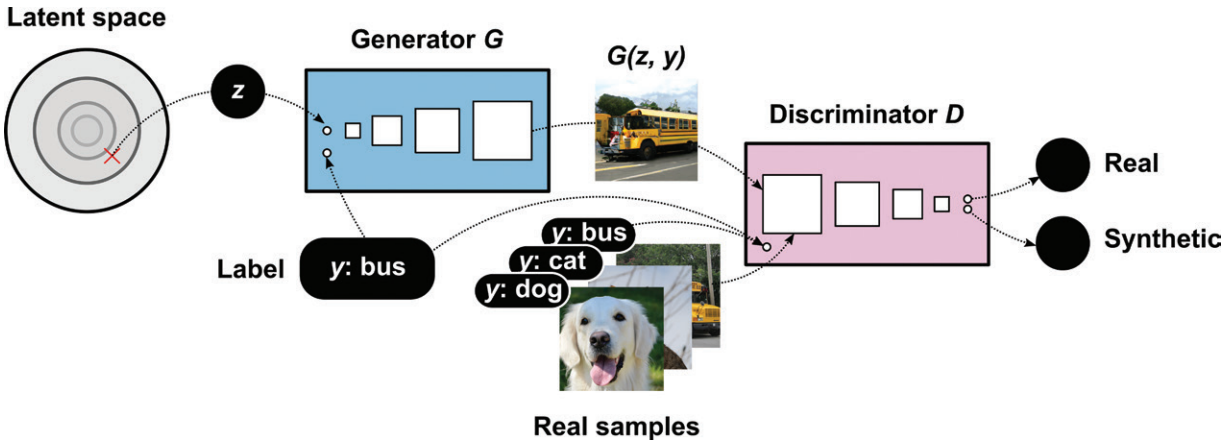

Figure 7. Diagram illustrates a conditional GAN. In addition to the latent space vector, *z*, the generator uses conditional information, *y*, as input—in this case, the desired object in the image. To synthesize a convincing image of a bus, it should synthesize an image that the discriminator considers to be a plausible image of a bus. The discriminator has access to a training dataset of images and labels that indicate the objects on the images—for example, "bus," "cat," "dog." Thus, it learns the characteristics of real images, as well as the correspondence between images and labels. The synthetic image of a bus, shown here as *G(z, y)*, was generated by using a conditional BigGAN model (21).

discriminator CNN, then, not only is trained to decide whether an image comes from the real dataset but also now has to determine for each combination of an image and a label whether it is a valid combination.

Thus, the question answered by the discriminator CNN is no longer "Is this a real image?" but rather "Is this a real image, and does it match what the label prescribes?" The discriminator CNN is trained to answer this question not only for real images but also for images synthesized by the generator. This also makes the task of the generator CNN more complex. The generator now has to synthesize an image that looks realistic *and* shows what is described in the image label.

Figure 7 shows an example of a conditional GAN model. Input to the generator, *G*, is still a point in the latent space, but now in combination with a label, which is "bus" in this example. Based on these two inputs, the generator has to synthesize an image (21). For all samples in the real dataset, there is now also a label such as "dog," "cat," or "bus" available. Both the image and its (desired) label are input to the discriminator, *D*, which decides whether this is a plausible image for the given label. To include this additional input information, modifications to the CNN architectures of both the generator and the discriminator are required.

Conditional GANs are trained in much the same way that normal GANs are trained, with the exception that the loss function is modified to force the discriminator and generator to take the image labels into account. During each training iteration, the discriminator is presented with minibatches of randomly selected real images and synthetic images based on randomly drawn latent space points and random labels. State-of-the-art conditional GANs such as BigGANs (big GANs) (21) can synthesize images from more than a thousand different classes.

Conditioning on Images

While the GAN in Figure 7 is conditioned on image labels, a GAN could be conditioned on anything, such as continuous values and image captions (22). A very valuable consequence of this for radiologic applications is that GANs can even be used to synthesize images conditioned on other images. In this case, we train the generator to synthesize a new image on the basis of an input image. Consequently, the discriminator determines for pairs of images whether they form a realistic combination. This makes it possible to use GANs for problems such as deblurring, colorization, and segmentation, all of which can be considered image-to-image translation problems. *Image-to-image translation* is an umbrella term for tasks in which an image in a source domain—for example, the domain of gray-scale images—is translated into a corresponding image in a target domain—for example, the domain of color images. As we describe later in the "Clinical Applications" section, many radiologic image analysis tasks can be considered image-to-image translation tasks.

Depending on the available training data, there are two main approaches to training GANs for image-to-image translation: with use of paired training data, or without use of paired training data. In paired training data problems, perfectly aligned pairs of images in both domains are available. A popular approach to such problems is the pix2pix framework (23). Figure 8 shows

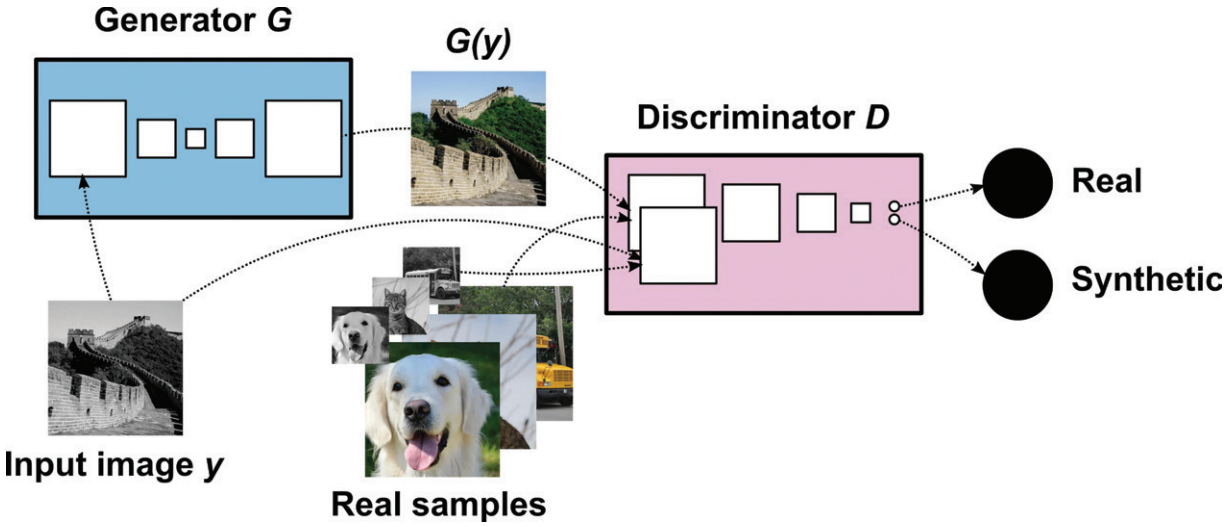

Figure 8. Example of an image-to-image translation model. The generator transforms the input (gray-scale) image, *y,* into a colorized image, *G(y)*. For this transformation, the generator uses an image-to-image CNN, such as encoder-decoder architecture (rectangular blocks). The discriminator takes the combination of both images as input and determines whether this is a feasible combination. To do this, the discriminator has access to a training set with pairs of images.

an example of this framework applied to an image-to-image translation task—namely, image colorization. Given a gray-scale input image, the generator, *G,* transforms the image into a colorized image. The real dataset consists of paired gray-scale and color images, which align perfectly because the transformation of color images into gray-scale images is trivial. The generator learns how to transform a gray-scale image into a color image, and the discriminator determines for each combination of a gray-scale and a color image whether this is a realistic image pair. Moreover, there is always a reference colorized image for the input gray-scale image. This allows the use of an additional loss term to optimize the generator, *G*—namely, one that quantifies the difference between the output color image and the reference color image. This way, the generator learns to synthesize images that both resemble real color images and maintain the content of the input image.

For the successful application of pix2pix-like models, perfectly aligned training images are required. However, in many cases, in radiologic applications in particular, no paired training data are available. For example, when MR and CT images obtained in patients are available, these images in two domains may originate from the same patient, but there is no pixel-perfect alignment between them because they were acquired in two separate acquisitions. This is an unpaired data setting in which a pix2pix-like model breaks down: The discriminator cannot inspect pairs of images, and no pixel-wise loss can be computed between the output of the generator, *G,* and a target image. Consequently, the discriminator can only determine whether the output image belongs to the second domain; however, there is no way to guarantee that the semantic content of the output image corresponds to that of the input image. A risk of this is that the image in the second domain will show different content. In radiology, this means that disease could be accidentally introduced or removed. To mitigate this risk, the GAN needs to be regularized during the training process; this means that additional loss terms are included during training. There are two ways to regularize a GAN and thus encourage the content of the images to stay the same even when no pairs of training images are available.

The first approach to unpaired training data is to enforce the pixel-wise correspondence between the input and output images to be strong. This can be achieved by minimizing the difference between these two images (24). However, this only works in cases in which the images are expected to be similar before and after translation, such as in image denoising or artifact removal.

The second, more generic and popular approach is to use multiple models in a cycle-consistent GAN, or CycleGAN, setup (25). Instead of two CNNs, cycle-consistent GANs use CNN model cycles, whereby each cycle contains two generator models and a discriminator model (Fig 9). The first generator model, G_B, translates data from the source domain, *A,* to the target domain, *B*. A discriminator model, D_B, in domain *B* determines how well the synthesized data match real images in that domain. Note that this is similar to a GAN without additional regularization. The regularization is provided by

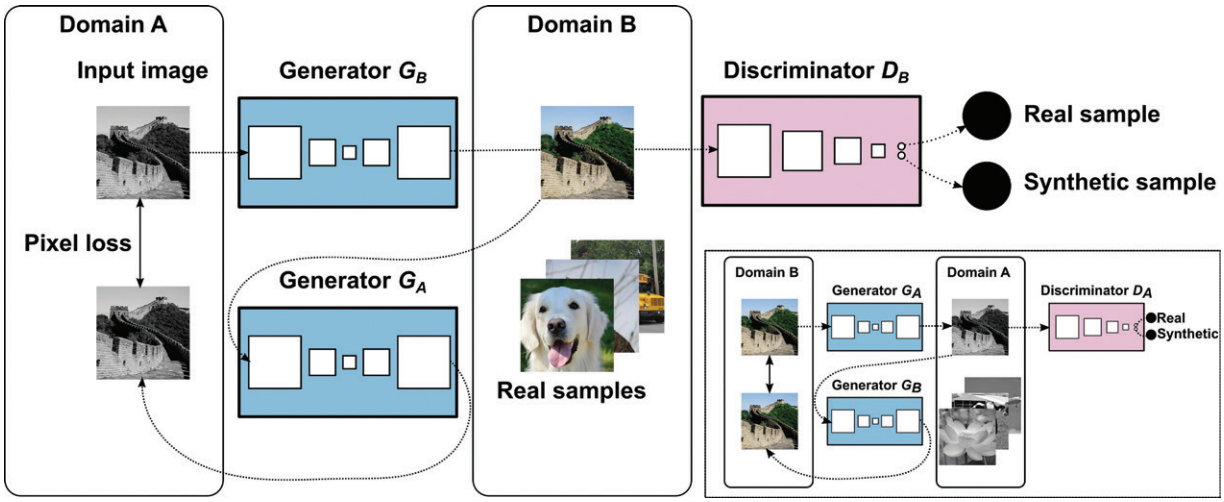

Figure 9. Diagram illustrates a cycle-consistent GAN, which combines a discriminator loss with a pixel-wise cycle-consistency loss. As in a conditional GAN, in a cycle-consistent GAN, an image is transformed from domain *A* to domain *B*. The synthetic image in domain *B* is then transformed back to domain *A* by a second generator. The difference between the initial image and the twice-transformed image is minimized. Bottom right inset: A second, backward cycle is also optimized during training.

the second generator, G_A, which translates the image back to domain A. The goal is to make the resulting image match the original image as closely as possible. If this can be achieved, it is assumed that the corresponding synthetic image in domain B at least contained the semantic content required to reconstruct the original image. To improve stability, a counterclockwise cycle is trained at the same time. This counterclockwise cycle uses the same generators and an additional discriminator, D_A, in domain A to translate data from domain B to domain A and then back to domain B.

Note that when using GANs, there is a lot of flexibility to adapt, for example, the CNN architectures of the generator and discriminator models, the objective of the discriminator, and additional regularization loss terms such as cycle-consistency loss to the problem one aims to solve. The GANs described in this section form the basis of many of the applications in the next section; however, more often than not, domain-specific design choices are made to best tackle a clinical problem. Similarly, as for other deep learning models, the number of required training images and the amount of required computing time depend heavily on the problem that is being solved.

Clinical Applications

As with deep learning in general, GANs have been applied to a wide range of problems in the field of radiology. To give the reader an idea of the applicability of GANs in the radiologic workflow, we provide examples of applications in image synthesis, image reconstruction, cross-domain synthesis, image analysis, and pseudohealthy synthesis (Fig 10) (26). For systematic reviews of all GAN applications in radiology, the user is referred to various sources in the literature (7–9).

Image Synthesis

In supervised deep learning, the need for a sufficiently large and well-annotated high-quality dataset remains a bottleneck, especially when tackling rare diseases and new applications for which such a dataset is not yet available. One common strategy for addressing this issue is to use data augmentation such as translation, rotation, scaling, and flipping of available samples to create "new" samples. In addition to these deterministic transformations, GANs allow the synthesis of completely new images to enlarge datasets. Just as the images in Figure 6 show portraits of people who do not exist, GANs can be used to generate medical images that do not exist. A prerequisite for the effective use of synthetic images is that they visually resemble real images. In several studies, blinded observer studies have been performed to verify that synthetic medical images visually resemble real images. For example, in a random blinded comparison of GAN-generated synthetic brain MR images and real MRI series, a group of radiologists and nonspecialized readers were unable to reliably distinguish synthetic images from real images (27).

Synthetic samples for deep learning are more useful when a reference label along with an image is provided—that is, when data are conditioned on a label. Conditional GANs (Fig 7) allow the synthesis of images based on a condition, which could be the disease that should be visible on the images or outlines of structures that should be shown. An example of this is the

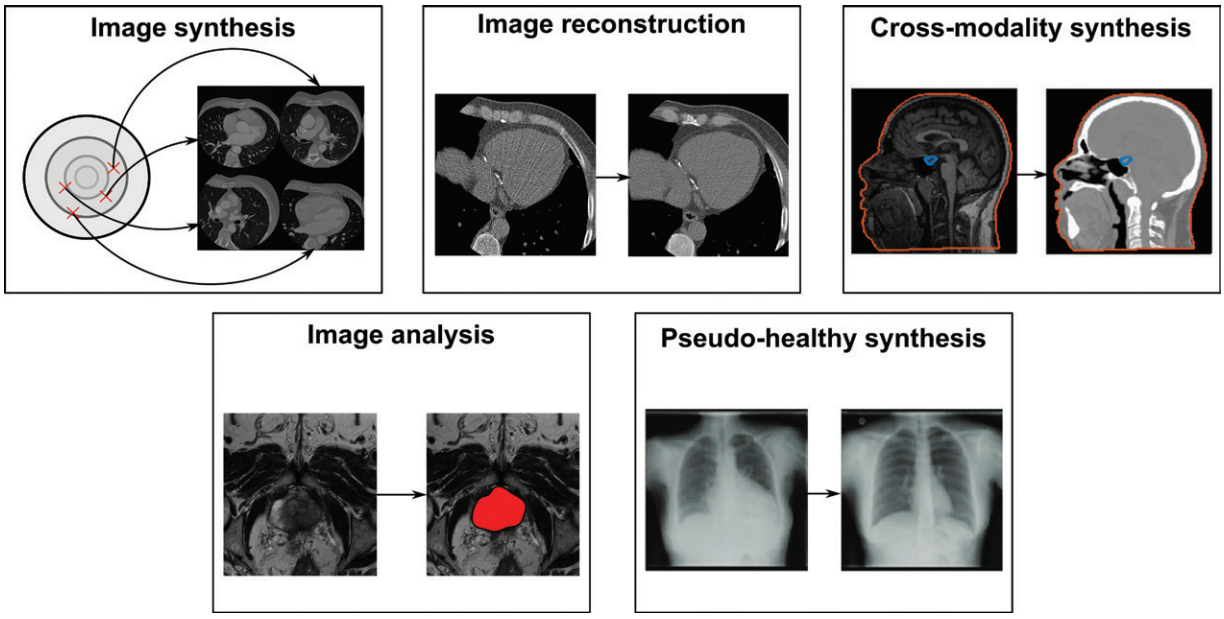

Figure 10. The applications of GANs for various syntheses in radiology are illustrated on axial CT images (top left and middle), sagittal head MR (left) and CT (right) images (top right), axial prostate MR images (bottom left), and frontal chest radiographs (bottom right. (Frontal chest images reprinted, with permission, from reference 26.) GANs have been widely used in combination with radiologic images. Here, applications are grouped into those that synthesize completely new data (image synthesis), those that improve the reconstruction of images and remove artifacts (image reconstruction), those that translate a medical image from one modality into another (cross-modality synthesis), those that improve existing image analysis techniques such as segmentation (image analysis), and those that allow the removal of disease findings from images to highlight abnormalities (pseudohealthy synthesis).

work by Frid-Adar et al (28), who used a deep convolutional GAN to synthesize additional CT images of cysts, metastases, and hemangiomas of the liver. A similar idea has been used to synthesize benign and malignant lung nodules at CT with a Wasserstein GAN (29).

Alternatively, image synthesis could be conditioned on the output of a weaker image simulator. In the nonmedical domain, this principle was used to improve the photograph realism of simulated images (24). In the medical domain, existing simulators for medical images include the XCAT model, which allows the synthesis of CT sections that are anatomically correct but lack the look and feel of "real" CT images. Hence, a cycle-consistent GAN (Fig 9) could be used to add a realistic representation to the XCAT phantom's simplified anatomy (30).

Despite their potential visual quality, de novo synthesized samples are of limited practical value if they do not enhance the performance of a downstream classifier model. Hence, the next step is to use the synthesized data to enlarge the training dataset for a deep learning–based method and thus improve its performance. The classification accuracy can then be easily tested in a comparative way between the enlarged dataset and a model trained only on the base dataset. Frid-Adar et al (28) showed how GAN-generated liver lesions could boost the classification sensitivity of a CNN from 78.6% to 85.7% and boost the classification specificity from 88.4% to 92.4%. Similarly, a CNN was pretrained with 60 000 synthetic lung nodule images and then trained on a relatively small dataset of 60 real images of lung nodules. This significantly enhanced the performance of the CNN classifier, from accuracies of 51.9% (benign nodules) and 84.9% (malignant nodules) to 66.7% and 93.9%, respectively (29). Russ et al (30) showed how the addition of synthetic data to a real dataset could improve the accuracy of a deep learning–based segmentation method for blood vessels. In addition, Salehinejad et al (31) showed how a CNN that is trained with real as well as synthetic chest radiographs outperforms a CNN trained with only real images.

Image Reconstruction

Medical image acquisition and reconstruction often include a trade-off between diagnostic image quality and adverse factors such as increased acquisition times and higher radiation dose. In practice, a compromise that puts realistic constraints on image acquisition parameters is chosen. However, this compromise could also lead to reduced image quality in the form of high noise levels, motion artifacts, and other artifacts. Moreover, in cases such as metal implant imaging at CT and patient motion, artifacts would persist regardless of the acquisition parameters used. In these scenarios, GANs could bridge the gap

between technical limitations and desirable image quality while minimizing adverse imaging effects.

The dose levels used at CT have substantially decreased in the past decades owing to advances in hardware and reconstruction techniques. In recent years, deep learning–based reconstruction of CT images with CNNs has contributed to these advances (32). One challenge in CNN-based CT denoising is the definition of an adequate loss function. Minimization with quantitative criteria such as the mean squared error might lead to overly smoothed images with an unrealistic or "blotchy" appearance, similar to those often criticized in iterative reconstruction techniques (33). Instead, a discriminative loss could encourage the preservation of texture details so that the appearance and noise statistics of denoised CT images resemble those of real CT images. On cardiac CT images, denoising by a single generator network with discriminator feedback allowed coronary calcium scoring on CT images acquired with an 80% radiation dose reduction (Fig 11) (4). Alternatively, two generator and discriminator networks for the low-dose and routine-dose domains, respectively, could be used in a cycle-consistent GAN to transform between these two domains (34).

Like CT images, MR images can be subject to deterioration by noise. Ran et al (35) simulated paired samples of reconstructed noisy and noise-free brain MR images and used these to train a pix2pix-like model in which the generator outputs a denoised image for a noisy input image. For the acceleration of MRI acquisitions, the (under)sampling of k space exploited in compressed sensing techniques is a key manipulatable factor. Deep learning has further accelerated compressed sensing, allowing even higher speed-ups with improved image quality (36). One of the first applications of GANs to compressed sensing involved the use of a discriminator loss in combination with a cyclic data-consistency loss to transform zero-filled MRI reconstructions based on undersampled k-space data into fully sampled images (37).

Some of the most exciting applications of GANs address limitations of image acquisition that would otherwise necessitate a hardware innovation such as detector resolution or motion tracking. For example, similar to applications for nonmedical images (38), a GAN could be trained for image super-resolution, increasing image matrix sizes above those originally acquired (39). Here, the input image of the generator network would be a low-resolution image, and the output image of that network would be a high-resolution image. Alternatively, motion artifacts could be removed from MR images by using a conditional GAN when it is trained with paired images (40) or by using a cycle-consistent GAN when it is trained with unpaired images (41). This approach has the potential to rescue otherwise impeded studies.

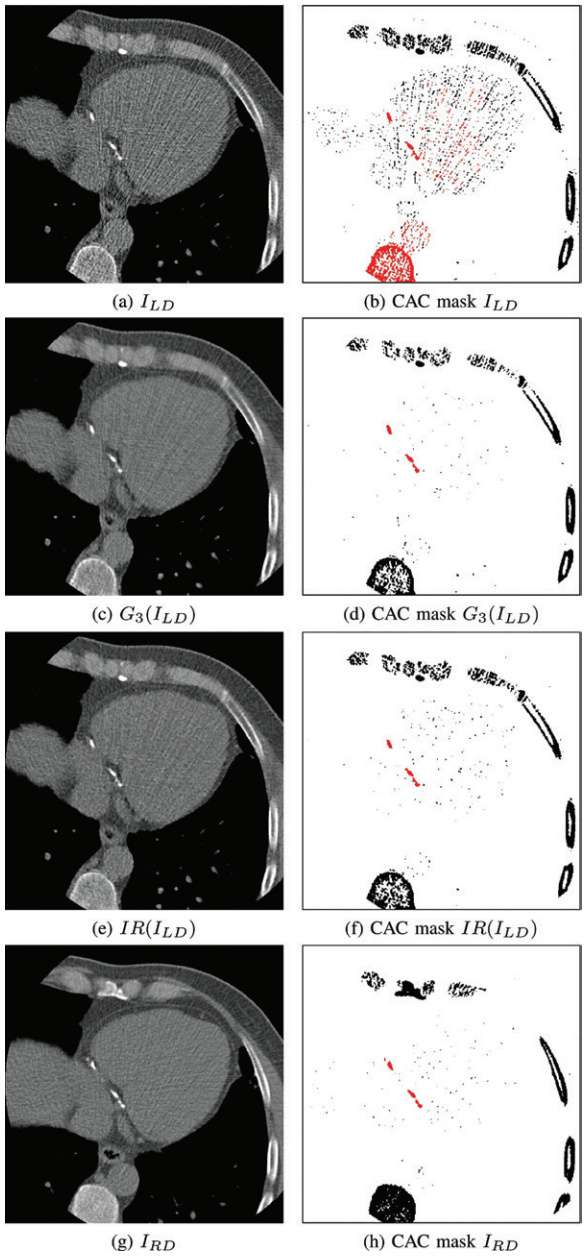

Figure 11. Axial CT and coronary artery calcium (CAC) mask images. **(a)** Low-dose image *(I_{LD})* acquired with 20% of the routine dose. **(c)** GAN-denoised *(G_3)* low-dose image. **(e)** Low-dose image processed with conventional iDose iterative reconstruction *(IR)* technique. **(g)** Routine-dose image *(I_{RD})*. **(b, d, f, h)** Corresponding CAC mask images show a mask (>130 HU) for coronary calcium scoring (black) and calcified lesions identified with manual scoring (red). The GAN-based **(c, d)** and iterative reconstruction **(e, f)** techniques enabled calcium scoring on the low-dose image. (Reprinted, with permission, from reference 4.)

Cross-Modality Synthesis

With the large variety of imaging modalities available to radiologists, some images intentionally or

unintentionally may not be acquired for a patient. To some extent, GANs allow the synthesis of missing image modalities on the basis of those images that are acquired. Possible benefits include time, radiation, and cost savings. A generator CNN can be trained to transform an image of one modality (the source domain) into an image of another modality (the target domain). Such a transformation is typically nonlinear, and a discriminator could be used to encourage characteristics of the target domain on the output image. An example of this is the synthesis of multiple MRI sequences from a single MR image acquisition, such as T2-weighted images from T1-weighted images, with use of a conditional GAN (5).

Relevant-use cases for cross-modality synthesis can be found in radiation therapy treatment planning, in which an MR image for tissue delineation and a CT image for electron density estimation are typically acquired. In MRI-only–guided radiation therapy treatment planning, the CT image is replaced with a synthetic CT image derived from the MR image acquisition (42). Nie et al (43) proposed using a GAN approach for brain MR image–to–brain CT image synthesis, in which a discriminative loss is combined with a pixel-wise loss. Notably, their model is different from a pix2pix model in that the discriminator is not conditioned on the input image. Largent et al (44) performed a systematic comparison of conditional GANs and standard CNNs with different loss functions for pelvic MR image–to–pelvic CT image synthesis. While conditional GANs such as pix2pix require pixel-perfect alignment between images in both domains, this is hardly ever the case with real medical images. Therefore, the use of cycle-consistent GANs has been popular for MR image–to–CT image synthesis (Fig 12) (45).

Conversely, Lei et al (46) used cycle-consistent GANs for CT-only guided radiation therapy planning—namely, to omit MR image acquisition but obtain tissue delineations at CT. Given the poor visibility of the prostate at CT, in this work, CT images were first transformed into synthetic MR images by using a cycle-consistent GAN; then, a pretrained MRI prostate segmentation CNN was applied to these images. Similarly, image conversion cycle-consistent GANs have been used to enlarge training sets for the segmentation of cardiac structures on MR and CT images (47).

Image Analysis

Deep learning has had a large effect on traditional postprocessing methods in image analysis such as image segmentation and image registration (2). In image segmentation, the performance of algorithms is typically evaluated by using overlap measures such as the Dice similarity coefficient. For a particular segmentation, these overlap measures can be high, but the anatomic plausibility of the segmentation, and thus its clinical value, could be low. The challenge is that criteria such as anatomic plausibility are hard to quantify in an objective function used to optimize a CNN (Fig 13a). This is where GANs come in: The discriminator network can be used to determine what is a good segmentation and what is not on the basis of a reference dataset of real segmentations (50).

Moeskops et al (48) proposed one of the first applications of GAN-based segmentation on medical images. A discriminator was trained to differentiate between automatic and reference segmentation masks on brain MR images (48). As image segmentation is a paired training data problem, the discriminator loss could be combined with a conventional pixel-wise segmentation loss (Fig 13b). The trained model made fewer small spurious segmentation errors, as these were penalized by the discriminator model, resulting in anatomically more plausible segmentations. Further applications of conditional pix2pix–like models for segmentation can be found in knee cartilage structure segmentation from MRI (51). A slightly different approach to GAN-based image segmentation is proposed in the popular SeGAN model (Fig 13c) (49). In this work, the discriminator is called a critic, as it does not provide a binary classification but rather aims to quantify the differences between predicted and reference segmentations. Xue et al (49) found that this model achieved state-of-the-art performance in tumor segmentation at brain MRI.

Similar to image segmentation, image alignment, or registration, is currently undergoing a deep learning renaissance (52,53), including novel methods that involve the use of GANs. As in segmentation, deep learning–based registration is guided by a loss function based on a metric of the alignment between two images. The lower this loss, the better images are aligned. Several groups have proposed the use of a discriminator CNN to distinguish well-aligned images from poorly aligned images and thus provide a registration loss that is different from the ones typically used (54,55).

Pseudohealthy Synthesis

In pseudohealthy synthesis, a medical image obtained in a patient with a disease is translated into a corresponding image that shows what the patient's imaging findings would look like if the disease were not present. This allows the visualization of disease on an image and the detection of abnormalities that could be related to the disease. Training data for such problems are by definition unpaired, as it is highly unlikely that aligned pairs containing normal-finding and disease-depicting

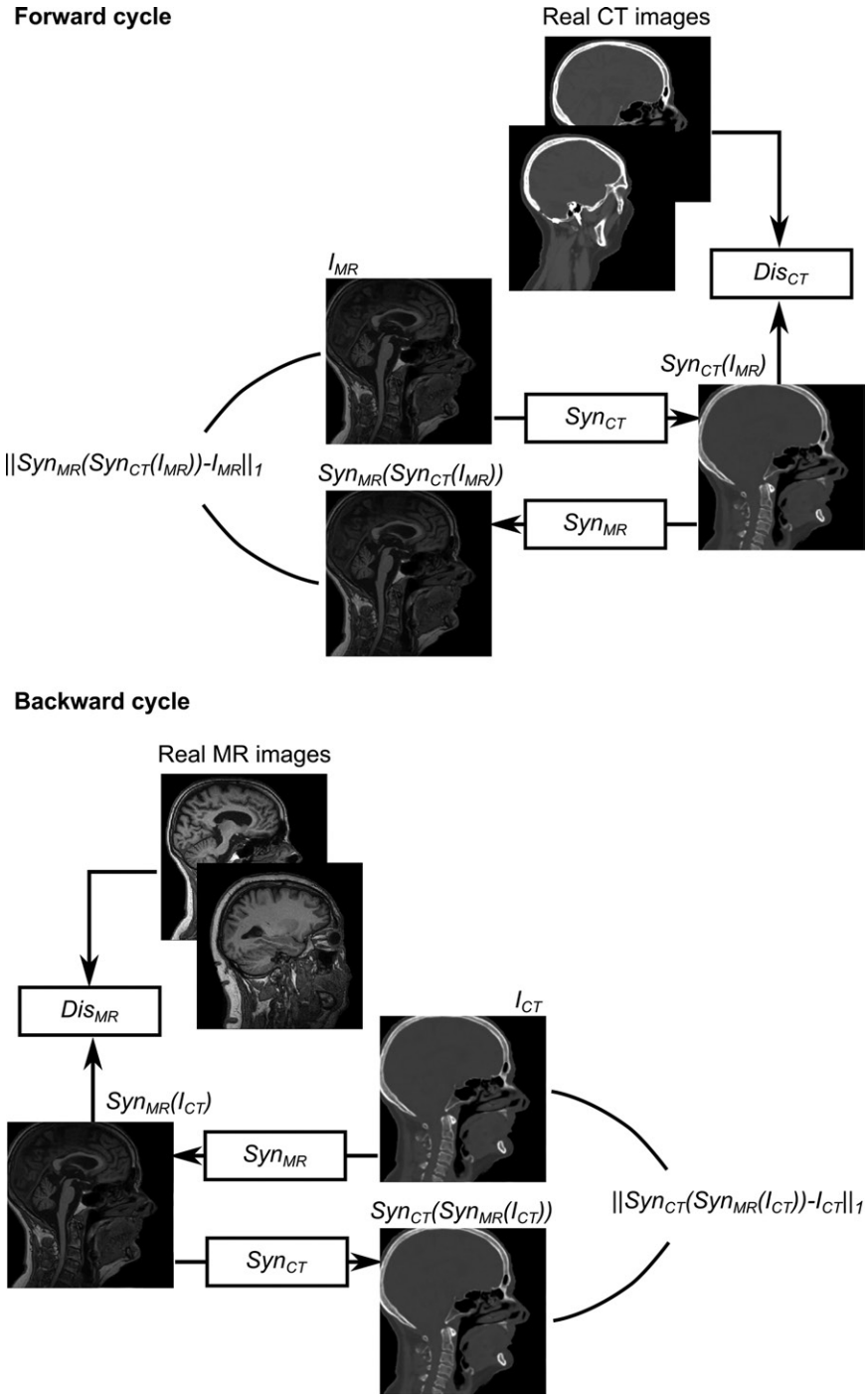

Figure 12. Diagram illustrates how a cycle-consistent GAN can be used to transform brain MR images (I_{MR}) into brain CT images (I_{CT}), and vice versa. The generator models, Syn_{CT} and Syn_{MR}, perform the cross-domain synthesis, while the domain-specific discriminator models, Dis_{MR} and Dis_{CT}, aim to distinguish real from synthetic images. (Reprinted, with permission, from reference 45.)

images obtained in the same patient are present to form training samples. Therefore, data from healthy subjects are combined with data on patients in whom the disease is present.

One example is the work by Baumgartner et al (56), who trained a Wasserstein GAN model to transform brain MR images in patients with Alzheimer disease into images that show the findings that would be seen in these patients if Alzheimer disease were not present. By subtracting the resulting image from the original image, the effect of disease on the brain could be visualized. Similarly, a pseudohealthy synthesis model was trained by using radiographs with and without congestive heart failure to visualize the effect of the disease. As expected, the model

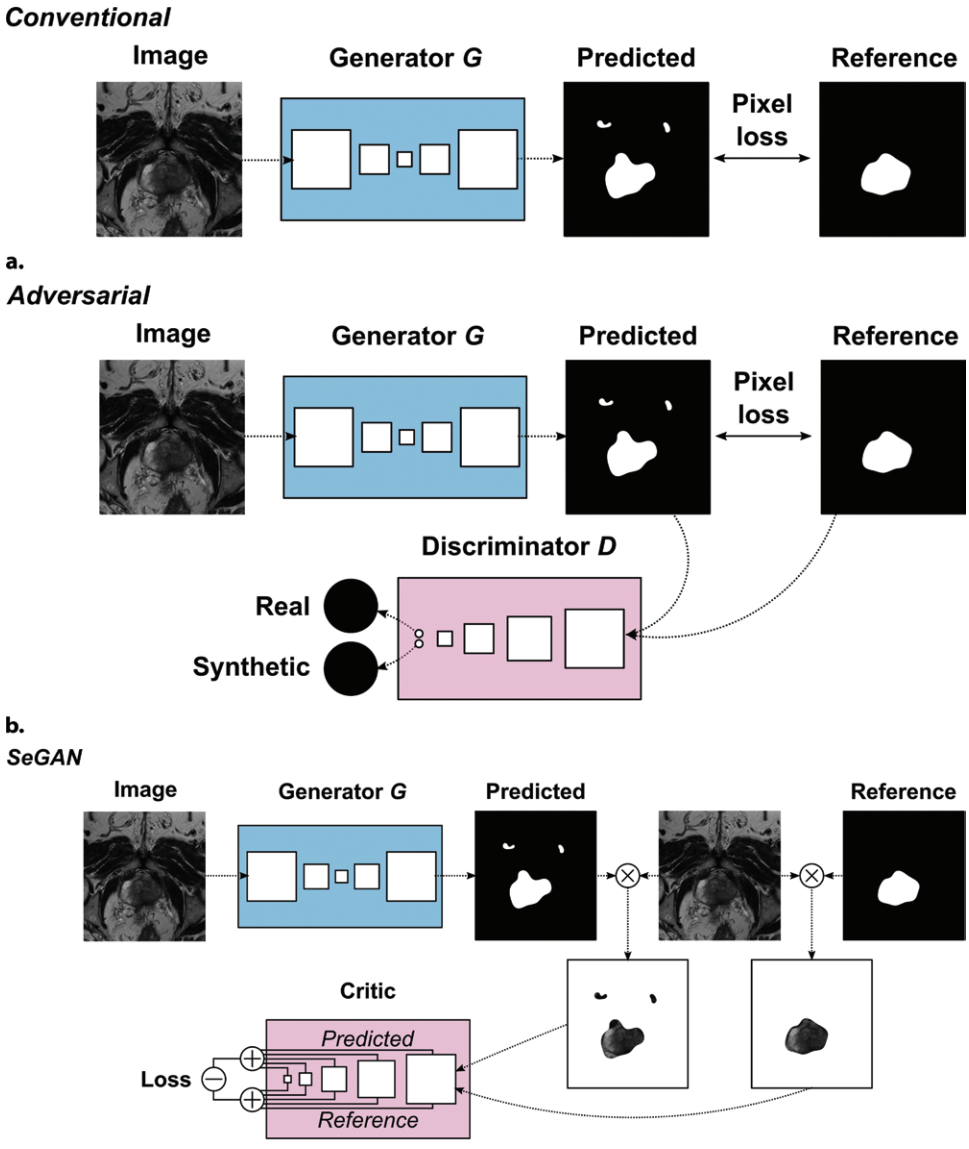

Figure 13. Diagrams illustrate three deep learning–based segmentation approaches for prostate segmentation in MRI. **(a)** With the conventional deep learning approach, a pixel-wise loss (eg, binary cross entropy) is computed between the predicted mask and the reference mask. **(b)** With the adversarial approach, a discriminator is additionally trained to distinguish real from synthetic segmentation masks (48). **(c)** With the *SeGAN* approach, the input image is multiplied by the predicted mask and reference mask. The discriminator is considered a critic that extracts features at multiple scales to compute a multiscale loss that converges to 0, as the masked images are identical (49).

shrank enlarged cardiac silhouettes on the synthesized images (26).

A third example of this is the work by Van Velzen et al (57), who used a cycle-consistent GAN to translate chest CT images obtained in patients with calcified atherosclerotic plaque into images without plaque, highlighting the location of calcified lesions. This selective image modification by means of training on coarse disease labels is a previously unreachable feat that is now made possible by GANs. Using a GAN to first generate images with normal-appearing findings from images with abnormal findings enabled image segmentation, tumor classification, and tumor generation on datasets of healthy patients. This is the inverse process to pseudohealthy synthesis (58). This technique can be used to train for the detection of lesions that are otherwise rarely encountered. It also opens up a field of interactively manipulated imaging datasets that can convincingly substitute for their real counterparts.

Discussion

GANs are powerful deep learning models. In the past few years, it has been shown that training a generator CNN jointly with a discriminator CNN can lead to improvements in radiology tasks such as image synthesis, cross-modality image conversion,

and pseudohealthy synthesis. The synthesis of images from other images has the potential to directly affect patient care, as fewer images need to be acquired and the information present on acquired images could be better exploited. Improvements in image acquisition and reconstruction based on GANs could affect the imaging time in the setting of MRI or the irradiation dose in the setting of CT. The use of GAN-based image analysis methods such as segmentation and registration could improve the accuracy and efficiency of many existing clinical tasks.

However, as with all deep learning models, several factors influence the effectiveness of GANs. As with any deep learning model, predictions will be suboptimal if the input data that are provided are not from the same distribution as the input data used to train the model. Questions about the instability of deep learning methods for inverse problems such as image denoising also apply to GANs (59). For instance, in the case of MR image reconstruction from undersampled data, minor perturbations in the measured signal could lead to large changes in the reconstructed image. These effects might not be directly visible and could lead to medical images that do not correspond to a patient's actual anatomy or pathologic condition. To detect the risk of such suboptimal predictions, efforts in uncertainty quantification for other deep learning techniques might also be applied to GANs. For example, studying the agreement of an ensemble of cycle-consistent GANs can aid in the detection of unsuitable input CT images (60).

Moreover, ethical and legal issues that underlie the use of deep learning in clinical applications are as relevant to GANs as they are to other deep learning approaches. Thus, for the foreseeable future, GAN-generated results should be checked by expert radiologists, and fully automatic applications will be limited to certain tasks. Like other applications, GAN-based applications should be well validated (61).

Perhaps the most relevant risk when using GANs for radiologic applications is that of hallucination. The synthesis of new data by itself raises risks of unintentionally synthesizing content that does not exist, or conversely, of removing relevant information from an image. GANs such as cycle-consistent GANs that are trained with unpaired data are particularly susceptible to these risks, as only an indirect check to verify that the synthesized image shows the same content is performed. There have been concerns about the use of cycle-consistent GANs for image-to-image translation, particularly when there is a mismatch between the distribution of disease in both domains. This might result in the spurious suggestion, or "hallucination," of disease (62).

One limitation of GANs is the concern about introducing false disease. In a series of experiments, Cohen et al (63) showed that a cycle-consistent GAN will remove a brain tumor from images during fluid-attenuated inversion-recovery MR image to T1-weighted MR image synthesis if none of the T1-weighted training MR images show tumors. Conversely, tumors will always be inserted if the T1-weighted training set contains only images with tumors. To mitigate this, the location of such structures could be explicitly factored in (64).

When using GANs for image-to-image translation, the generator CNN captures information about the relationship between the two domains. For example, when using a GAN for MR image–to–CT image synthesis, the generator CNN may locally make a direct mapping between MR image signal intensity values and CT attenuation values but also add nonlocal patterns that make the output image look like a CT image. Thus, a single MR image does not necessarily contain all of the information that a CT image contains, but the combination of that MR image with the information captured in the CNN can lead to realistic CT images.

Generator CNNs can also encode certain information into their output images. It has been found that cycle-consistent GAN generators hide information on images. This hidden information is subsequently used to transform the image back to its original domain in a process called steganography (62). This process could lead to visually unnoticeable differences between synthesized and real images in the target domain that could affect downstream image analysis.

An outstanding issue that arises when using GANs in medical image analysis is that of evaluation. While quantitative metrics such as the peak signal-to-noise ratio or mean square error are often used as both training objective and evaluation criteria, the downstream task should also be considered. For example, to use a CT denoising method, lesions must still be present on the denoised image. Scores based on pretrained CNNs for natural images such as the Frèchet inception distance (19) and inception (18) scores may not translate directly to medical images.

While this review is focused on GAN applications for medical images, other medical applications for using GANs for nonimaging data such as electrocardiographic signals (65) and electronic health records (66) have been found. Future applications of GANs in imaging are likely to include cross-modality image synthesis, improved detection of abnormalities, and synthesis of newly obtained images for training radiologists. Improvements in computer vision tend to find their way into medical image analysis, and it

is not unlikely that computational challenges will be overcome and image synthesis models will be developed to synthesize high-resolution three-dimensional medical images of patients who do not exist, akin to the portraits in Figure 6.

Moreover, advances in GANs for image analysis will go hand in hand with developments in CNN architectures. Finally, GANs are likely to be used to further accelerate and improve MR and CT image acquisitions, and the acquisition of other images such as US scans.

Conclusion

Deep learning has facilitated many incredible feats in the more tedious parts of radiologists' tasks. In the past few years, it has been shown that GANs are powerful deep learning models and that training a generator CNN jointly with a discriminator model can lead to improvements in a broad range of radiology tasks. These include image synthesis, cross-domain image synthesis, and abnormality detection. While there are limitations to what GANs can achieve reliably, when used with care, they could enable many exciting new applications of AI in radiology.

***Disclosures of Conflicts of Interest.*—A.M.** *Activities related to the present article:* disclosed no relevant relationships. *Activities not related to the present article:* employee of Technische Universität Darmstadt, Zuse Institute, Berlin, Germany. *Other activities:* disclosed no relevant relationships. **T.L.** *Activities related to the present article:* disclosed no relevant relationships. *Activities not related to the present article:* institution receives grant from the Dutch Technology Foundation (grant P15–26) and Netherlands Heart Foundation (grant 14741), with participation from Pie Medical Imaging and Philips Healthcare; institution receives payment for lectures from Philips Healthcare, Bracco, and Bayer Healthcare; coinventor (U.S. patent no. 10,395,366) and royalty agreement with Pie Medical (no royalties received to date); cofounder and shareholder in Quantib-U (no money paid to author or institution). *Other activities:* disclosed no relevant relationships. **A.M.B.** *Activities related to the present article:* disclosed no relevant relationships. *Activities not related to the present article:* payment for lectures and travel support for presentations at 2019 RSNA Annual Meeting from Guebert, Bayer Healthcare, and Siemens Healthineers. *Other activities:* disclosed no relevant relationships. **I.I.** *Activities related to the present article:* disclosed no relevant relationships. *Activities not related to the present study:* institution received grant from Dutch Technology Foundation (P15-26, 12726), with participation of Pie Medical Imaging and Philips Healthcare; institutional research grant, Pie Medical Imaging; the Netherlands Organisation for Health Research and Development - Institutional Research - grant with participation of Pie Medical Imaging (104003009); coinventor, U.S. patent no. 10,176,575 and U.S. patent no. 10,699,407 (royalty agreement with Pie Medical Imaging; no royalties received to date); cofounder and shareholder in Quantib-U (no money paid to author or institution). *Other activities:* disclosed no relevant relationships.

References

1. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ 2020;368:m689.
2. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
3. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Advances in Neural Information Processing Systems—NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems. 2014;2:2672–2680. https://dl.acm.org/doi/10.5555/2969033.2969125. Accessed May 8, 2020.
4. Wolterink JM, Leiner T, Viergever MA, Isgum I. Generative adversarial networks for noise reduction in low-dose CT. IEEE Trans Med Imaging 2017;36(12):2536–2545.
5. Dar SUH, Yurt M, Karacan L, Erdem A, Erdem E, Cukur T. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Trans Med Imaging 2019;38(10):2375–2388.
6. Wolterink JM, Kamnitsas K, Ledig C, Išgum I. Deep learning: generative adversarial networks and adversarial methods. In: Zhou SK, Rueckert D, Fichtinger G, eds. Handbook of Medical Image Computing and Computer Assisted Intervention. London, England: Academic Press/Elsevier, 2020; 547–574. https://doi.org/10.1016/B978-0-12-816176-0.00028-4.
7. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. Med Image Anal 2019;58:101552.
8. Sorin V, Barash Y, Konen E, Klang E. Creating artificial images for radiology applications using generative adversarial networks (GANs): a systematic review. Acad Radiol 2020;27(8):1175–1185.
9. Kazeminia S, Baur C, Kuijper A, et al. GANs for Medical Image Analysis. Cornell University Computer Science Machine Learning website. https://arxiv.org/abs/1809.06222.
10. Chartrand G, Cheng PM, Vorontsov E, et al. Deep learning: a primer for radiologists. RadioGraphics 2017;37(7):2113–2131.
11. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional neural networks for radiologic images: a radiologist's guide. Radiology 2019;290(3):590–606.
12. Bishop CM. Pattern recognition and machine learning. New York, NY: Springer, 2006.
13. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, Mass: MIT Press, 2016.
14. Radford A, Metz L, Chintala S. Unsupervised representation learnig with deep convolutional generative adversarial networks. In: 4th International Conference on Learning Representations, ICLR 2016: conference track proceedings, 2016. Semantic Scholar website. https://www.semanticscholar.org/paper/Unsupervised-Representation-Learning-with-Deep-Radford-Metz/8388f1be26329fa45e5807e968a641ce170ea078#matched. Accessed May 8, 2020.
15. Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning, Volume 70. ICML'17. JMLR 2017;214–223.
16. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. Improved training of Wasserstein GANs. In: Advances in neural information processing systems 30. NIPS'17: Proceedings of the 27th International Conference on Neural Information Processing Systems. https://papers.nips.cc/paper/2017/hash/892c3b1c6dccd52936e27cbd0ff683d6-Abstract.html. Accessed May 8, 2020.
17. Karras T, Laine S, Aila T. A Style-based generator architecture for generative adversarial networks. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, June 15–20, 2019. Piscataway, NJ: IEEE, 2019; 4396–4405.
18. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training GANs. Cornell University Computer Science Machine Learning website. https://arxiv.org/abs/1606.03498. Accessed May 8, 2020.
19. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. Advances in Neural Information Processing Systems 2017;6627–6638.
20. Mirza M, Osindero S. Conditional generative adversarial nets. Cornell University Computer Science Machine Learning website. https://arxiv.org/abs/1411.1784. Published November 2014. Accessed May 8, 2020.

21. Brock A, Donahue J, Simonyan K. Large scale GAN training for high fidelity natural image synthesis. Cornell University Computer Science Machine Learning website. https://arxiv.org/abs/1809.1109. Accessed May 8, 2020.
22. Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H. Generative Adversarial Text to Image Synthesis. In: Balcan MF, Weinberger KQ, eds. In: Proceedings of The 33rd International Conference on Machine Learning, vol 48. New York, New York, 2016:1060-1069.
23. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition, 2017; 5967–5976.
24. Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W, Webb R. Learning from simulated and unsupervised images through adversarial training. In: Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition, 2017; 2242–2251.
25. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. 2017; 2242–2251.
26. Seah JCY, Tang JSN, Kitchen A, Gaillard F, Dixon AF. Chest radiographs in congestive heart failure: visualizing neural network learning. Radiology 2019;290(2):514–522.
27. Kazuhiro K, Werner RA, Toriumi F, et al. Generative adversarial networks for the creation of realistic artificial brain magnetic resonance images. Tomography 2018;4(4):159–163.
28. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 2018;321:321–331.
29. Onishi Y, Teramoto A, Tsujimoto M, et al. Automated pulmonary nodule classification in computed tomography images using a deep nonvolutional neural network trained by generative adversarial networks. BioMed Res Int 2019:6051939.
30. Russ T, Goerttler S, Schnurr AK, et al. Synthesis of CT images from digital body phantoms using CycleGAN. Int J CARS 2019;14(10):1741–1750.
31. Salehinejad H, Colak E, Dowdell T, Barfett J, Valaee S. Synthesizing chest x-ray pathology for training deep convolutional neural networks. IEEE Trans Med Imaging 2019;38(5):1197–1206.
32. Willemink MJ, Noël PB. The evolution of image reconstruction for CT: from filtered back projection to artificial intelligence. Eur Radiol 2019;29(5):2185–2195.
33. Geyer LL, Schoepf UJ, Meinel FG, et al. State of the art: iterative CT reconstruction techniques. Radiology 2015;276(2):339–357.
34. Kang E, Koo HJ, Yang DH, Seo JB, Ye JC. Cycle-consistent adversarial denoising network for multiphase coronary CT angiography. Med Phys 2019;46(2):550–562.
35. Ran M, Hu J, Chen Y, et al. Denoising of 3D magnetic resonance images using a residual encoder-decoder Wasserstein generative adversarial network. Med Image Anal 2019;55:165–180.
36. Knoll F, Zbontar J, Sriram A, et al. fastMRI: a publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. Radiol Artif Intell 2020;2(1):e190007.
37. Quan TM, Nguyen-Duc T, Jeong WK. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. IEEE Trans Med Imaging 2018;37(6):1488–1497.
38. Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. 2017; 105–114.
39. Kim KH, Do WJ, Park SH. Improving resolution of MR images with an adversarial network incorporating images with different contrast. Med Phys 2018;45(7):3120–3131.
40. Johnson PM, Drangova M. Conditional generative adversarial network for 3D rigid-body motion correction in MRI. Magn Reson Med 2019;82(3):901–910.
41. Khalili N, Turk E, Zreik M, Viergever MA, Benders MJNL, Išgum I. Generative adversarial network for segmentation of motion affected neonatal brain MRI. In: Shen D, Liu T, Peters TM, et al, eds. Medical image computing and computer assisted intervention: MICCAI 2019—lecture notes in computer science, vol 11766. Cham, Switzerland: Springer, 2019; 320–328.
42. Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. Radiat Oncol 2017;12(1):28.
43. Nie D, Trullo R, Lian J, et al. Medical image synthesis with context-aware generative adversarial networks. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins D, Duchesne S, eds. Medical image computing and computer assisted intervention: MICCAI 2017—lecture notes in computer science, vol 10435. Cham, Switzerland: Springer, 2017; 417–425.
44. Largent A, Barateau A, Nunes JC, et al. Comparison of deep learning-based and patch-based methods for pseudo-CT generation in MRI-based prostate dose planning. Int J Radiat Oncol Biol Phys 2019;105(5):1137–1150.
45. Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, van den Berg CAT, Išgum I. Deep MR to CT synthesis using unpaired data. In: Tsafaris S, Gooya A, Frangi A, Prince J, eds. Simulation and synthesis in medical imaging: SASHIMI 2017—lecture notes in computer science, vol 10557. Cham, Switzerland: Springer, 2017; 14–23.
46. Lei Y, Dong X, Tian Z, et al. CT prostate segmentation based on synthetic MRI-aided deep attention fully convolution network. Med Phys 2020;47(2):530–540.
47. Chartsias A, Joyce T, Dharmakumar R, Tsaftaris SA. Adversarial image synthesis for unpaired multi-modal cardiac data. In: Tsafaris S, Gooya A, Frangi A, Prince J, eds. Simulation and synthesis in medical imaging: SASHIMI 2017—lecture notes in computer science, vol 10557. Cham, Switzerland: Springer, 2017; 3–13.
48. Moeskops P, Veta M, Lafarge MW, Eppenhof KAJ, Pluim JPW. Adversarial training and dilated convolutions for brain MRI segmentation. In: Cardoso J, Arbel T, Carneiro G, et al, eds. Deep learning in medical image analysis and multimodal learning for clinical decision support: DLMIA 2017, ML-CDS 2017—lecture notes in computer science, vol 10553. Cham, Switzerland: Springer, 2017; 56–64.
49. Xue Y, Xu T, Zhang H, Long LR, Huang X. SegAN: adversarial network with multi-scale L1 loss for medical image segmentation. Neuroinformatics 2018;16(3-4):383–392.
50. Luc P, Couprie C, Chintala S, Verbeek J. Semantic Segmentation using adversarial networks. Cornell University Computer Science Machine Learning website. https://arxiv.org/abs/1611.08408. Published November 2016. Accessed May 29, 2020.
51. Gaj S, Yang M, Nakamura K, Li X. Automated cartilage and meniscus segmentation of knee MRI with conditional generative adversarial networks. Magn Reson Med 2020;84(1):437–449.
52. Haskins G, Kruger U, Yan P. Deep learning in medical image registration: a survey. Mach Vis Appl 2020;31:8.
53. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. Med Image Anal 2019;52:128–143.
54. Mahapatra D, Antony B, Sedai S, Garnavi R. Deformable medical image registration using generative adversarial networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, April 4–7, 2018. Piscataway, NJ: IEEE, 2018; 1449–1453.
55. Elmahdy MS, Wolterink JM, Sokooti H, Išgum I, Staring M. Adversarial optimization for joint registration and segmentation in prostate CT radiotherapy. In: Shen D, Liu T, Peters TM, et al, eds. Medical image computing and computer assisted intervention: MICCAI 2019—lecture notes in computer science, vol 11766. Cham, Switzerland: Springer, 2019; 366–374.
56. Baumgartner CF, Koch LM, Tezcan KC, Ang JX, Konukoglu E. Visual feature distribution using Wasserstein GANs. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, June 18–23, 2018. Piscataway, NJ: IEEE, 2018; 8309–8319.

57. Van Velzen S, de Vos BD, Verkooijen H, Leiner T, Viergever M, Išgum I. Coronary artery calcium scoring: can we do better? In: Išgum I, Landman BA, eds. Proceedings of SPIE: medical imaging 2020—image processing, vol 11313. Bellingham, Wash: International Society for Optics and Photonics, 2020; 113130G.
58. Sun L, Wang J, Huang Y, Ding X, Greenspan H, Paisley J. An Adversarial Learning Approach to Medical Image Synthesis for Lesion Detection. IEEE J Biomed Health Inform 2020;24(8):2303–2314.
59. Antun V, Renna F, Poon C, Adcock B, Hansen AC. On instabilities of deep learning in image reconstruction and the potential costs of AI. Proc Natl Acad Sci U S A 2020;117(48):30088–30095.
60. van Harten LD, Wolterink JM, Verhoeff JJC, Išgum I. Automatic online quality control of synthetic CTs. In: Išgum I, Landman BA, eds. Proceedings of SPIE: medical imaging 2020—image processing, vol 11313. Bellingham, Wash: International Society for Optics and Photonics, 2020; 113131M.
61. Angus DC. Randomized Clinical Trials of Artificial Intelligence. JAMA 2020;323(11):1043–1045.
62. Chu C, Zhmoginov A, Sandler M. CycleGAN, a master of steganography. Cornell University Computer Science Machine Learning website. http://arxiv.org/abs/1712.02950. Published December 2017. Accessed May 31, 2020.
63. Cohen JP, Luck M, Honari S. Distribution matching losses can hallucinate features in medical image translation. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G, eds. Medical image computing and computer assisted intervention:MICCAI 2018.l—lecture notes in computer science, vol 11070. Cham, Switzerland: Springer, 2018; 529–536.
64. Jiang J, Hu YC, Tyagi N, et al. Tumor-aware, adversarial domain adaptation from CT to MRI for lung cancer segmentation. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G, eds. Medical image computing and computer assisted intervention: MICCAI 2018—lecture notes in computer science, vol 11071. Cham, Switzerland: Springer, 2018; 777–785.
65. Shaker AM, Tantawi M, Shedeed HA, Tolba MF. Generalization of convolutional neural networks for ECG classification using generative adversarial networks. IEEE Access 2020;8:35592–35605.
66. Baowaly MK, Lin CC, Liu CL, Chen KT. Synthesizing electronic health records using improved generative adversarial networks. J Am Med Inform Assoc 2019;26(3):228–241.