Artificial Intelligence In Medicine 109 (2020) 101938

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

GANs for medical image analysis

Salome Kazeminia [a,1], Christoph Baur [b,1,*], Arjan Kuijper [c], Bram van Ginneken [d], Nassir Navab [b], Shadi Albarqouni [b,e], Anirban Mukhopadhyay [a]

[a] Department of Computer Science, TU Darmstadt, Germany
[b] Computer Aided Medical Procedures (CAMP), TU Munich, Germany
[c] Fraunhofer IGD, Darmstadt, Germany
[d] Radboud University Medical Center, Nijmegen, The Netherlands
[e] Department of Computing, Imperial College London, United Kingdom

ARTICLE INFO

Keywords:
Generative adversarial networks
Deep learning
Medical imaging
Survey`

ABSTRACT

Generative adversarial networks (GANs) and their extensions have carved open many exciting ways to tackle well known and challenging medical image analysis problems such as medical image de-noising, reconstruction, segmentation, data simulation, detection or classification. Furthermore, their ability to synthesize images at unprecedented levels of realism also gives hope that the chronic scarcity of labeled data in the medical field can be resolved with the help of these generative models. In this review paper, a broad overview of recent literature on GANs for medical applications is given, the shortcomings and opportunities of the proposed methods are thoroughly discussed, and potential future work is elaborated. We review the most relevant papers published until the submission date. For quick access, essential details such as the underlying method, datasets, and performance are tabulated. An interactive visualization that categorizes all papers to keep the review alive is available at http://livingreview.in.tum.de/GANs_for_Medical_Applications/.

1. Introduction

From the early days of medical image analysis, machine learning (ML) and artificial intelligence (AI) driven systems have been a key component for complex decision making—a brief history of which can be found in [1]. Across generations of development, the focus was mostly on decision making at different granularity levels. These techniques range from low-level pixel processing over feature engineering combined with supervised classifier learning to the recent wave of feature learning using convolutional neural networks (CNNs).

Until recently, the driving focus of the medical image analysis (MIA) community has been on the supervised learning of decision boundaries, while generative tasks have been on the back seat. This changed dramatically with the advent of generative adversarial networks (GANs) [2], which lead to a new age of generative modeling and distribution learning. With their abilities to mimic data distributions and to synthesize images at yet unprecedented levels of realism, GANs have carved open new ways to bridge the gap between *supervised learning* and *image generation*.

The mentioned improvements are essentially due to the following properties: (1) GANs maximize the probability density over the data generating distribution by exploiting density ratio estimation [3] in an indirect fashion of supervision; (2) GANs can discover the high dimensional latent distribution of data, which has led to significant performance gains in the extraction of visual features.

This review summarizes GAN-based architectures proposed for medical image processing applications published until the end of 2018.

We categorized the reviewed papers into seven categories according to the following applications: synthesis, segmentation, reconstruction, detection, de-noising, registration, and classification. The categorical distribution of the reviewed papers can be seen in Fig. 1. Methods based on GANs were applied to a variety of different medical imaging modalities such as MRI (magnetic resonance imaging), CT (computed tomography), OCT (optical coherence tomography), chest X-ray, dermoscopy, ultrasound, PET (positron emission tomography) and microscopy.

* Corresponding author.
E-mail addresses: salome.kazeminia@gris.tu-darmstadt.de (S. Kazeminia), c.baur@tum.de (C. Baur), arjan.kuijper@mavc.tu-darmstadt.de (A. Kuijper), bram.vanginneken@radboudumc.nl (B. van Ginneken), navab@cs.tum.edu (N. Navab), shadi.albarqouni@tum.de (S. Albarqouni), anirban.mukhopadhyay@gris.tu-darmstadt.de (A. Mukhopadhyay).
[1] The authors contributed equally to this work.

https://doi.org/10.1016/j.artmed.2020.101938
Received 29 October 2019; Received in revised form 30 July 2020; Accepted 7 August 2020
Available online 9 August 2020

To find the papers, we searched for the keywords "medical" and "GAN" (or "generative adversarial network") along with the aforementioned applications in Google Scholar, Semantic Scholar, PubMed, and CiteSeer. Also, we checked references and citations of selected papers. Since GANs are rather new, and a significant number of articles are still in the publication process of different journals and conferences, we covered pre-prints published in arXiv as well.

For quick access, we summarized key information on the reviewed methods, such as metrics, datasets, image modality and the employed architectures in tables. Also, we provide a live tree at http://livingreview.in.tum.de/GANs_for_Medical_Applications/ for ease of categorization of papers. Further, we thoroughly discuss the advantages and shortcomings of the methods and specify clear directions for future work.

Thus, we ended up with 79 papers, which we consider the most relevant ones covering a broad spectrum of applications and varieties of GANs. The remainder of this paper is structured as follows: In Section 3, we introduce the architecture of the GAN and its subclasses, which are used in medical image applications. In Section 4, different contributions of GANs in medical image processing applications (de-noising, reconstruction, segmentation, registration, detection, classification, and synthesis) are described, and Section 5 provides a conclusion about the investigated methods, challenges, and open directions for the employment of GANs in medical image processing.

2. Opportunities for medical image analysis

Supervised deep learning is currently the state of the art in many computer vision and medical image analysis tasks, but its success depends heavily on the large-scale availability of labeled training data. In the medical field, this is a fundamental problem as often there is a severe lack of labeled data. This *data scarcity* arises from the tedious, time-consuming and costly nature of medical image acquisition and labeling. Furthermore, in the medical field, many datasets suffer from severe *class imbalance* due to the rare nature of some pathologies. GANs can potentially alleviate these two limitations by generating realistic-looking images from an implicit distribution that follows the real data distribution. The community has recognized this potential early and has been eagerly investigating GAN's suitability for tackling these problems. Applied research on GANs for such medical image synthesis can be broadly classified into two categories: (i) unconditional and (ii) conditional image synthesis. The GAN is originally an unconditional, unsupervised generative model that generates data from noise and leaves the user with limited influence on the generated output (a review of such methods can be found in Section 4.1.1). Conditional variants have also been proposed, which allow seeding the generation process with prior information such as class labels, image attributes, or even images themselves (see Section 4.1.2).

While labeled data might be scarce in the medical field, unlabeled data is often readily available. To leverage both labeled and arbitrary amounts of unlabeled data, numerous so-called *semi-supervised deep learning* methods have been proposed throughout the last years. Recent advances in this field make use of adversarial networks, the core-concept behind GANs, to harness both labeled and unlabeled data in the training of classifiers and segmentation models. A related issue hampering deep learning is the effect of so-called *domain shift*, i.e. a shift between the training and testing data distribution. Such a shift frequently occurs in the medical field and often leads to models that generalize poorly beyond the training data, with potentially unpredictable behavior. Prominent examples are (i) MR data, where images from different device vendors exhibit very different characteristics, or (ii) histopathological images, which vary significantly due to different staining protocols and types of equipment. GAN-based domain adaptation methods have recently shown great potential to close this gap, some of which are reviewed in Sections 4.2 and 4.7.

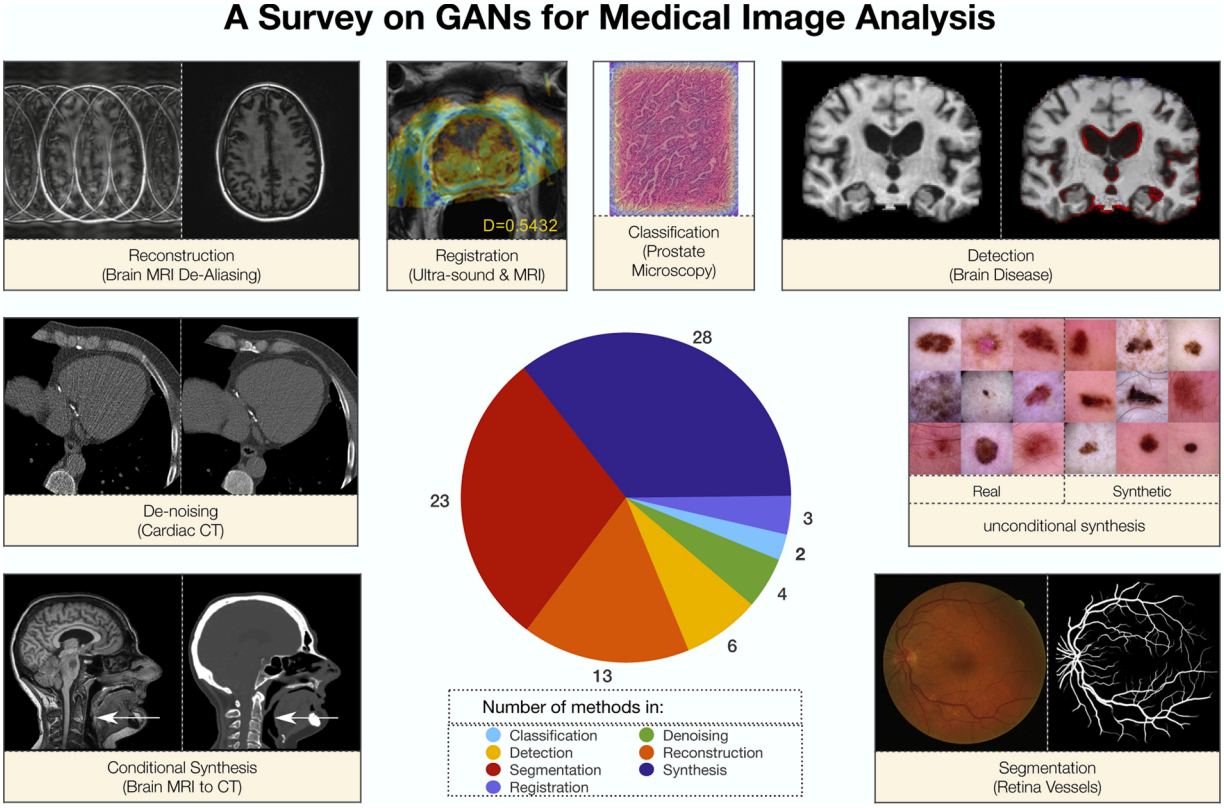

Fig. 1. The pie chart of distribution of papers and visual examples of GAN functionality among the different applications. Examples are taken from papers as the following: conditional synthesis [4], denoising [5], reconstruction [6], registration [7], classification [8], detection [9], unconditional synthesis [10], and segmentation [11].

This potential comes primarily from the introduction of an adversarial network into the overall optimization process. According to [12], such a network can be understood as a learned, rich similarity metric: traditional similarity metrics such as the ℓ_1 or ℓ_2-distance are known to produce blurry results and lack the incorporation of spatial context, whereas such a learned similarity metric allows to optimize for concepts in images beyond the pixel level. This appealing property has recently been exploited in medical image denoising (see Section 4.5), super-resolution, image-to-image translation (see Section 4.1.2) and even medical image segmentation (see Section 4.2).

The newly achieved levels of realism promoted by adversarial networks also open up new opportunities for tackling problems with high time-complexity, such as medical image reconstruction or registration. These problems are mathematically well modeled, but usually require a costly iterative optimization. GANs have been successfully used to learn a mapping from raw signals to reconstructed images (see Section 4.3) or to register medical images (see Section 4.6) in a single forward pass while ensuring coherence and high levels of realism.

3. Basic GAN models employed in medical applications

In this section, we introduce the general concept behind GANs, their conditional variants as well as a variety of prominent extensions. These extensions comprise the DCGAN, Markovian GAN, conditional GAN, CycleGAN, auxiliary classifier GAN, Wasserstein-GAN as well as least squares GAN, which are used as the inspiring models in medical applications.

In the context of this work, there are three "*adversarial*" concepts, which should be adequately understood by their different meanings. "*Adversarial attack*" as a general term means to make imperceptible changes to data such that a classifier misclassifies it. Usually, the modified sample, called "*adversarial example*" [13], is perceptually not distinguishable from its original [14]. "*Adversarial training*" proposed by [14] is a strategy that increases the robustness of neural networks against adversarial attacks by training the model with both typical and adversarial examples. Due to the state of existing neural networks at that time, implementing adversarial training was not a practical solution. The effectiveness of this idea became apparent when Goodfellow et al. employed it in GANs [2].

3.1. GAN

The GAN framework [2] consists of a training dataset **X**, whose underlying distribution we denote p_{real}, and a pair of competing networks: a generator (G) with parameters θ_G, and a discriminator (D) with weights θ_D (see Fig. 2). G aims to find a mapping $\hat{x} = G(z;\theta_G)$ that maps latent random variables $z \sim p_z(z)$—drawn from a prior distribution p_z—to generated data $\hat{x} \in \hat{\mathbf{X}}$, which is supposed to follow the distribution $p_\theta(\hat{x}|z)$. The primary goal is to optimize this mapping such that the distribution of generated data $\hat{\mathbf{X}}$ resembles the distribution of the training data **X**, i.e. $p_\theta(\hat{x}|z) \sim p_{\text{real}}$. In other words, G is supposed to generate fake data which must not be distinguishable from real data. This is achieved with the help of the discriminator network D, whose task is to classify between fake and real samples. Essentially, D is a binary classifier which yields $D(x) = 1$ for real samples and $D(\hat{x}) = 0$ for fake data. Both networks are adversaries as G attempts to gradually synthesize more and more realistic samples which D would misclassify as real, while D constantly learns to differentiate between real and synthesized samples. Mathematically speaking, D and G play a two-player minimax game with the following value function $V(G, D)$:

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

To optimize for Eq. (1), D is trained to maximize the probability of correct label assignment for fake and real data, while G is trained to trick D into thinking a generated sample is real by minimizing $log(1 - D(G(z)))$. In practice, these networks are usually implemented as multi-layer perceptrons (MLPs) or convolutional neural networks and trained with minibatch stochastic gradient descend in an alternating fashion. Once trained, it is sufficient to sample a random z and feed it through the generator to synthesize data. This "*adversarial training*" framework exhibits a few interesting properties: (1) the generator G is updated only through gradients back-propagated from the discriminator and (2) no explicit correspondences between z, x and $\hat{x}$ are required, such that input is not explicitly memorized by G. (3) Furthermore, as proven by the authors, optimization for Eq. (1) minimizes the Jensen-Shannon (JS) divergence between the distributions of real and synthetic data.

Although theoretically well-grounded, the vanilla GAN has proven to be quite hard to train. General convergence is heavily dependent on hyperparameter tuning to avoid vanishing or exploding gradients, and they are prone to *mode collapse*. This term describes a phenomenon where GANs map all z to very similar synthetic samples covering only a single mode of the data distribution. During optimization, these modes might also change (so-called *mode hopping*). A plethora of extensions and subclasses have been proposed to cope with these problems, a selection which is introduced in the following subsections.

3.2. DCGAN

To address the instability of the basic GAN architecture and increase the resolution of GANs in the synthesis of image data, [15] propose the deep convolutional GAN (DCGAN). In this model, both the generator and discriminator follow a deep convolutional network architecture, exploiting the efficacy of spatial kernels and hierarchical feature learning. Concepts such as batch-normalization and leaky-ReLU have been included to improve training stability, but issues such as mode collapse were not entirely resolved.

3.3. cGAN

Since in the original GAN no explicit control on the actual data generation is given, [16] proposed the conditional GAN (cGAN) to incorporate additional information like class labels in the synthesis process (Fig. 3). In the cGAN, the generator is presented with random noise z jointly with some prior information c. Additionally, the prior knowledge c is fed into the discriminator together with the corresponding real or fake data. Mathematically speaking, the cGAN framework is given as follows:

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

In image data, the authors showed that conditioning the GAN not only improves the generation of detailed features in the target image but also improves training stability.

Another conditional GAN framework is the Markovian GAN (MGAN) [17], proposed for fast and high-quality style transfer in images. The MGAN, as depicted in Fig. 4, heavily utilizes a pre-trained VGG19

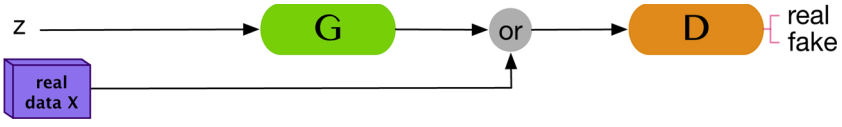

Fig. 2. GAN.

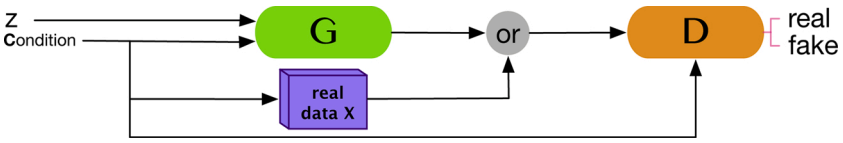

Fig. 3. cGAN.

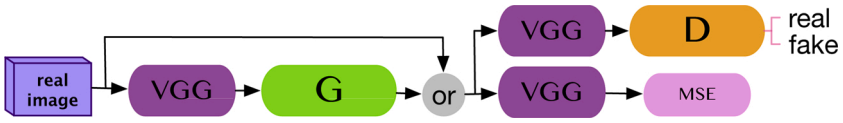

Fig. 4. MGAN.

network with fixed weights to extract high-level features for both style transfer and preserving the image content. Both the discriminator and generator network are prepended with a VGG19 network to extract feature maps. An additional perceptual loss on those extracted feature maps forces the generator to generate images with realistic VGG19 feature activations–as would have been obtained on real data.

[3] proposed Pix2Pix, a very successful variant of the cGAN for high-resolution image-to-image translation. The Pix2Pix generator follows the U-Net [18] architecture, while the discriminator–similar to MGAN–follows a fully convolutional architecture for differentiating between the real and fake high resolution data. The authors showed that the skip connections within the U-Net generator are beneficial for the global coherence of the synthesized images. In contrast to the original GAN framework, Pix2Pix requires pairs of corresponding input and desired output images. This allows the use of the ℓ_1 loss between the generators' output and actual ground-truth image to stabilize the training.

3.4. CycleGAN

For image transformation between two domains, the model should be able to extract characteristic features of both domains and to discover the underlying relationship between them. [19] proposed the CycleGAN to provide these principles. In essence, the framework combines two GANs to find a mapping from domain X to domain Y and vice versa. These consist of a generator $G: X \rightarrow Y$, trained by discriminator D_Y, and generator $F: Y \rightarrow X$, trained by discriminator D_X (Fig. 5). The two GANs are chained together, and a cyclic loss function forces them to reduce the space between their possible mapping functions. More precisely, this cyclic loss function minimizes the discrepancy between the original image and the reconstruction obtained from the chained generators. The final loss function of CycleGAN is defined as:

$$L(G, F, D_X, D_Y) = L_{\text{GAN}}(G, D_Y, X, Y) + L_{\text{GAN}}(F, D_X, Y, X) + \lambda L_{\text{cyc}}(G, F) \tag{3}$$

with

$$L_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim P_{\text{data}}(x)}[\| F(G(x)) - x \|_1] + \mathbb{E}_{y \sim P_{\text{data}}(y)}[\| G(F(y)) - y \|_1] \tag{4}$$

3.5. AC-GAN

The auxiliary classifier GAN (AC-GAN) proposed by [12] describes a different approach toward constructing a GAN conditioned on classes. Authors report that instead of providing both the generator and the discriminator networks with prior information (as shown in the cGAN), the discriminator can be additionally tasked with respectively classifying its input. More precisely, the discriminator architecture is modified such that after a few layers, it splits into a standard discriminator network as well as an auxiliary classifier network (see Fig. 6), which aims at classifying samples into different categories. According to the authors, this allows to use (partially) pre-trained discriminators and appears to stabilize training.

3.6. WGAN

In the discussed frameworks, the distributions of generated and real data are matched by means of the Jensen-Shannon (JS) divergence. This divergence measure potentially makes gradients vanishing and the saddle-point of optimization unreachable, which are the underlying reasons behind GAN failures. [20] proposed the Wasserstein-GAN (WGAN) which uses the Earth Mover (ME) or Wasserstein-1 distance as a more optimal divergence measure to avoid vanishing gradients. This

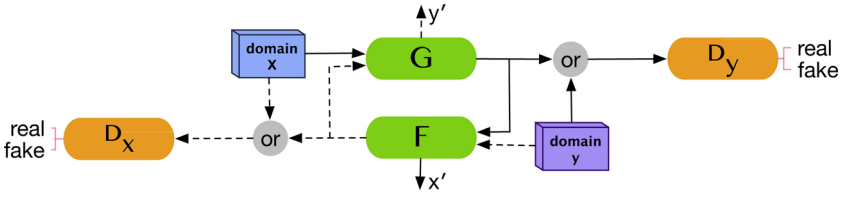

Fig. 5. CycleGAN.

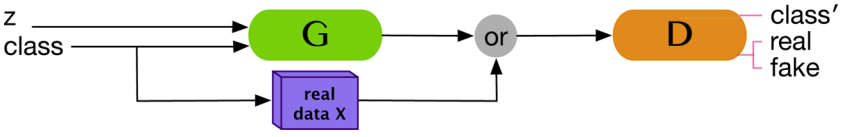

Fig. 6. AC-GAN.

Table 1
Metric explanation. (seg: segmentation, rec: reconstruction, det: detection, den: de-noising, reg: registration, clas: classification, TP: true positive, TN: true negative, FP: false positive, and FN: false negative.)

Abbreviation	Metric	Application	Abbreviation	Metric	Application
ROC	Receiver operating characteristics	seg	AUC	Area under the ROC curve	seg, det
IoU	Intersection over union: $\frac{\text{areaofoverlap}}{\text{areaofunion}}$	seg	Accuracy	$\frac{TP+TN}{TP+FP+FN+TN}$	seg, clas
Precision	$\frac{TP}{TP+FP}$	seg, clas	Specificity	$\frac{TN}{TN+FP}$	det
DSC	Dice similarity coefficient: $2\frac{TP}{2TP+FP+FN}$	seg, det, reg	Sensitivity	$\frac{TP}{TP+FN}$	seg, det
Recall	Sensitivity	seg, det, clas	F-score	$2\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$	seg
ASD	Average symmetric surface distance	seg	MSD	Mean surface distance	seg
HSD	Hausdorff surface distance	seg	OH	Object Hausdorff	seg
Correlation of EF	Correlation of ejection fraction	seg	MAE of infarct size	Mean absolute error on the surface of infarction	seg
NCC	Normalized cross correlation	det	Correct detection	(Subjective)	det
MSE	Mean squared error	rec	NMSE	Normalized MSE	rec
RMSE	Root MSE	rec	NRMESE	Normalized root MSE	rec
TRE	Target registration error	rec	SIS	Semantic interpretability score	seg
SNR	Signal to noise ratio	rec	PSNR	Peak SNR	rec, den
MOS	Mean opinion score (subjective)	rec	GCF	Global contrast factor	rec
SSIM	Structural similarity	rec, den	Agatston score	clinical intensity weighted measure of calcified area	den
Noise suppression	scores in [1, 5] (subjective)	den	Artifact reduction	Scores in [1, 5] (subjective)	den
Overall quality	Scores in [1, 5] (subjective)	den	P2PE	Point-to-point errors	den

model has proven to be much more robust, while easy to implement. The downside of the WGAN is its slow optimization.

3.7. LSGAN

[21] also tried to tackle the training instability of GANs with their so-called least squares GAN (LSGAN). Similarly to WGAN, the loss function is modified to avoid vanishing gradients:

$$min_G max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[(D(x)-b)^2] + \mathbb{E}_{z\sim p_z(z)}[(D(G(z))-a)^2] \quad (5)$$

This loss-function operates directly on the logits of the discriminator network, where $a=0$ and $b=1$ are the controlling parameters for fake and real data, respectively. This way, the fake samples which are discriminated as real, even if they are far away from the dense distribution of real data, will be penalized more during optimization. On the other hand, the gradient faces the least value only if the distribution of fake data perfectly matches the distribution of real data.

4. Applications in medical image processing

In this section, we summarize GAN-based methods proposed to solve medical imaging problems in 7 application categories: synthesis, segmentation, reconstruction, detection, de-noising, registration, and classification. In every subsection, a table summarizes the essential details in reviewed methods: the fundamental architectures and loss functions proposed, image modality and dataset properties, the evaluated performance of the proposed method, and if the paper underwent peer-review (PR) or provided any source code.[2] Since various metrics are used for the evaluation of different methods, we provided Table 1 to explain metrics briefly.

4.1. Synthesis

Originally, GANs have been proposed as an entirely unsupervised generative framework, with the ability to map from random noise to synthetic, realistically looking images. With the conditional GAN, the framework has also been successfully turned into a supervised generative framework by conditioning on prior knowledge, rather than noise alone. For clarity, we refer to the original GAN framework as the *unconditional* or *unsupervised* GAN, in contrast to the *conditional* GAN. The generative property of both frameworks has been exploited in various ways for synthesizing certain types of medical images. In the following, a broad overview of works from both categories will be given. In the particular case for conditional approaches, we further classify the contributions based on the imaging modality.

4.1.1. Unconditional image synthesis

A great variety of works has recently appeared in the field of unsupervised medical image generation using GANs, which allows us to tackle problems such as data scarcity and class imbalance [22], facilitates data simulation [23] and aids to gain deeper insights into the nature of data distributions and their latent structure.

Initial results have shown that the DCGAN can be used to synthesize realistically looking, small patches of prostate lesions [24], retinal images [25] or lung cancer nodules [23]. The synthesized lung cancer nodules could hardly be distinguished from real patches in a visual Turing test involving two radiologists. [22] also uses the DCGAN, but for the synthesis of focal CT liver lesion patches from different classes. For each class, i.e., cysts, metastases, and hemangiomas, they train a separate generative model. As the training dataset is originally quite small, they used heavily augmented data to train the GANs. In a set of experiments for liver lesion classification, the authors demonstrate that synthetic samples—in addition to data augmentation—can considerably improve a CNN classifier.

[26] has shown that the DCGAN can learn to mimic the distribution of entire MR data at considerably high resolution, even from a surprisingly small amount of samples. After training for 1500 epochs, the authors obtained visually compelling results which human observers could not reliably distinguish from real MR midline slices.

[10] compares the DCGAN, LAPGAN and some LAPGAN-modifications for skin lesion synthesis at high resolution. Due to the high variance within the training data, the small number of samples turned out not to be sufficient to train a reliable DCGAN. However, the hierarchical LAPGAN and its variants showed promising synthesis results. The synthetic samples have also been successfully used in the training of a skin lesion classifier. [27] employed the recently proposed progressive GAN growing [28] concept to synthesize high-resolution images of skin lesions and showed stunning, highly realistic synthetic images that expert dermatologists could not reliably tell apart from real samples.

[2] A list of papers with published source code and corresponding links is available on https://github.com/bumuckl/GANs-for-Medical-Image-Analysis/.

4.1.2. Conditional image synthesis

4.1.2.1. CT from MR. In many clinical settings, the acquisition of CT images is required. However, CT imaging puts the patient at risk of cell damage and cancer because of radiation exposure. This motivates the synthesis of CT images from MR acquisitions. [29] synthesizes CT images from corresponding MR images with a cascade of 3D fully convolutional networks. In order to improve realism of the synthetic CT images, they train the model with a pixel-wise reconstruction loss and an image gradient loss in addition to the adversarial training. The idea of utilizing a cascade of generators originates from the so-called auto--context model (ACM). In the ACM, a network provides its output as an additional input to a succeeding network to provide context information and allow for refinements (Fig. 7).

While [29] requires corresponding pairs of CT and MR images for training, [4] successfully utilizes CycleGANs to transform 2D MR images to CT images without the need for explicit, co-registered data pairs. Interestingly, their training led to even better results as the mapping is not affected by co-registration artifacts. [30] uses conditional GANs to map 3D MR data of the head to its CT counterpart to facilitate the segmentation of craniomaxillofacial bony structures. To obtain feasible image-to-image translation results, they propose so-called "deep supervision discrimination", which—similarly to the perceptual loss—utilizes the feature representations of a pretrained VGG16 model to (i) tell real and synthetic CT images apart and (ii) provide gradient updates to the actual generator network.

4.1.2.2. MR from CT. Similar to [4,31], successfully leverage CycleGANs for unpaired image-to-image translation to generate pairs of cardiac MR image and its corresponding segmentation mask from pairs of cardiac CT slices and its ground-truth segmentation. The authors have shown that the performance of a segmentation model can be improved by 16% when the model is additionally trained with the synthetic data. They also showed that a model trained with only synthetic data performs only 5% worse than a model trained on real data. A major limitation of GANs, as pointed out by [32], is the lack of a guarantee that tumors/lesions are preserved during image-to-image translation. To cope with this issue, [33] proposes a tumor-aware loss function for CycleGAN to synthesize MR images from CT images.

4.1.2.3. Retinal image synthesis. [34] utilizes a slight modification of the Pix2Pix framework [3] to generate high-resolution eye fundus image from binary images of vessel trees. In follow-up work, [35] further introduces an adversarial autoencoder (AAE), which is trained to compress vessel tree images into a multivariate normal distribution and also consecutively to reconstruct them. The resulting model synthesizes arbitrary high-resolution vessel tree images by sampling from the multivariate normal distribution. The synthetic vessel tree images, in turn, can be fed into the image-to-image translation model, leading to an end-to-end framework for realistic, high-resolution retinal image synthesis.

Very similarly, [36] proposes a two-stage approach, consisting of a GAN trained to synthesize vessel tree images from noise, and a Pix2Pix network [3] to generate realistic, high-resolution pairs of ground-truth vessel segmentation and corresponding eye fundus image. In succession, the authors investigate the performance of a U-Net trained for segmentation using real data pairs and another model trained only on the synthetic samples. They find that training from only the synthetic data leads to a slightly inferior model.

[37] also leverages the Pix2Pix framework for synthesizing filamentary structured images, i.e. eye fundus images and neurons from binary segmentation masks. In comparison to [34,35], the authors also provide their framework with a reference image for style. To enable style transfer, they train the generator with feedback from a VGG-network and the discriminator. Opposed to Pix2Pix, they do not introduce noise with the help of dropout, but by augmenting noise to the encoder-decoder network's bottleneck.

4.1.2.4. PET from CT. PET images are frequently used for diagnosis and staging in oncology. The combined acquisition of PET and anatomical CT images is a standard procedure in clinical routine. Furthermore, PET/CT imaging is becoming a valuable evaluation tool for new drug therapies. However, PET devices are expensive and involve radioactivity, thus put patients at risk. Consequently, the medical imaging analysis community has been working on synthesizing PET images directly from CT data. [38] has obtained initial promising results to synthesize liver PET images from CT data with a conditional GAN. The cGAN, inspired by [3], can synthesize very realistic looking PET images. However, its performance has a low response to underrepresented tumor regions, which leads to poor tumor detection performance in a set of use-case experiments. In comparison, an FCN for PET image synthesis is capable of synthesizing tumors but produces blurry images in general. By blending corresponding synthetic PET images coming from the conditional GAN and the FCN, they are able to achieve very high tumor detection performance.

Similarly, [39] utilizes a cGAN for synthesizing high-resolution PET images from pairs of CT images and binary labelmaps. While CT images alone would be sufficient as input, the authors emphasize that adding a labelmap leads to a globally more realistic, synthetic output. Because of the two-channel input to the generator, they refer to their network as the multi-channel GAN. Further, the authors validated their synthetic PET images with a tumor detection model trained on synthetic data and obtained comparable results to a model trained with real data.

4.1.2.5. PET from MRI. Measuring the myelin content in PET images of the human brain is valuable for monitoring disease progression, understanding physiopathology, and evaluating the treatment efficacy of multiple sclerosis (MS). Unfortunately, PET imaging for MS is costly and invasive as it requires the injection of a radioactive tracer. [40] successfully utilizes a cascade of two conditional GANs, based on 3D U-Nets for the generators and 3D CNNs for the discriminators, to synthesize PET images from a set of different MR volumes. The authors noted that a single cGAN was insufficient for the task at hand as it produced blurry images. Splitting the synthesis task into smaller, more stable subproblems seemed to improve the results drastically.

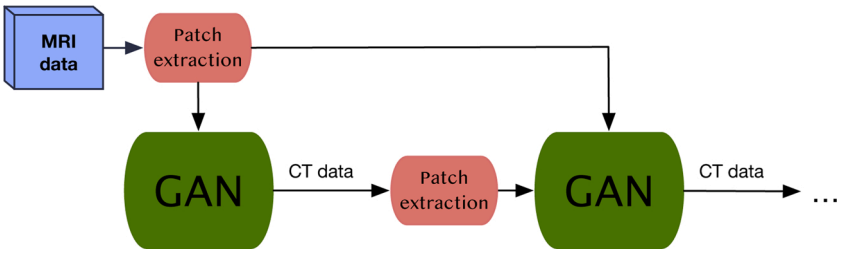

Fig. 7. Synthesis: Proposed architecture by [29].

4.1.2.6. Ultrasound. [41] proposes a cGAN conditioned on 3D voxel locations for synthesizing 2D freehand ultrasound (US) images of a fetus phantom. In comparison to the standard cGAN, the authors find it necessary to transform the pixel locations into feature-maps and inject these at various layers of the generator to facilitate training. In their experiments, they demonstrate the capability to simulate US images at locations unseen to the network. Further, they quantify the generation of sound images by comparing the location of clinically relevant anatomical landmarks in synthetic and real images and verifying the realism of the generated images in a usability study.

[42] applies GANs for intravascular US (IVUS) simulation in a multi-stage setup. A first generator, conditioned on physically simulated tissue maps, produces speckle images. The second generator maps the speckle images to low resolution, synthetic 64×64 pixels sized US images. A third generator transforms these low-resolution images into high-resolution samples at a resolution of 256×256 pixels. In a visual Turing test, the synthetic images could not reliably be distinguished from real ones.

4.1.2.7. X-ray. [43] trains a cGAN for chest X-ray synthesis from perturbed segmentation maps and combine it with Bayesian Neural Networks for Active Learning.

4.1.2.8. Stain normalization. Conditional GANs have also been leveraged for coping with the variance in digital histopathology staining, which is well known to cause problems for CAD systems. [44] points out that tumor classifiers do not only generalize poorly across data with different stains but also that existing stain-normalization approaches fail to preserve important image features. To overcome this issue, they propose a feature-preserving cGAN for stain style transfer. First, they map histopathology images to a canonical gray-scale representation. In succession, they leverage a cGAN to transform these gray-scale images into RGB images with the desired staining. The benefits of employing an additional feature-preserving loss on the hidden layers of the discriminator are experimentally validated by obtaining the smallest performance degradation among a variety of State-of-the-Art (SOTA) stain transfer methods.

[45] leverages the Pix2Pix framework for virtual H&E staining on unstained hyperspectral microscopy patches. The authors report promising quantitative results, but also point out to require expert feedback to draw a valid conclusion.

[46] tries to tackle the stain transfer problem with the help of an AC-GAN by simultaneously training a conditional GAN for stain-transfer and a task-specific network (i.e. a segmentation or classification model). The joint optimization of the generator, the discriminator, and the task-specific network drives the generator to produce images with relevant features being preserved for the task-specific model.

The aforementioned methods rely on paired training data to map from a source to a target staining. Such a dataset is often hardly available and requires preprocessing, such as co-registration. Moreover, co-registration itself is not perfect and is prone to produce artifacts. [47] alleviates the need for paired training data and co-registration by employing CycleGANs for the task of stain transfer. In a broad set of experiments on different datasets, they obtain visually more compelling stain transfer results than previous deep-learning and non-deep learning-based methods.

4.1.2.9. Microscopy. [48] proposes a framework similar to Pix2Pix for transferring phase contrast (PC) and differential interference contrast (DIC) in microscopy images to one another. A U-net like a generator is trained to synthesize a certain modality image from a source modality and the corresponding cell mask. Interestingly, two discriminators are employed to differentiate between (i) pairs of (real source, real target) versus (real source, synthesized target), and (ii) pairs of (cell mask real source) versus (cell mask, synthesized target). In a set of qualitative and quantitative evaluations, they rank their two-discriminator approach against the single-discriminator Pix2Pix framework. They report improved quantitative results when transferring from DIC image to phase contrast, and comparable results when trying to map from phase contrast to DIC. The authors amount the latter's comparable performance to the details already present in PC images, which leaves the cell mask with very little impact on the synthesis outcome.

4.1.2.10. Blood vessels. Machine learning driven analysis methods for detecting atherosclerotic plaque or stenosis in coronary CT angiography (CCTA) are powerful, but data-hungry. To deal with the lack of labeled data, [49] proposes to synthesize plausible 3D blood vessel shapes with the help of a WGAN fed with noise and attribute vectors. To facilitate the synthesis in 3D at appropriately high resolution, the authors generate 1D parameterizations of primitives which characterize blood vessels as a proxy for the final vessel rendering. Alternatively to CT, magnetic resonance angiography (MRA) has evolved as an important tool for visualizing vascular structures However, often it is not acquired alongside the standard protocols. [50] proposes the so-called steerable GAN to synthesize MRA images from T1 and T2-weighted MR scans, potentially alleviating the need for additional MR scans. Their conditional, steerable GAN combines a ResNet-like generator with a PatchGAN-discriminator. Also, they propose an ℓ_1-loss between the real and synthesized image as well as a steerable filter loss to promote faithful reconstructions of vascular structures.

The unconditional methods are summarized in Table 2, whereas the conditional GAN variants are summarized in Tables 3 and 4. In particular, we report the method containing the underlying GAN architecture, the image modalities on which the particular method operates, the datasets used, and the resolution of the synthesized images. We further mark whether a paper has been peer-reviewed or not. Since losses are a substantial part of the underlying GAN framework, we do not explicitly report them here. Further, we do not report any quantitative results since they (i) are in many cases unavailable, (ii) hardly interpretable, and (iii) overall hardly comparable.

In general, many interesting GAN-based approaches have been proposed for both unconditional and conditional image synthesis. To measure the soundness and degree of realism—certainly of great importance in medical data—the reviewed works have taken many different paths, some of which are debatable. If available, synthesis results can be checked against ground-truth using common measures like PSNR, SSIM, or MSE [48,45], which may not grasp the benefits introduced by GANs: Rendering of coherent image features beyond the pixel-level. The majority of works relies on human expert judgment (visual Turing tests) to measure realism. In some cases, the validity of the evaluation method or study design is questionable, though. For instance, in many visual Turing tests, it seemed to have been fairly easy to distinguish between real and generated images [22,23,41] due to artifacts in synthetic samples. [41] and [23] tackle this problem by applying anisotropic or Gaussian filtering to both real and fake samples before presenting them to the raters [41,23]. This is only valid as long as blurry images still contain the required amount of information for the task at hand. In some applications, domain-specific measures can be applied to judge the level of realism, e.g., by looking for anatomical landmarks and their relative positioning [34,41] in the synthetic data. Deep learning models themselves can also be used as a proxy to measure fidelity and information content by training one model on synthetic data, a second model on real data, and comparing their performance on a held-out test set of real images [31,36,39]. If no domain-specific means to measure soundness are available, enigmatic measures such as the Inception-Score [51], Frechet-Inception-Distance [52], or Sliced-Wasserstein-Distance [28], that quantify the distributional overlap of real and synthetic data in feature space, might come to the rescue. Another problem of GANs, which is hardly discussed in the reviewed literature is the phenomenon of mode collapse, under which GANs can

Table 2
Unconditional GANs for medical image synthesis.

Method	Architecture	Modality	Dataset	Resolution	PR	Code
[24]	DCGAN	MRI prostate lesions	SPIE ProstateX Challenge 2016 (330 MRI scans)	16 × 16	No	No
[23]	DCGAN	CT lung cancer nodules	LIDC-IDRI (1145 nodules)	56 × 56	Yes	No
[22]	DCGAN	Focal CT liver lesion patches	Non-public (182 CT scans)	64 × 64	Yes	No
[26]	DCGAN	2D axial brain MR slices	Baltimore Longitudinal Study of Aging (BLSA) (528 healthy subjects)	220 × 172	Yes	No
[10,27]	DCGAN, LAPGAN, PGAN	Skin lesions (dermoscopy)	ISIC2017 (2000 samples) ISIC2018 (10,000 samples)	256 × 256	No	No

Table 3
Conditional GANs for medical image synthesis.

Method	Architecture	Modality	Dataset	Resolution	PR	Code
[29]	3D Autocontext FCN with adversarial loss, image gradient loss and ℓ_2-loss	MR to CT	ADNI (16 subjects) and 22 non-public pelvic image pairs	32 × 32 × 32 (MRI) to 16 × 16 × 16 (CT)	Yes	Yes
[4]	CycleGAN	2D saggital brain MR and CT slices	Non-public (24 subjects)	256 × 256	No	No
[30]	3D cond. GAN	MR to CT	ADNI (16 subjects)	152 × 184 × 149	Yes	No
[31]	CycleGAN	2D cardiac MR w. segmentation mask to cardiac CT w. segmentation mask	Non-public (20 CT/CTA and 20 MRI)	232 × 232	No	No
[33]	Tumor-Aware CycleGAN	CT to MR	NSCLC (The Cancer Imaging Archive, 377 scans) 256 × 256 &non-public (42 scans)		Yes	No
[34,35]	AAE and Pix2Pix	2D binary vessel tree images to retinal images	DRIVE (40 samples) MESSIDOR (1200 samples)	512 × 512	Yes	Yes
[53]	3D cond. GAN	CT (lung nodules)	LIDC (1018 scans)	64 × 64 × 64	Yes	No
[36]	GAN and Pix2Pix	2D binary vessel tree, images to retinal	DRIVE (40 samples) MESSIDOR (1200 samples)	512 × 512	No	Yes
[37]	Pix2Pix w. Style Transfer	Eye fundus, microscopic neuronal	DRIVE (40 samples) STARE (20 samples) HRF (45 samples) NeuB1 (112 samples)	512 × 512 and higher	No	Yes
[38]	Pix2Pix and FCN	2D liver tumor CT to PET	Non-public (25 pairs)	n/a	No	No

Table 4
Conditional GANs for medical image synthesis.

Method	Architecture	Modality	Dataset	Resolution	PR	Code
[39]	Cond. multi-channel GAN	CT and segmentation pairs to PET images	Non-public (50 subjects)	200 × 200	No	No
[41]	Spatially cond. GAN	2D US	Non-public fetus phantom (26,396 images)	160 × 120	No	Yes
[42]	Multi-stage cond. GAN	Simulated tissue maps to 2D Intravascular US	IVUS challenge (2175 images)	256 × 256	Yes	No
[43]	Cond. GAN	Segmentation maps to synthetic X-ray images	SCR chest XRay database (247 images) NIH (400 images)	512 × 512	Yes	No
[44]	Feature-preserving cond. style-transfer GAN	Digital histopathology	CAMELYON16 (400 slides)	n/a	No	No
[45]	Pix2Pix	Hyperspectral microscopic to H&E stained	Non-public (2838 image pairs)	64 × 64	No	Yes
[46]	ACGAN	Digital histopathology	MITOS-ATYPIA14 (11 slides) MICCAI'16 GlaS challenge (165 slides) Non-public ovarian carcinoma (135 slides)	250 × 250	Yes	No
[47]	CycleGAN	Digital histopathology	MITOS-ATYPIA14 (11 slides) Camelyon16 (400 slides)	256 × 256	Yes	Yes
[48]	Cond. GAN with two discriminators	DIC &phase contrast microscopy	Non-public (1600 pairs)	256 × 256	No	No
[49]	WassersteinGAN	Geometric parameters extracted from CCTA	Non-public (4412 centerlines)	n/a	No	No
[50]	Cond. steerable GAN	MRA from T1 &T2w MRI axial slices	IXI Dataset (578 pairs)	n/a	No	No

only produce very similar looking samples. Particularly in the works of [24] and [25], where samples look fairly similar, it should be considered if mode collapse has occurred. In general, the community still lacks a meaningful, universal quantitative measure for judging synthetic images' realism. Regardless of realism, the aforementioned works have shown that GANs can be used successfully for data simulation and augmentation in classification and segmentation tasks. How exactly realism, artifacts, and other specific properties of generated samples affect a machine learning model generally remains an open question.

4.2. Segmentation

Segmentation of objects and organs in medical images is an essential pre-requisite for many applications such as detection, classification, and shape analysis. The tedious and time-consuming nature of manual segmentation made automatic methods the most active field of Deep-Learning research in medical image analysis [1]. However, in many scenarios, the pixel-wise based optimization and evaluation procedure in deep networks is not sufficient to capture notions of anatomical structures. Additional corrections on top of CNNs such as conditional random fields (CRFs) and statistical shape models (SSMs) are required to regularize the shape and appearance of segmentations, which are not easy to optimize [54,55].

In this context, transfer of the GAN concept to the domain of image segmentation has proven to be very useful. As a primary example, forcing the segmentation network to generate segmentations which

cannot be distinguished from manual segmentations through adversarial training constitutes an effective and efficient alternative to building a manual post-processing pipeline [56]. A different approach [57] successfully demonstrates—at the example of liver and brain tumor segmentation in MRI as well as cell segmentation in microscopic data—that an ensemble of adversarial networks can even be used to overcome class imbalance issues.

In the following subsection, we focus on segmentation approaches that leverage GAN-concepts, categorized based on specific parts of the anatomy.

4.2.1. Brain

The proposed methods for brain structures and abnormality segmentation comprise both supervised and unsupervised methods. [56] demonstrates that using the GAN training strategy in CNNs enhances not only the performance of semantic segmentation methods but also brings the performance of non-semantic segmentation methods closer to semantic ones. [58] highlights the superior performance of GANs in the segmentation of normalized/equalized patches of brain tumors. [59] proposes the SegAN framework, which employs the U-Net as the generator architecture of a GAN. They show that pixel-dependencies are learned better by using an adversarial loss in addition to a multi-scale, pixel-wise losses. One of the known challenges with most of the supervised segmentation methods is the performance degradation on unseen images. [60] proposes to leverage an adversarial framework to address this problem for unsupervised domain adaptation in brain lesion segmentation. In this method, a domain discriminator network is employed to make the segmentation network invariant to input from different domains. The varying appearance of anatomical tissues in multi-modal images leads to improved segmentation performance. For the segmentation of bony structures in brain images, [61] proposes to synthesize CT images from MRI images using GANs and then use both of them as the input to a segmentation network. They proposed a segmentation architecture called Deep-supGAN, optimized with five different losses: an adversarial loss, a voxel-wise loss, and three perceptual losses defined on differences between VGG extracted features. For multi-class classification of brain tumors, [62] proposes combining the cGAN and MGAN, where class labels define conditions.

4.2.2. Chest

Bad quality, local artifacts and the overlap of lung and heart areas are the main obstacles for the segmentation of chest X-ray images. As existing approaches in this field do not provide consistency between global and local features, [63] proposes the SCAN architecture, achieving human levels of performance in heart/lung segmentation. In this method, the generator is pre-trained with a pixel-wise loss function to address the instability problem of GANs.

4.2.3. Eye

Many CNN-based approaches perform even better than human experts in retinal vessel segmentation. However, blurriness and false-positive segmentations near minuscule or faint branches constitute a problem which is not solved yet. [11,64] proposes GANs as a solution to this problem. For optical disc and cup segmentation in 2D color fundus images, [65] successfully leverages a cGAN.

4.2.4. Abdomen

[66] proposes GANs which employ U-Net as the generator to segment the liver in 3D CT images of the abdomen. [67] proposes to use a CycleGAN for both liver and liver tumor segmentation. To address the problem of the miss-segmentation of tiny tumors, they propose a new architecture based on the U-Net (polyphase U-Net) for the generator. Spleen segmentation in MRI images is challenging due to the varying size and shape of this organ. To address this problem, [68] proposes a cGAN combined with the global convolutional network (GCN) [69] as the generator architecture. They showed that larger convolutional kernels in the GCN, in addition to adversarial training, enhance the segmentation performance on objects with considerable variability. For the challenging task of organ segmentation in MRI of the pelvis, [70] proposes the so-called STRAINet, a well-engineered, fully convolutional segmentation network, and successfully combine it with adversarial training to obtain more consistent realistic organ delineations.

4.2.5. Microscopic images

The automatic segmentation of microscopic images is challenging due to the variety of size, shape, and texture [71,72]. [71] proposes to use GANs with special training loss function, which considers a weight to specify which pixels in the foreground/background are more important. [72] uses GANs with special blocks (convolution followed by batch normalization) in the discriminator for the same problem. [73] proposes DAN—a combination of DCAN [74] and VGG16—which is trained with both supervised and unsupervised strategies to provide high-quality segmentation masks for unseen images.

4.2.6. Cardiology

Low contrast, high level of noise, and cardiac motion are challenges for segmenting cardiology images. To segment the left ventricle (LV) in low-contrast cardiology images, [75] proposes the VoxelAtlasGAN, which employs a V-Net [76] atlas-based segmentation in the generator of a cGAN framework (Fig. 8). Also, [77] proposes to benefit from the cGAN on top of atlas-based segmentation to facilitate feature extraction of all time sequences frames for precise segmentation of myocardial infarction. Since in segmentation, residual (non-RoI) information in addition to RoI features can make segmentation results more realistic, [78] and [79] propose a reconstruction strategy based on both of these features. They added a reconstruction loss to the optimization procedure of GAN as a controlling parameter. For semi-supervised and unsupervised segmentation, [78] and [79] propose to use the CycleGAN and LSGAN architectures, respectively. They also proposed some loss functions to address over-segmentation in large regions with varying intensity.

4.2.7. Spine

In medical image analysis of the spine, machine learning based approaches suffer from improperly learning the anatomy of the discs in vertebrae for segmentation and localization. [80] proposes a butterfly shape GAN model to segment disc regions in two views of vertebrae.

Tables 5–11 summarize GAN-based segmentation methods. These mainly use GAN/cGAN/DCGAN in addition to pixel/voxel-wise optimization loss functions. CycleGAN and reconstruction loss strategies are also proposed to consider non-RoI features for more precise segmentation. In the reviewed methods, U-Net and ResNet—due to providing general identification features—are the most popular segmentation networks for the generator architecture. However, in the reviewed papers, some limitations exist which trouble a clear judgment on the proposed methods. The ability of GANs to consider both global and local information performs close to semantic segmentation, even though difficulties related to semantic methods, unsupervised learning, and detail preservation are addressed. Tables show that GAN-based methods are trainable using a varying amount of data from 10 to 1000+ input samples. However, there are some limitations in the information provided in the reviewed papers. Some papers only used the DICE similarity score for evaluation, which its practical meaning is debatable.

Experimental results generally show that the competition between segmentation methods using and not using adversarial training is tightly closed. However, in microscopic image segmentation, GANs made a significant enhancement.

4.3. Reconstruction

Fast MR reconstruction without sacrificing details is a core problem in medical imaging. Fast acquisition and reconstruction directly reduces

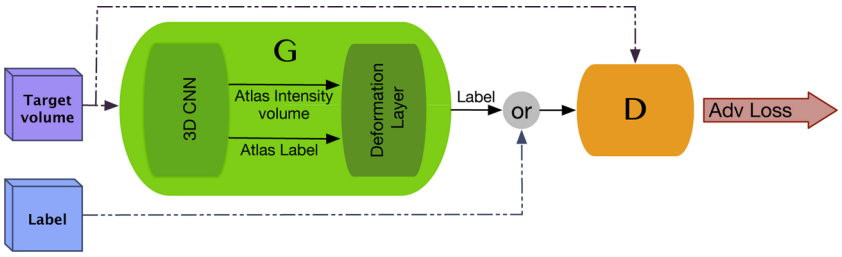

Fig. 8. Segmentation: Proposed architecture by [75].

Table 5
Segmentation GAN-based methods—Brain.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
[56]	GAN	Adv, cross entropy	MRI	MICCAI 2012 challenge Train: 15 Test: 20 MRBrainS13 challenge Train: 5 Test: 15	DSC = 0.92 ± 0.03 DSC = 0.85 ± 0.01	Yes	No
[58]	GAN	Adv	MRI	BRATS 2017 Train: 285	(Whole, Core, Enhancing) DSC = 0.87, 0.72, 0.68 Sensitivity = 0.87, 0.72, 0.68	Yes	No
SeGAN [59]	U-Net, GAN	Adv, multiScale	MRI	BRATS 2013 Train: 25 BRATS 2015 Train: 274	(Whole, Core, Enhanced) DSC = 0.84, 0.70, 0.65 Precision = 0.87, 0.80, 0.68 Sensitivity = 0.83, 0.74, 0.72 DSC = 0.85, 0.70, 0.66 Precision = 0.92, 0.80, 0.69 Sensitivity = 0.80, 0.65, 0.62	Yes	No
[60]	GAN, 3D-CNN	Adv, SGD	MRI (TBI)	Unknown Train: 61	DSC = 0.62 Recall = 0.58 Precision = 0.71	Yes	Yes
Deep-supGAN [61]	GAN, VGG16	Adv, perceptual, voxel-wise	MRI 3D	ADNI Train: 16	DSC = 94.46	Yes	No
[62]	c-GAN, MGAN	Adv	MRI	BRATS 2017 Train: 285	(Whole, Core, Enhanced) DSC = 0.70, 0.55, 0.40 Sensitivity = 0.68, 0.52, 0.99 Specificity = 0.99, 0.99, 0.99	Yes	No

Table 6
Segmentation GAN-based methods—Chest.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
SCAN [63]	VGG, ResNet, GAN	Multi-class cross-entropy, Adv	X-ray	JSRT Montgomery Train: 382	(Lungs, Heart) DSC = 0.973, 0.927 IoU = 0.947, 0.866	Yes	No

Table 7
Segmentation GAN-based methods—Eye.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
[11]	U-Net, GAN	Adv, Cross entropy	Funduscopy (Retina)	DERIVE Train: 20 Test: 20 STARE Train: 10 Test: 10	DSC = 0.829 ROC = 0.9803 Precision = 0.9149 DSC = 0.834 ROC = 0.9838 Precision = 0.9167	No	Yes
[64]	DCGAN	Adv	Funduscopy (Retina)	DERIVE (blood vessels) Train: 20 Test: 20	AUC = 0.945	Yes	No
[65]	c-GAN, ResU-net	Adv, $L1$	Funduscopy (Retina)	RIM-ONE Train: 159	(Disc, cup) F-score = 0.97, 0.94 IOU = 0.89, 0.76	Yes	No

any motion artifacts and are thus highly desirable.

Classic compressed sensing-based solutions directly use k-space information to reconstruct images [82]. The ability to promote realism in images with fast inference makes GANs an obvious candidate for solving the MR reconstruction problem. GAN-based MR reconstruction research has a major focus on modifying well-known architectures and

Table 8
Segmentation GAN-based methods—Abdominal.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
[66]	U-Net, auto-encoder	Adv, Cross-entropy	CT 3D (Liver)	Unknown MICCAI (SLiver07) Train: 1000+	DSC = 0.95 ASD = 1.90	Yes	No
[67]	U-Net, CycleGAN	CycleGAN, Cross-entropy, L2	CT 3D (Liver)	LiTS2017 Train: 73 Test: 9	(Liver, lesion) DSC = 0.89, 0.46 Recall = 0.94, 0.5 Precision = 0.86, 0.48	Yes	No
SSNet [68]	GCN, cGAN	Adv, DSC	MRI	Unknown Train: 45 Test: 15	DSC = 0.9260	Yes	No
[70]	STRAINet	Adv, Cross-entropy	MRI (Pelvis)	Unknown Train: 35 Test: 10	Bladder, prostate, rectum DSC = 0.968, 0.907, 0.905 ASD = 0.884, 1.317, 1.386	Yes	Yes

Table 9
Segmentation GAN-based methods—Microscopic.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
[71]	GAN, U-net, Res-Net, MS CNN	Adv, weighted loss	Cell 2D	Unknown Train: 5×6000	F-score = 0.70 Precision = 0.74 Recall = 0.69	Yes	No
[72]	GAN (rib cage)	Adv	Cell 2D	H1299 Train: 2–11	F-score = 0.89 Precision = 0.82 Recall = 0.85	Yes	Yes
DAN [73]	GAN, DCAN, VGG	Adv, Multi-scale cross entropy	Fungus 3D	2015 MICCAI Gland Challenge Train: 85 Test: 20	F-score = 0.88 DSC = 0.865 OH = 74.55	Yes	No

Table 10
Segmentation GAN-based methods—Cardiology.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
VoxelAtlasGAN [75]	cGAN, V-Net	Adv, intensity, label	Echo 3D	Unknown Train: 25 Test: 35	DSC = 0.95 MSD = 1.85 HSD = 7.26 Corr-of-EF = 0.91 Time = 0.1	Yes	No
MuTGAN [77]	GAN, ConvLSTM, 3DConv	Adv, MAE, DSC	3T MR Cardiac cine, DE-MR	Unknown Train: 140 Test:	DSC = 0.90 Accuracy = 96.46 Infarct size = 22.3	Yes	No
[78]	CycleGAN	Adv, DSC, MAE	Cine MR 3D	2017 ACDC Challenge Edinburgh Imaging Facility QMRI Train: 128 Test: 50	F-score = 0.771	Yes	Yes
DAN [79]	LSGAN, U-net	Adv, Intensity-Var, Over-Seg penalty, Recons	CT or MR Cardiac 2D	2017 MM-WHS Challenge Train: 20 Test: 3MR, 3CT	(MR) DSC = 0.66 (CT) DSC = 0.5	Yes	No

Table 11
Segmentation GAN-based methods—Spine.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
Btrfly Net [80]	GAN, Btrfly-Net	Adv, Btrfly-Net	CT 3D	[81] Train: 242 Test: 60	Precision = 0.84 Recall = 0.83 F1-score = 0.84	Yes	No

combining them with appropriate loss functions.

4.3.1. DAGAN-based strategies

Early research on GAN-based MR reconstruction focuses on the DAGAN architecture [83]. In this method, a perceptual loss is added to adversarial and pixel-wise losses to compare deep extracted features in real and generated information, which also enhances the stability of the model. A series of work refines the DAGAN architecture over time [82,83] by manipulating loss functions to preserve frequency information (Fig. 9). A combined loss function is defined as:

$$L_G = \alpha L_{\text{image-MSE}} + \beta L_{\text{freq-MSE}} + \gamma L_{\text{VGG}} + L_{\text{GAN}} \quad (6)$$

The next update of DAGAN is introduced in [84] which propose to add a refinement network (Fig. 10) to separate pixel-wise and perceptual

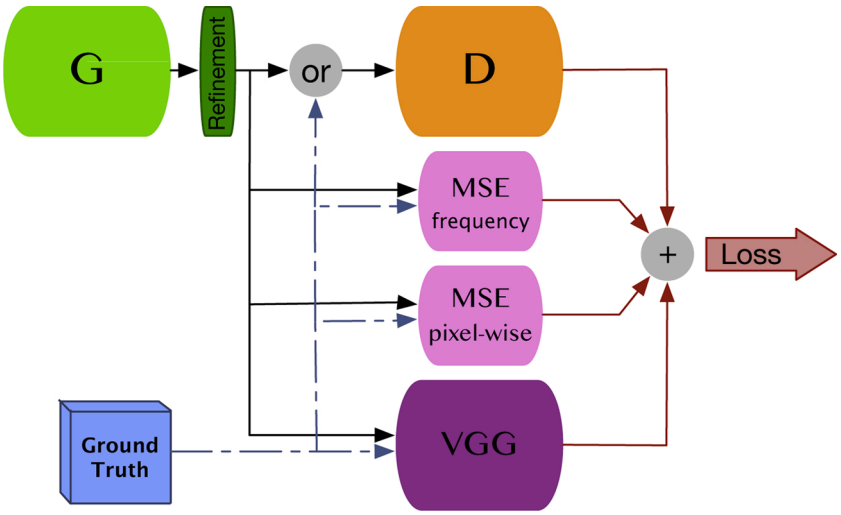

Fig. 9. Reconstruction: DAGAN architecture [83].

information-based training.

A similar architecture (Fig. 11) is proposed in [85], where the authors consider a cyclic training strategy based on the reconstruction of data in the lost frequencies. Moreover, they propose to use a chain of generators to address the ambiguities made in earlier generators.

4.3.2. 3D super-resolution strategies

[86] adapts the SRGAN [87] with 3D convolutional layers to deal with volumetric information and enhance the stability of the proposed GAN model. Their loss function combines a pixel-wise loss with a gradient-based loss (GDL) [88] to address the blurring effect in the reconstructions. [89] propose a 3DSRGAN along with two loss functions to control data interpolation and prevent over-fitting. In addition to the adversarial and the MSE losses, two other objectives are defined to overcome over-fitting and control data interpolation. The considerable memory footprint of 3D convolutions is a well-known challenge. To address this problem, [90] proposes a multi-level densely connected super-resolution network (mDCSRN), a combination of the WGAN model [20] and a modified version of DenseNet [91].

4.3.3. Other methods

Other GAN-based reconstruction methods mostly introduce additional loss functions to the original framework. In [92] and [93] pixel-wise losses, in [6] and [94] perceptual losses, in [95] a saliency loss, and in [96] Voronoi-vectorization and regularization losses are proposed to be added to the adversarial loss. [92] proposes to use the LSGAN as a solution to address training instability.

Tables 12–14 summarize properties of mentioned methods and their performance. While many architectural modifications of the GAN are proposed, it seems that ResNet is the most popular architecture for the generator. All the methods introduce a pixel-wise loss in addition to the adversarial loss. Also, other loss functions are introduced to preserve essential information in the lost data reconstruction. Due to their synthesis abilities, GANs can provide good performance in the reconstruction of lost data in medical images. Textural features are essential for diagnosis, and GANs should be trained with additional loss functions to satisfy these aspects. Generally, results of reviewed papers show that GANs, compared with other methods, perform faster and more accurately in data reconstruction. Most of the methods need a considerable amount of data to provide the convergence through all loss functions defined. Moreover, there is a lack of generally applicable evaluation metrics which explore the preservation of essential features in reconstructed data.

4.4. Detection

The reliable detection of anomalies from images with supervised deep learning requires a large amount of annotated training data. For very rare pathologies, such data is not readily available. Further, relying only on labeled data potentially limits supervised methods to detect only anomalies similar to those whose appearances are known at training time. In this context, GANs have carved open a completely new paradigm toward unsupervised anomaly detection, not requiring any pathological data a-priori at all.

The underlying approaches leverage the distribution-modeling capabilities of GANs to first model the distribution of normal anatomical variability. Then, the generative GAN is tasked to reconstruct the most similar, anomaly-free counterpart to a query image (potentially containing an anomaly), such that anomalies can be detected from discrepancies between the query and the reconstructed image.

Pioneering work was introduced in [25], the so-called AnoGAN, and showed that such an idea could be useful in detecting anomalies in OCT images of the retina. Succeeding work in [99] and [100] transferred this concept to anomaly detection in brain MR images. Inspired by the AnoGAN, [9] proposes the visual attribution GAN (VA-GAN) for Alzheimer's disease detection. This model extracts the map of changes that

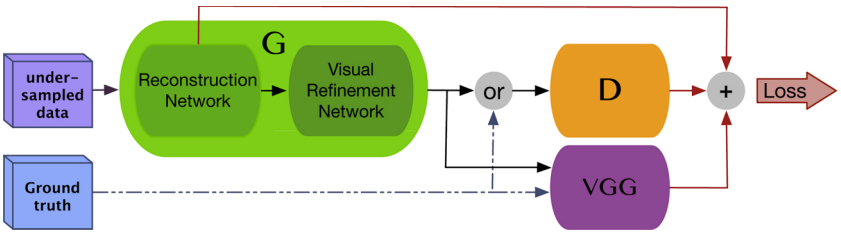

Fig. 10. Reconstruction: Proposed architecture by [84].

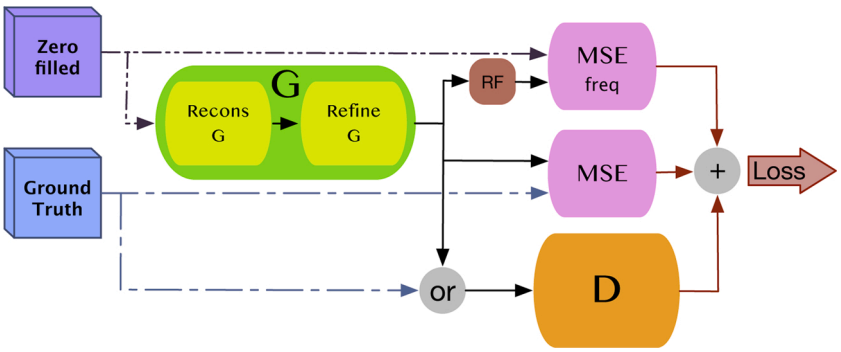

Fig. 11. Reconstruction: RefineGAN architecture [85].

Table 12
Reconstruction GAN-based methods—DAGAN based.

++ Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
[82]	cGAN, U-Net	Adv, Pix-wise, Perceptual, Refinement	MRI	IXI, MICCAI Grand Challenge 2013 Train: 1605 + 100 Test: 50	Mask 30%: NMSE = 0.09 ± 0.02 PSNR = 39.53 ± 4.12 (CPU, GPU) time = 0.2 ± 0.1, 5.4 ± 0.1 (ms)	No	No
DAGAN [83]	cGAN, U-Net	Adv, Pix-wise, Frequency, Perceptual, Refinement	MRI	MICCAI Grand Challenge 2013 Train: 21128 Test: 9854	Mask 30%: NMSE = 0.08 ± 0.02 PSNR = 40.20 ± 4.07 (CPU, GPU) time = 0.2 ± 0.1, 5.4 ± 0.1 (ms)	Yes	No
[84]	cGAN, U-Net	Adv, feature matching, Perceptual, penalty	MRI Cardiac	Unknown Train: 3000 3D test: 1200	PSNR = 31.82 ± 2.28 MOS = 3.24 ± 0.63 (max = 3.78 ± 0.45) SIS = 0.94 (max = 1)	Yes	No
RefineGAN [85]	GAN chain, ResNet	Adv, Cyclic	MRI	Brain: IXI Train: 100 Test: 100 Chest: Data Science Bowl challenge Train: 100 Test: 100	Mask 30%, Time: 0.16(s) SSIM = 0.97 ± 0.01 PSNR = 38.71 ± 2.57 Mask 30%, time: 0.18 (s) SSIM = 0.97 ± 0.01 PSNR = 38.64 ± 2.76	No	Yes

Table 13
Reconstruction GAN-based methods—3D super-resolution.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
[86]	SRGAN, subpixel-NN	LSGAN, GDL, Pixel-wise	MRI (Brain)	ADNI database Train: 470 Test: 119	(Scale 2, Scale 4) PSNR = 39.28, 33.58 SSIM = 0.98, 0.95	Yes	No
[89]	ResNet, GAN	Adv, Pixel-wise 3D variation	MRI (Brain)	Glioma patients Train: 30 + 10	PSNR = 24.2 MSE = 262.2	Yes	No
mDCSRN [90]	DenseNet, WGAN	MSE, WGAN	MRI (Brain)	Unknown Train: 891 Test: 111	PSNR = 35.88 SSIM = 0.94 NRMSE = 0.0852	Yes	No

convert the class of the image from healthy to diseased and use it for abnormality detection. [101] shows the better performance of the VA-GAN in detecting low contrast lesions arising from Alzheimer's in comparison with CNNs.

For aggressive prostate cancer detection [102] and skin lesion detection [103] propose to use the U-Net architecture as the generator of a GAN and a cGAN, respectively. In prostate US, the low contrast of tumor boundaries challenges contour detection. [104] addresses this problem using GANs.

Table 15 summarizes these papers. Papers proposed in anomaly detection by GANs have more structural complexity than previous applications because they benefit from different aspects of GANs. The role of the discriminator is highlighted in detection methods. The aforementioned methods show good performance in anomaly detection while they reduce the number of training data significantly. However, varying datasets and metrics employed for the experiments challenge a fair comparison between the methods.

4.5. De-noising

Imaging in diagnostic radiology typically involves a trade-off between image contrast and radiation hazard. A higher contrast might lead to better diagnosis, but exposes the patient to unwanted excessive radiation, whereas reduced radiation exposure leads to lower contrast and

Table 14
Reconstruction GAN-based methods—Other.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
GANCS [92]	ResNet, LSGAN	Adv	MRI (Chest)	Contrast-enhanced MRI abdomen pediatric patients Train: 300 Test: 50	SNR = 20.48 SSIM = 0.87 Time = 0.02	No	No
[93]	ResNet, GAN	Adv	MRI (Brain)	Unknown Train: 1560 Test: 346	PSNR = 37.95	No	No
GANCS [6]	ResNet, GAN, VGG	Adv	MRI (Brain 2D)	Unknown Train: 170 Test: 43	PSNR = 32.32 SSIM = 0.88 Time = 0.37	Yes	No
[95]	ResNet, GAN	Adv, CNN saliency	Retinal Funduscopy	Unknown (5000 data))	(Scale 4, Scale 8) SSIM = 0.89, 0.84 RMSE = 6.2, 7.5 PSNR = 44.3, 39db	Yes	No
[94]	[97], GAN	Adv, Perceptual	Microscopy (Cell)	Unknown Train: 11,000 Test: 500	PSNR = 27.8591	Yes	No
[96]	[87], GAN, Cyclic	Adv, Regular	Endo-microscopy	[98] Train: 202 Test: 36	SSIM = 0.8.7 $\Delta GCF_{\widehat{HR}} = 0.66$ $\Delta GCF_{\widehat{LR}} = 0.37$ $Tot_{cs} = 0.66$	Yes	No

Table 15
Detection GAN-based methods in medical image processing.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
AnoGAN [25]	DCGAN	Adv	SD-OCT scans	Unknown Train: 270 Test: 20	Precision = 0.8834 Recall = 0.7277 Sensitivity = 0.7279 Specificity = 0.8928 AUC = 0.89	Yes	No
[99]	AnoGAN, WGAN-GP	WGAN-GP, Regular	MRI (brain)	BRATS Train: 35 Test: 42	AUC = 0.92	No	No
VA-GAN [9]	WGAN, U-Net	Adv	MRI (brain)	ADNI Train: 80 Test: 20	NCC = 0.27	Yes	Yes
[102]	U-Net, GAN	MSE, GAN	MRI (prostate)	(NCT) Heidelberg Train: 188	Specificity = 0.98 ± 0.14 DSC = 0.41 ± 0.28 Sensitivity = 0.55 ± 0.36	No	No
[103]	cGAN, U-net	–	Natural skin	Unknown Train: 2417 Test: 583	(Subjective) Correct detect = 0.914	Yes	No
[104]	GAN	Adv, Local, Contour	Ultra-sound (prostate)	Unknown Train: 4570 Test: 229	DSC = 0.92 ± 0.3	Yes	No

lower signal-to-noise ratios. Deep learning has been successfully used to denoise low-contrast images and to enhance their resolution. However, these methods tend to produce blurry images. GANs, known to facilitate the generation of sharp, realistic-looking images, provide the means to mitigate this problem. A variety of works have recognized this potential and proposed different ways to exploit GANs for denoising images with perceptually higher quality.

For example, [5] proposes to learn tissue texture information from a small amount of paired data and address the blurring effect using GANs. Similarly, [105] proposes to use the cGAN to remove metal artifacts from CT images. [106] utilizes a combination of the W-GAN and a perceptual loss to improve training stability without losing perceptual

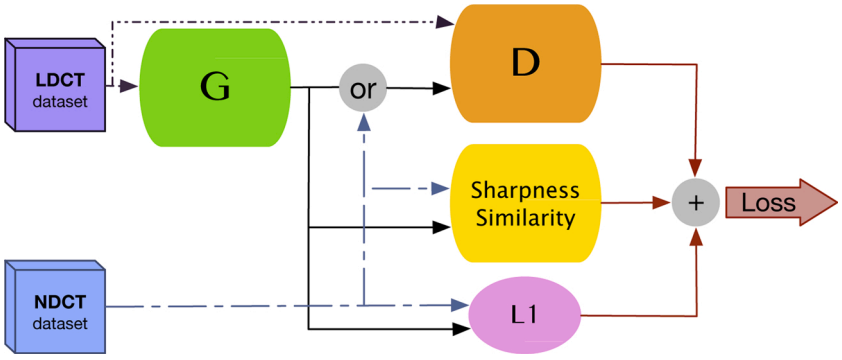

Fig. 12. Denoising: SAGAN architecture [107].

quality. The sharpness of the denoised image is also the factor that the sharpness aware generative adversarial network (SAGAN) [107] worked on (Fig. 12).

Table 16 summarizes major GAN-based denoising methods. While visually, the results look compelling, an adequate metric for evaluating the strength of methods in preserving important information in medical images is not yet available. SNR, MSE, SSIM, SD, and mean constitute the most commonly used metrics in the evaluation of de-noising methods but are not sensitive enough to consider texture details. Metrics applied on a level of regions-of-interest (RoI) provided by semantic segmentation could be a potential, but computationally costly solution. Generally, more research is needed toward a new, meaningful metric.

Despite this limitation, results obtained in reviewed papers visually benefit from the ability of GANs to learn the principal features of a domain of images. Modifications of the GAN loss function to incorporate specific textural features have proven to be beneficial for medical image de-noising.

However, finding a faster and more accurate framework is an open direction for future work.

4.6. Registration

Traditional registration methods suffer from parameter dependency and heavy optimization load. CNNs have been successfully used to align medical images in a single forward-pass through the network. Therein, GANs with their excellent image transformation capabilities have emerged as a candidate to extract a more optimal registration mapping.

[108] proposes an unsupervised GAN to register structural patterns (defined in patches) among different brain images. In 3D prostate MRI and intra-procedural transrectal ultrasound (TRUS), [7] and [109] propose GANs for registration and deformation correction, respectively. In the method proposed in [7], the discriminator serves as a certainty evaluator during testing.

Table 17 summarizes properties of the mentioned methods and their performance. In registration tasks, both local and global features are of importance. GANs learn features at various scales to model the discrepancy between distributions, hence provide this beneficial information. While GANs significantly enhance the performance of registration methods, in some real medical settings, the required performance can still not be reached.

4.7. Classification

In recent years, CNNs have proven to be a powerful tool for medical image classification tasks, but they are also known to require vast amounts of labeled training data. In the medical field, difficulties in access to such data hamper their productive employment. In Section 4.1 on image synthesis, numerous works have been reviewed which focus on training data augmentation with synthetic samples to overcome this limitation. However, besides artificially increasing training data, a few works leveraged the adversarial training concept to make patch-based classifiers domain invariant [8] or embed classifier training directly into the GAN framework for improved robustness and generalization [110].

For unsupervised domain adaptation in the challenging task of classification of whole-slide prostate histopathology images into two Gleason grades, [8] introduces a domain-discriminator that learns to distinguish between feature representations of a source and a target domain classifier. In turn, the adversarial optimization is used to minimize the discrepancy among their feature representations, leading to improved generalization, although no labeled target domain data was involved.

For separating useful from non-informative images in cardiac ultrasound (US), [110] proposes to embed a classifier training directly into a semi-coupled GAN-framework. In this work, the discriminator constitutes the actual classifier, and two generators are employed to generate adversarial examples of both useful and non-informative cardiac samples; consequently, the discriminator is extended to differentiate both real/fake and useful/non-informative samples (Fig. 13).

Table 18 provides insights into the performance and summarizes properties of the mentioned methods. The results indicate that robustness and generalization capabilities of classification tasks benefit from the adversarial training concept.

5. Discussion

GANs have been receiving significant attention from the medical imaging community—this is evident from the sudden spike in the number of papers published using GANs.

5.1. Advantages

They are capable of mimicking data distributions, producing realistic-looking images and learning rich similarity metrics—which are beneficial for discriminative deep learning frameworks.

Scarcity of labeled data and class imbalance: Often, annotations are expensive to obtain in medical imaging. This impairs supervised deep learning methods. At the same time, medical data often suffer from class imbalance due to the rare nature of some pathologies.

Table 16
De-noising GAN-based methods in medical image processing

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
[5]	CNN, GAN	CNN, Adv	CT (phantom) (cardiac)	Unknown Train: 48 Train: 28	Agatston score: Median = 20.7 Min = 6.1 Max = 145.1	Yes	No
[105]	U-net, ResNet, cGAN	Adv, *L*1	CT (ear)	Unknown Train: 14,346 Test: 74	P2PEs: Median = 0.409 STD = 0.133 Max = 0.912	Yes	No
[106]	WGAN, VGG	features distance, WGAN	CT	Unknown Train: 4000	Noise suppression = 3.20 ± 0.25 Artifact reduction = 3.45 ± 0.25 Overall quality = 3.70 ± 0.15	Yes	No
SAGAN [107]	MGAN, ResNet	Pixel-wise, MGAN, Sharpness	CT	CT phantom (Catphan 600) Train: 4×708 Test: 4×142	($N = 10^4$) PSNR = 26.77 SSIM = 0.8454 ($N = 10^5$) PSNR = 28.25 SSIM = 0.87	Yes	No

Table 17
Registration GAN-based methods in medical image processing.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
[108]	U-net, GAN	Adv, Regular	MRI 3D Brain	LPBA40 IBSR18 CUMC12 MGH10 Train: 30 Test: 10	DSC = 71.8 ± 2.3% DSC = 57.8 ± 2.7% DSC = 54.4 ± 2.9% DSC = 61.7 ± 2.1%	Yes	No
[7]	CNN, WGAN	Adv	Prostate 3D MRI and TRUS	Unknown Train: 636 Test: 127	TRE = 3.84 mm DSC = 0.58	Yes	No
[109]	3D GAN	Adv, DSC, Regular	Prostate 3D MRI TRUS	Unknown 108 pairs	TRE = 6.3 mm DSC = 0.82	Yes	No

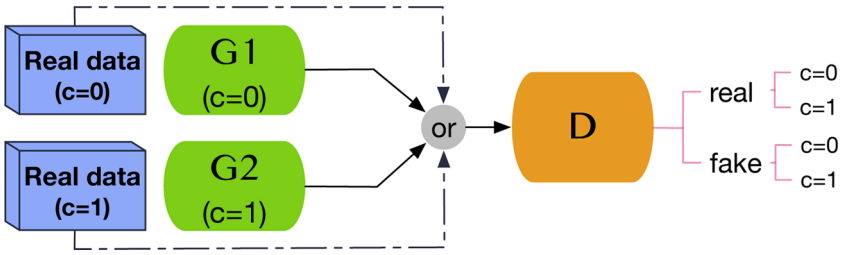

Fig. 13. Classification: SCGAN architecture [110].

Table 18
Classification GAN-based methods in medical image processing.

Method	Arch	Loss	Modality	Dataset	Performance	PR	Code
SCGAN [110]	GAN	Adv	MR (Cardiac)	UKBB 3400	(MAS − MBS) Acc = 92.5, 89.3 Precision = 87.6, 89.1 Recall = 90.5, 91.7	Yes	No
[8]	[97], GAN	Adv, patch-class	Prostate histopathology	TCGA, CINJ	(TCGA/TCGA to CINJ) Accuracy = 0.77, 0.75	Yes	No

GAN's ability to generate realistic-looking images can be employed not only to enlarge training datasets but also to be exploited in semi- and unsupervised settings.

Considerable success has been made in image-to-image translation for cross-modality image synthesis, such as mapping from MR to CT data. By synthesizing across modalities, conditional GANs reduce the cost and risk factors of medical image acquisition (Section 4.1.2) and offer multi-modal information which can be fused for better diagnostic decision making. In unsupervised realms, the GAN's ability to learn data distributions opens up the possibility to detect unseen abnormal cases in real datasets (Section 4.4).

Rich feature extraction: Learning distinctive patterns in medical images plays a vital role in the diagnosis of diseases. GANs, equipped with the adversarial training concept, are powerful in extracting semantically meaningful features from images which traditional pixel-wise losses fail to grasp. This property has been successfully leveraged in segmentation (Section 4.2), registration (Section 4.6), and de-noising (Section 4.5). On the other hand, the hierarchical discriminative potential of GANs has also been utilized for classification applications (Section 4.7).

5.2. Drawbacks

We identify three major drawbacks in the current form of GANs that might hinder their acceptance in the medical community:

Trustability of the generated data: In healthcare, where gaining the clinician's trust is the biggest challenge for any technology, images generated by GANs provide little comfort. The basic networks (generator and discriminator) are still deep neural networks, the mechanism of which is not sufficiently understood. In medical images, intensities are typically associated with some meanings, e.g. in CT data, every intensity can be mapped to the Houndsfield-scale and thus characterizes certain tissue. Such an association and mapping are currently missing from the GAN reconstruction—a shortcoming severe enough for clinicians to distrust images synthesized with GANs. In computer vision, where the overall perception is the main concern, these results are more adequate.

Unstable training: The typical GAN training is unstable because of numerical and conceptual reasons [111]. This may result in convergence issues such as mode collapse or mode-hopping. Fundamental theoretical work focuses on solving this problem and usually provides benchmarks on computer vision datasets, where generated images are easy to interpret. However, in medical imaging, where the modes of images are unclear, the identification of such unstable situations and unrealistic results can be very challenging. In the reviewed methods, architectural modifications and customized loss functions are proposed to address these problems. Yet, there is a lack of proper evaluation tools and medical benchmarks which measure the efficacy of the proposed methods in a comparable way.

Evaluation: In medical imaging, researchers mostly rely on traditional pixel-wise metrics to evaluate GANs performance. This is tricky since GANs overcome the limitations of these traditional

measures. Most of these metrics are usable only when ground truth images are available and this challenges the evaluation of unsupervised methods. On the other hand, as [112] mentioned, specific metrics should be employed to evaluate the performance of GAN-based methods due to the randomness of the initialization/optimization procedure and model instability problem. In the reviewed papers, such a metric is not explored.

Uninterpretability: A model will be reliable in a medical environment if it follows features that clinicians consider in diagnosis and prognosis. Even if a model provides information which is not simply recognized by medical experts, reasons for its decision should be interpretable. While GANs show superior performance in many applications, they suffer from the same un-interpretability as other deep models. This is the main obstacle to their practical application in medical environments. Generally, solving deep networks' interpretability would open up a new direction to discover shortages in the models.

We believe GANs need to address the significant drawbacks discussed before being considered a trustworthy technology in practical medical image analysis. To this end, we can think of GANs as a technical building block rather than a stand-alone piece of technology for the future. For example, in the case of synthesizing CT data, enveloping GANs synthesis with a physics-based simulation might ensure realistic HU values. Training instabilities need to be addressed as well, which means rigorous experimentation to understand the convergence of GANs in the medical imaging context. Important steps have already been made in the computer vision field. In short, along with exciting results, GANs open up many possible research questions for the next few years. Proper understanding and answering those hold the key to their successful deployment in the real clinical scenario.

Author contributions

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgments

S.A. is supported by the PRIME programme of the German Academic Exchange Service (DAAD) with funds from the German Federal Ministry of Education and Research (BMBF).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.artmed.2020.101938.

References

[1] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
[2] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Advances in neural information processing systems 2014:2672–80.
[3] Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. 2017 IEEE conference on computer vision and pattern recognition (CVPR) 2017:5967–76.
[4] Wolterink JM, Dinkla AM, Savenije MH, Seevinck PR, van den Berg CA, Išgum I. Deep MR to CT synthesis using unpaired data. International workshop on simulation and synthesis in medical imaging 2017:14–23.
[5] Wolterink JM, Leiner T, Viergever MA, Išgum I. Generative adversarial networks for noise reduction in low-dose CT. IEEE Trans Med Imaging 2017;36(12):2536–45.
[6] Zhang P, Wang F, Xu W, Li Y. Multi-channel generative adversarial network for parallel magnetic resonance image reconstruction in k-space. International conference on medical image computing and computer-assisted intervention 2018:180–8.
[7] Yan P, Xu S, Rastinehad AR, Wood BJ. Adversarial image registration with application for MR and TRUS image fusion. International workshop on machine learning in medical imaging 2018:197–204.
[8] Ren J, Hacihaliloglu I, Singer EA, Foran DJ, Qi X. Adversarial domain adaptation for classification of prostate histopathology whole-slide images. International conference on medical image computing and computer-assisted intervention 2018:201–9.
[9] Baumgartner CF, Koch LM, Can Tezcan K, Xi Ang J, Konukoglu E. Visual feature attribution using Wasserstein GANs. Proceedings of the IEEE conference on computer vision and pattern recognition 2018:8309–19.
[10] Baur C, Albarqouni S, Navab N. MelanoGANs: high resolution skin lesion synthesis with GANs. 2018. CoRR abs/1804.04338.
[11] Son J, Park SJ, Jung K-H. Retinal vessel segmentation in fundoscopic images with generative adversarial networks. 2017. CoRR abs/1706.09318.
[12] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier GANs. Proceedings of the 34th International conference on machine learning, vol. 70 of proceedings of machine learning research, PMLR 2017:2642–51. http://proceedings.mlr.press/v70/odena17a.html.
[13] Goodfellow I, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. International conference on learning representations 2015. http://arxiv.org/abs/1412.6572.
[14] Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, et al. Intriguing properties of neural networks. International conference on learning representations 2014. http://arxiv.org/abs/1312.6199.
[15] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. 2015. CoRR abs/1511.06434.
[16] Mirza M, Osindero S. Conditional generative adversarial nets. 2014. CoRR abs/1411.1784.
[17] Li C, Wand M. Precomputed real-time texture synthesis with Markovian generative adversarial networks. European conference on computer vision 2016:702–16.
[18] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. International conference on medical image computing and computer-assisted intervention 2015:234–41.
[19] Zhu J-Y, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017 IEEE international conference on computer vision (ICCV) 2017:2242–51.
[20] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: Precup D, Teh YW, editors. Proceedings of the 34th international conference on machine learning, vol. 70 of proceedings of machine learning research, PMLR. Sydney, Australia: International Convention Centre; 2017. p. 214–23. http://proceedings.mlr.press/v70/arjovsky17a.html.
[21] Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP. Least squares generative adversarial networks. 2017 IEEE international conference on computer vision (ICCV) 2017:2813–21.
[22] Frid-Adar M, Klang E, Amitai M, Goldberger J, Greenspan H. Synthetic data augmentation using GAN for improved liver lesion classification. 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) 2018:289–93.
[23] Chuquicusma MJM, Hussein S, Burt JR, Bagci U. How to fool radiologists with generative adversarial networks? A visual Turing test for lung cancer diagnosis. 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) 2018:240–4.
[24] Kitchen A, Seah J. Deep generative adversarial neural networks for realistic prostate lesion MRI synthesis. 2017. CoRR abs/1708.00129.
[25] Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. International conference on information processing in medical imaging 2017:146–57.
[26] Bermudez C, Plassard AJ, Davis LT, Newton AT, Resnick SM, Landman BA. Learning implicit brain MRI manifolds with deep learning. Medical imaging 2018: image processing, vol. 10574 2018:105741L.
[27] Baur C, Albarqouni S, Navab N. Generating highly realistic images of skin lesions with GANs. OR 2.0 context-aware operating theaters, computer assisted robotic endoscopy, clinical image-based procedures, and skin image analysis 2018:260–7.
[28] Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. International conference on learning representations 2018. https://openreview.net/forum?.id=Hk99zCeAb.
[29] Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, et al. Medical image synthesis with context-aware generative adversarial networks. International conference on medical image computing and computer-assisted intervention 2017:417–25.
[30] Zhao M, Wang L, Chen J, Nie D, Cong Y, Ahmad S, et al. Craniomaxillofacial bony structures segmentation from MRI with deep-supervision adversarial learning. International conference on medical image computing and computer-assisted intervention 2018:720–7.
[31] Chartsias A, Joyce T, Dharmakumar R, Tsaftaris SA. Adversarial image synthesis for unpaired multi-modal cardiac data. International workshop on simulation and synthesis in medical imaging 2017:3–13.

[32] Cohen JP, Luck M, Honari S. Distribution matching losses can hallucinate features in medical image translation. International conference on medical image computing and computer-assisted intervention 2018:529–36.

[33] Jiang J, Hu Y-C, Tyagi N, Zhang P, Rimner A, Mageras GS, et al. Tumor-aware, adversarial domain adaptation from CT to MRI for lung cancer segmentation. International conference on medical image computing and computer-assisted intervention 2018:777–85.

[34] Costa P, Galdran A, Meyer MI, Abràmoff MD, Niemeijer M, Mendonça AM, et al. Towards adversarial retinal image synthesis. 2017. CoRR abs/1701.08974.

[35] Costa P, Galdran A, Meyer MI, Niemeijer M, Abràmoff M, Mendonça AM, et al. End-to-end adversarial retinal image synthesis. IEEE Trans Med Imaging 2017;37(3):781–91.

[36] Guibas JT, Virdi TS, Li PS. Synthetic medical images from dual generative adversarial networks. 2017. CoRR abs/1709.01872.

[37] Zhao H, Li H, Cheng L. Synthesizing filamentary structured images with GANs. 2017. CoRR abs/1706.02185.

[38] Ben-Cohen A, Klang E, Raskin SP, Amitai MM, Greenspan H. Virtual pet images from CT data using deep convolutional networks: initial results. International workshop on simulation and synthesis in medical imaging 2017:49–57.

[39] Bi L, Kim J, Kumar A, Feng D, Fulham M. Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs). Molecular imaging, reconstruction and analysis of moving body organs, and stroke imaging and treatment 2017:43–51.

[40] Wei W, Poirion E, Bodini B, Durrleman S, Ayache N, Stankoff B, et al. Learning myelin content in multiple sclerosis from multimodal MRI through adversarial training. International conference on medical image computing and computer-assisted intervention 2018:514–22.

[41] Hu Y, Gibson E, Lee L-L, Xie W, Barratt DC, Vercauteren T, et al. Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks. Molecular imaging, reconstruction and analysis of moving body organs, and stroke imaging and treatment 2017:105–15.

[42] Tom F, Sheet D. Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning. 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) 2018:1174–7.

[43] Mahapatra D, Bozorgtabar B, Thiran J-P, Reyes M. Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network. 2018. arXiv preprint arXiv:1806.05473.

[44] Cho H, Lim S, Choi G, Min H. Neural stain-style transfer learning using GAN for histopathological images. 2017. CoRR abs/1710.08543.

[45] Bayramoglu N, Kaakinen M, Eklund L, Heikkilä J. Towards virtual H&E staining of hyperspectral lung histology images using conditional generative adversarial networks. 2017 IEEE international conference on computer vision workshops (ICCVW) 2017:64–71.

[46] Bentaieb A, Hamarneh G. Adversarial stain transfer for histopathology image analysis. IEEE Trans Med Imaging 2018;37(3):792–802.

[47] Shaban MT, Baur C, Navab N, Albarqouni S. Staingan: Stain style transfer for digital histological images. 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019) 2019:953–6.

[48] Han L, Yin Z. Transferring microscopy image modalities with conditional generative adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition workshops 2017:99–107.

[49] Wolterink JM, Leiner T, Isgum I. Blood vessel geometry synthesis using generative adversarial networks. MIDL conference book, MIDL mIDL 2018 medical imaging with deep learning 2018.

[50] Olut S, Sahin YH, Demir U, Unal G. Generative adversarial training for MRA image synthesis using multi-contrast MRI. Predictive intelligence in medicine 2018:147–54.

[51] Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training GANs. Advances in neural information processing systems 2016:2234–42.

[52] 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 2017:6626–37.

[53] Jin D, Xu Z, Tang Y, Harrison AP, Mollura DJ. CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. Medical image computing and computer assisted intervention—MICCAI 2018 2018:732–40.

[54] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 2018;40:834–48.

[55] Tack A, Mukhopadhyay A, Zachow S. Knee menisci segmentation using convolutional neural networks: data from the osteoarthritis initiative. Osteoarthr Cartil 2018;26(5):680–8.

[56] Moeskops P, Veta M, Lafarge MW, Eppenhof KA, Pluim JP. Adversarial training and dilated convolutions for brain MRI segmentation. Deep learning in medical image analysis and multimodal learning for clinical decision support 2017:56–64.

[57] Rezaei M, Yang H, Meinel C. Conditional generative refinement adversarial networks for unbalanced medical image semantic segmentation. 2018. arXiv preprint arXiv:1810.03871.

[58] Li Z, Wang Y, Yu J. Brain tumor segmentation using an adversarial network. International MICCAI brainlesion workshop 2017:123–32.

[59] Xue Y, Xu T, Zhang H, Long LR, Huang X. SegAN: adversarial network with multi-scale L_1 loss for medical image segmentation. Neuroinformatics 2018:1–10.

[60] Kamnitsas K, Baumgartner C, Ledig C, Newcombe V, Simpson J, Kane A, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. International conference on information processing in medical imaging 2017:597–609.

[61] Zhao M, Wang L, Chen J, Nie D, Cong Y, Ahmad S, et al. Craniomaxillofacial bony structures segmentation from MRI with deep-supervision adversarial learning. International conference on medical image computing and computer-assisted intervention 2018:720–7.

[62] Rezaei M, Harmuth K, Gierke W, Kellermeier T, Fischer M, Yang H, et al. A conditional adversarial network for semantic segmentation of brain tumor. International MICCAI brainlesion workshop 2017:241–52.

[63] Dai W, Dong N, Wang Z, Liang X, Zhang H, Xing EP. SCAN: structure correcting adversarial network for organ segmentation in chest X-rays. Deep learning in medical image analysis and multimodal learning for clinical decision support 2018:263–73.

[64] Lahiri A, Ayush K, Biswas PK, Mitra P. Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale miscroscopy images: automated vessel segmentation in retinal fundus image as test case. 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW) 2017:794–800.

[65] Shankaranarayana SM, Ram K, Mitra K, Sivaprakasam M. Joint optic disc and cup segmentation using fully convolutional and adversarial networks. Fetal, Infant and Ophthalmic Medical Image Analysis 2017:168–76.

[66] Yang D, Xu D, Zhou SK, Georgescu B, Chen M, Grbic S, et al. Automatic liver segmentation using an adversarial image-to-image network. International conference on medical image computing and computer-assisted intervention 2017:507–15.

[67] Kim B, Ye JC. Cycle-consistent adversarial network with polyphase u-nets for liver lesion segmentation. MIDL 2018 medical imaging with deep learning 2018. https://openreview.net/pdf?.id=SyQtAooiz.

[68] Huo Y, Xu Z, Bao S, Bermudez C, Plassard AJ, Liu J, et al. Splenomegaly segmentation using global convolutional kernels and conditional generative adversarial networks. Medical imaging 2018: image processing, vol. 10574 2018:45–51.

[69] Peng C, Zhang X, Yu G, Luo G, Sun J. Large kernel matters—improve semantic segmentation by global convolutional network. 2017 IEEE conference on computer vision and pattern recognition (CVPR) 2017:1743–51.

[70] Nie D, Wang L, Gao Y, Lian J, Shen D. Strainet: spatially varying stochastic residual adversarial networks for MRI pelvic organ segmentation. IEEE Trans Neural Netw Learn Syst 2018;30(5):1552–64.

[71] Sadanandan SK, Karlsson J, Wählby C. Spheroid segmentation using multiscale deep adversarial networks. 2017 IEEE international conference on computer vision workshops (ICCVW) 2017:36–41.

[72] Arbelle A, Riklin-Raviv T. Microscopy cell segmentation via adversarial neural networks. 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) 2018:645–8.

[73] Zhang Y, Yang L, Chen J, Fredericksen M, Hughes DP, Chen DZ. Deep adversarial networks for biomedical image segmentation utilizing unannotated images. International conference on medical image computing and computer-assisted intervention 2017:408–16.

[74] Chen H, Qi X, Yu L, Heng P-A. DCAN: deep contour-aware networks for accurate gland segmentation. 2016 IEEE conference on computer vision and pattern recognition (CVPR) 2016:2487–96.

[75] Dong S, Luo G, Wang K, Cao S, Mercado A, Shmuilovich O, et al. VoxelAtlasGAN: 3D left ventricle segmentation on echocardiography with atlas guided generation and voxel-to-voxel discrimination. International conference on medical image computing and computer-assisted intervention 2018:622–9.

[76] Milletari F, Navab N, Ahmadi S-A. V-net: fully convolutional neural networks for volumetric medical image segmentation. 2016 fourth international conference on 3D vision (3DV) 2016:565–71.

[77] Xu C, Xu L, Brahm G, Zhang H, Li S. MuTGAN: simultaneous segmentation and quantification of myocardial infarction without contrast agents via joint adversarial learning. International conference on medical image computing and computer-assisted intervention 2018:525–34.

[78] Chartsias A, Joyce T, Papanastasiou G, Semple S, Williams M, Newby D, et al. Factorised spatial representation learning: application in semi-supervised myocardial segmentation. International conference on medical image computing and computer-assisted intervention 2018:490–8.

[79] Joyce T, Chartsias A, Tsaftaris SA. Deep multi-class segmentation without ground-truth labels. MIDL conference book, MIDL mIDL 2018 medical imaging with deep learning 2018.

[80] Sekuboyina A, Rempfler M, Kukačka J, Tetteh G, Valentinitsch A, Kirschke JS, et al. Btrfly net: Vertebrae labelling with energy-based adversarial learning of local spine prior. International conference on medical image computing and computer-assisted intervention 2018:649–57.

[81] Glocker B, Zikic D, Konukoglu E, Haynor DR, Criminisi A. Vertebrae localization in pathological spine CT via dense classification from sparse annotations. International conference on medical image computing and computer-assisted intervention 2013:262–70.

[82] Yu S, Dong H, Yang G, Slabaugh GG, Dragotti PL, Ye X, et al. Deep de-aliasing for fast compressive sensing MRI. 2017. CoRR abs/1705.07137.

[83] Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, et al. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE Trans Med Imaging 2017;37(6):1310–21.

[84] Seitzer M, Yang G, Schlemper J, Oktay O, Würfl T, Christlein V, et al. Adversarial and perceptual refinement for compressed sensing MRI reconstruction. International conference on medical image computing and computer-assisted intervention 2018:232–40.

[85] Quan TM, Nguyen-Duc T, Jeong W-K. Compressed sensing MRI reconstruction with cyclic loss in generative adversarial networks. 2017. CoRR abs/1709.00753.
[86] Sánchez I, Vilaplana V. Brain MRI super-resolution using 3D generative adversarial networks. MIDL conference book, MIDL mIDL 2018 medical imaging with deep learning 2018.
[87] Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. 2017 IEEE conference on computer vision and pattern recognition (CVPR) 2017:105–14.
[88] Mathieu M, Couprie C, LeCun Y. Deep multi-scale video prediction beyond mean square error. International conference on learning representations 2016. http://arxiv.org/abs/1511.05440.
[89] Li Z, Wang Y, Yu J. Reconstruction of thin-slice medical images using generative adversarial network. International workshop on machine learning in medical imaging 2017:325–33.
[90] Chen Y, Shi F, Christodoulou AG, Xie Y, Zhou Z, Li D. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network. International conference on medical image computing and computer-assisted intervention 2018:91–9.
[91] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL, Deeplab:. Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 2018;40:834–48.
[92] Mardani M, Gong E, Cheng JY, Vasanawala SS, Zaharchuk G, Alley MT, et al. Deep generative adversarial networks for compressed sensing automates MRI. 2017. CoRR abs/1706.00051.
[93] Shitrit O, Raviv TR. Accelerated magnetic resonance imaging by adversarial neural network. Deep learning in medical image analysis and multimodal learning for clinical decision support 2017:30–8.
[94] Han L, Yin Z. A cascaded refinement GAN for phase contrast microscopy image super resolution. International conference on medical image computing and computer-assisted intervention 2018:347–55.
[95] Mahapatra D, Bozorgtabar B, Hewavitharanage S, Garnavi R. Image super resolution using generative adversarial networks and local saliency maps for retinal image analysis. International conference on medical image computing and computer-assisted intervention 2017:382–90.
[96] Ravì D, Szczotka AB, Pereira SP, Vercauteren T. Adversarial training with cycle consistency for unsupervised super-resolution in endomicroscopy. Med Image Anal 2019;53:123–31.
[97] Su H, Yin Z, Huh S, Kanade T. Cell segmentation in phase contrast microscopy images via semi-supervised classification over optics-related features. Med Image Anal 2013;17(7):746–65.
[98] André B, Vercauteren T, Buchner AM, Wallace MB, Ayache N. A smart atlas for endomicroscopy using automated video retrieval. Med Image Anal 2011;15(4):460–76.
[99] Chen X, Konukoglu E. Unsupervised detection of lesions in brain MRI using constrained adversarial auto-encoders. MIDL conference book, MIDL mIDL 2018 medical imaging with deep learning 2018.
[100] Baur C, Wiestler B, Albarqouni S, Navab N. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. International MICCAI brainlesion workshop 2018:161–9.
[101] Shwartz-Ziv R, Tishby N. Opening the black box of deep neural networks via information. 2017. CoRR abs/1703.00810.
[102] Kohl S, Bonekamp D, Schlemmer H-P, Yaqubi K, Hohenfellner M, Hadaschik B, et al. Adversarial networks for the detection of aggressive prostate cancer. 2017. CoRR abs/1702.08014.
[103] Udrea A, Mitra GD. Generative adversarial neural networks for pigmented and non-pigmented skin lesions detection in clinical images. 2017 21st international conference on control systems and computer science (CSCS) 2017:364–8.
[104] Tuysuzoglu A, Tan J, Eissa K, Kiraly AP, Diallo M, Kamen A. Deep adversarial context-aware landmark detection for ultrasound imaging. International conference on medical image computing and computer-assisted intervention 2018:151–8.
[105] Wang J, Zhao Y, Noble JH, Dawant BM. Conditional generative adversarial networks for metal artifact reduction in CT images of the ear. International conference on medical image computing and computer-assisted intervention 2018:3–11.
[106] Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, et al. Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE Trans Med Imaging 2018;37(6):1348–57.
[107] Yi X, Babyn P. Sharpness-aware low-dose CT denoising using conditional generative adversarial network. J Digit Imaging 2018:1–15.
[108] Fan J, Cao X, Xue Z, Yap P-T, Shen D. Adversarial similarity network for evaluating image alignment in deep learning based registration. International conference on medical image computing and computer-assisted intervention 2018:739–46.
[109] Hu Y, Gibson E, Ghavami N, Bonmati E, Moore CM, Emberton M, et al. Adversarial deformation regularization for training image registration neural networks. International conference on medical image computing and computer-assisted intervention 2018:774–82.
[110] Zhang L, Gooya A, Frangi AF. Semi-supervised assessment of incomplete LV coverage in cardiac MRI using generative adversarial nets. International workshop on simulation and synthesis in medical imaging 2017:61–8.
[111] Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: an overview. IEEE Signal Process Mag 2018;35(1):53–65.
[112] Lucic M, Kurach K, Michalski M, Gelly S, Bousquet O. Are GANs created equal? A large-scale study. Advances in neural information processing systems 2018:700–9.