

GANs for Medical Image Analysis

Salome Kazemina^{a,1}, Christoph Baur^{b,1}, Arjan Kuijper^a, Bram van Ginneken^c, Nassir Navab^b, Shadi Albarqouni^b, Anirban Mukhopadhyay^a

^a*Department of Computer Science, TU Darmstadt, Germany*

^b*Computer Aided Medical Procedures (CAMP), TU Munich, Germany*

^c*Image Sciences Institute, University Medical Center Utrecht, Netherlands*

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 denoising, 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. A total of 63 papers published until end of July 2018 are reviewed. For quick access, the papers and important details such as the underlying method, datasets and performance are summarized in tables.

Keywords: Generative Adversarial Networks, Medical, Image Synthesis, Segmentation, Reconstruction, Denoising, Superresolution

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 put on decision making at different granularity levels, with techniques ranging 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).

Email addresses: salome.kazemina@gris.tu-darmstadt.de (Salome Kazemina), c.baur@tum.de (Christoph Baur), arjan.kuijper@mavc.tu-darmstadt.de (Arjan Kuijper), bram.vanginneken@radboudumc.nl (Bram van Ginneken), navab@cs.tum.edu (Nassir Navab), shadi.albarqouni@tum.de (Shadi Albarqouni), anirban.mukhopadhyay@gris.tu-darmstadt.de (Anirban Mukhopadhyay)

¹The authors contributed equally to this work.

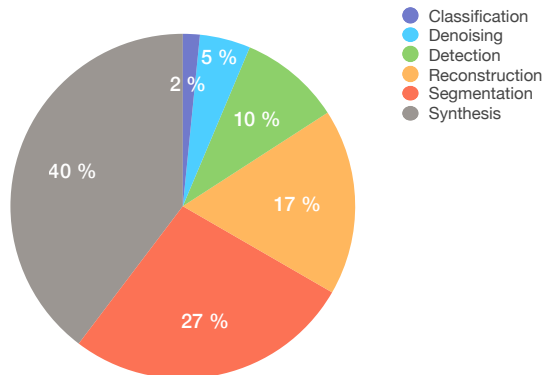


Figure 1: The distribution of papers among the different categories.

The driving focus of the machine learning-based Medical Image Analysis community has been on the supervised learning of decision boundaries, while generative tasks have been on the back seat. The unique ability of Generative Adversarial Networks (GANs) introduced in [2] by Goodfellow et al. to mimic data distributions has carved open the possibility to bridge the gap between *learning* and *synthesis*. The rapid enhancement of GANs [3] are facilitating the synthesis of realistic-looking images at unprecedented level. The reasons behind this superiority are related to two basic properties. First, GANs as an unsupervised training method aim to obtain pieces of information over data [4], in a fashion similar to the way human learns features of an image [1]. Second, GANs have shown significant performance gains in the extraction of visual features by discovering the high dimensional latent distribution of the data.

This review summarizes GAN-based architectures proposed for medical image processing applications, including de-noising, reconstruction, segmentation, detection, classification and image synthesis. The distribution of papers according to this classification can be seen in Fig. 1. We also provide tables to have quick access to key information like the performance of methods, metrics, datasets, modality of images and the general format of the proposed architecture. Moreover, we discuss the advantages and shortcomings of the methods and specify clear directions for future works.

In this review, we have covered medical imaging application of GAN published until December 2017, and MICCAI and MIDL 2018 accepted GAN-based papers, which were available on arXiv. Papers published in this time range propose using GANs in medical applications of de-noising, reconstruction (compressed sensing and super-resolution), segmentation, detection, classification and image synthesis. These papers were applied to different image modalities such as MRI, CT, OCT, chest X-Ray, Dermoscopy, Ultrasound, PET, and Microscopy. To find the papers we searched for 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.

We thus ended up with 63 papers which we consider the most relevant ones covering a broad spectrum of applications and variety of GANs. The rest 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, detection, classification, and synthesis) are described and Section 5 provides a conclusion about the investigated methods, challenges and open directions in employing GANs for 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. However, a major limiting factor for this paradigm, not only in the context of medical applications, is its dependence on vast amounts of annotated training data. In the medical field, this is particularly crucial, as the acquisition and labeling of medical images require experts, is tedious, time-consuming and costly, which leads to a severe lack of labeled training data. Besides, in the medical field, many datasets suffer from severe class imbalance due to the rare nature of some pathologies. In this context, generative modeling can potentially act as a reliever for resolving these well-known machine learning problems. GANs have shown the capabilities to generate images with unprecedented realism. Under the assumption that GANs can generate meaningful samples that enhance existing datasets and carry useful information, a variety of research has already been conducted for medical image synthesis, which is reviewed in Subsection 4.6.

Another issue hampering the machine learning community is the necessity to handcraft similarity measures for general tasks such as Superresolution, In-Painting, Segmentation or Image-to-Image translation. Traditional similarity objectives comprise pixel-wise losses such as the ℓ_1 or ℓ_2 -distance, both of which induce blurry results and lack the incorporation of context [2]. The adversarial training concept behind GANs theoretically eliminates the need to model explicit pixel-wise objective functions by learning a rich similarity metric to tell real and fake data apart. This allows optimizing for concepts in images beyond the pixel-level, leading to more realistic results. This appealing property has been recently exploited for improved medical image segmentation (reviewed in Subsection 4.3), Image-Enhancement such as Denoising (reviewed in Subsection 4.1) and tackling the general problem of domain shift in medical images using GAN-based Image-to-Image translation techniques (reviewed in Subsection 4.6.2).

The phenomenon of domain shift is in fact another major issue currently limiting the generalization capabilities of Deep Learning models. The assumption that training and inference data come from the same distributions and trained

models should thus also function properly on unseen data often does not hold and limits the applicability of the models. Domain Adaptation is concerned with making models robust to such domain shift, and adversarial training holds a lot of potential for this task.

In the deficiency of annotated and simultaneous abundance of unlabeled data, the paradigm of semi-supervised learning offers different frameworks for training machine learning models by ensuring similar or dissimilar model behavior for similar or dissimilar data points, where similarity needs to be defined appropriately. Again, the notion of similarity is a crucial parameter and often highly data-dependent. Under such conditions, GANs and adversarial training have also proven useful for training classifiers or dealing with domain shift in medical data, as the explicit formulation of similarity is not required (reviewed in 4.6.2).

3. Overview

In this section, we introduce the general concept behind GANs, their conditional variants as well as a variety of prominent extensions and follow-up works that have been successfully leveraged in Medical Image Analysis applications. These extensions comprise Wasserstein-GAN, conditional GAN (for example of Pix2Pix), CycleGAN, Least Squares GAN, Markovian GAN as well as Auxiliary Classifier GAN.

In the context of this work, there are three *adversarial* concepts, which should be understood properly by their different meanings. *Adversarial attack* means to make imperceptible changes to an image such that a classifier misclassifies it, while it could classify unmodified image successfully. Usually the modified image, called adversarial image or *adversarial examples*, is not recognizable from the original image visually. *Adversarial training* proposed by Szegedy et al. [5] is an idea that increases the robustness of neural networks against adversarial attacks by learning their characteristics. Due to the state of existing neural networks, at the time, implementing adversarial training was not a practical solution. The effectiveness of this idea becomes apparent when Goodfellow et. al employed it in GANs [2]. Sometimes GAN is mis-attributed as adversarial training, but it is necessary to differentiate between them. In reality, GANs consists of two types of networks and use the adversarial training concept, elaborated in the following section.

3.1. GAN

The GAN framework [2] consists of a generator (G), a discriminator (D) network as well as a training dataset of real data \mathbf{X} with an underlying distribution p_{real} . G, as a forger, is a multilayer network with parameters θ_G , which aims to find a mapping $\hat{x} = G(z; \theta_G)$ that relates latent random variables $z \sim p_z(z)$ to fake data following the distribution $p_\theta(\hat{x}|z)$. By discovering the mapping, G generates fake data, which is supposed to not be distinguishable from real data, i.e. $p_\theta(\hat{x}|z) \sim p_{real}$. On the other hand, discriminator $D(x; \theta_D)$ aims to

distinguish the fake samples from real ones. Thereby, $D(x)$ is the scalar output of the discriminator network that shows the probability that x is real rather than generated from $p_\theta(\hat{x}|z)$ (Fig. 2). D is trained to maximize the probability of correct label assignment to fake and real data, while G is trained to fool the discriminator by minimizing $\log(1 - D(G(z)))$. Mathematically speaking, D and G play a two-player minimax game with value function $V(G,D)$:

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

This way, the generator is updated only through gradients back-propagated from the discriminator. Goodfellow et al. [2] mentioned that if the generator is optimized to maximize $\log(D(G(z)))$ instead of minimizing $\log(1 - D(G(z)))$, much stronger gradients can be obtained in earlier steps (iterations) of training. In general, this indirect optimization procedure prevents input components to be explicitly memorized by the generator. The main advantage of GAN is to find similarities that map a candidate model to the distribution of real data by focusing on the underlying probability density of data. It leads to very sharp distributions around data, which can be used to degeneration of that [3].

Though GANs show such inherent advantages over discriminatively trained CNNs, there are some challenges as well: 1) *mode collapse*: when G collapses to map all latent space inputs to the same data and 2) *instability*: which leads to the generation of different outputs for same input. The main causes for these phenomena are related to vanishing gradients through the optimization procedure.

Although batch-normalization comes as a solution for the instability of GAN, it does not enough to improve the performance of GAN to the optimal stability. So, many subclasses of GANs have been introduced to resolve these drawbacks that some of the most common ones are introduced here. Furthermore, many GAN-based deep networks are proposed specifically for medical image processing projects, in which different architecture and loss functions are used to enhance the reliability and accuracy of the deep networks in the necessary level of health-care CAD systems. 4.

3.2. DCGAN:

To address the instability of GAN, Radford et. al propose the Deep Convolutional GAN (DCGAN) [6], in which both the generator and discriminator follow a deep convolutional network architecture. These networks are able to extract hierarchical features of the image by learning down/up-sampling due to the location of features existence. In this way, the extracted features of objects can be used to generate new ones. Key components of the DCGAN which affect the stability of the network, are batch normalization and leaky-ReLU. Although DCGANs are more stable than the vanilla GAN, they are still prone to mode collapse.

3.3. cGAN:

Mirza et al.[7] proposing conditional GAN (cGAN), have also shown that prior information can be incorporated into the GAN framework. In the cGAN,

the generator is presented with random noise z as well as some prior information c jointly. 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:

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

By conditioning the networks, it has been shown that both training stability and output generation can be improved (Fig.3). In [4], Isola et al. propose a very successful variant of the conditional GAN named “pix2pix” for the challenging task of image-to-image translation. In this architecture, the generator and discriminator are following the U-Net [8] and MGAN (PatchGAN) [9] networks which are demonstrated to provide a good framework for wide conditional transformation problems. In the proposed model, the $L1$ loss in combination with adversarial loss is considered to put more pressure on the generator to produce images more similar to the ground truth images.

3.4. MGAN

Another conditional GAN framework is Markovian GAN (MGAN) [9], which has been proposed by Li et al. for fast and high quality style transfer. The MGAN, as depicted in Fig. 4, heavily utilizes a pre-trained VGG19 network with fixed weights to extract high-level features for both transferring style to a target texture and simultaneously preserving the image content. In the MGAN, both discriminator and generator network are prepended with a VGG19 network to extract featuremaps. The generator transfers these featuremaps to an image with target texture, and the discriminator transforms the either input (real or texturized) image into VGG19 feature maps again, on which it finally discriminates with the help of a Fully Convolutional Network (FCN). Utilizing an FCN for classifying the input as real or fake ultimately amounts to classifying patches in VGG19 feature map space. By training the generator to fool this discriminator, it is forced to generate images which lead to realistic VGG19 feature activations as would have been obtained on real data and thus also to images with realistic style. An additional perceptual loss component (calculated using VGG) ensures that the image content does not change too much while the style is transferred.

3.5. cycleGAN:

Zhu et al. [10] propose a GAN architecture, which aims to discover the underlying relationship between two image domains through learning their definitive features from unpaired data. To achieve this goal a cycle training algorithm is used to capture main features of a domain of image for translating them to another domain. Since the map function learned by adversarial loss is not reliable to map input image to desired output, a cycle loss function is considered to reduce the space of possible mapping functions. In this way, two generators ($G : X \rightarrow Y$ and $F : Y \rightarrow X$) are considered to find the mapping from X

domain to Y domain and vice versa and also two discriminators (D_Y and D_X) to train them (Fig.5). This learning strategy stabilizes network performance and generates high quality translated images. Final loss function is defined as follows:

$$\begin{aligned} L(G, F, D_X, D_Y) = & L_{GAN}(G, D_Y, X, Y) \\ & + L_{GAN}(F, D_X, Y, X) \\ & + \lambda L_{cyc}(G, F) \end{aligned}$$

$$L_{cyc}(G, F) = \mathbb{E}_{x \sim P_{data}(x)} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim P_{data}(y)} [\|G(F(y)) - y\|_1] \quad (3)$$

3.6. AC-GAN

Odena et al.[11] report that instead of providing both the generator and the discriminator networks with side information as seen in the cGAN, the discriminator can be tasked with reconstructing such side information. In their auxiliary classifier GAN framework (AC-GAN, Fig. 6), the discriminator architecture is modified such that after a few of layers it splits into a standard sample discriminator network as well as an auxiliary classifier network, which aims at classifying samples into different categories. The authors show that this framework allows to use (partially) pre-trained discriminators and appears to stabilize training.

3.7. WGAN:

In the original GAN framework, the data distributions of generated and real images are compared using the Jensen-Shannon (JS) divergence. This kind of exact comparison can make the saddle-point of optimization unreachable and gradients vanishing, which leads to mode collapse and instability. So considering another, more approximate distance estimation between real and generated data distribution can be effective as a solution. Arjovsky et al. [12] propose the Wasserstein-GAN (WGAN) architecture that uses the Earth Mover (ME) or Wasserstein-1 distance estimation instead of the JS divergence. In addition, both the generator and discriminator follow the general DCGAN architecture. WGAN provides a robust adversarial generative model through a more meaningful learning procedure, which is able to find deeper relationships between distributions. Despite these theoretical advantages, WGAN leads to a slow optimization process in practical scenarios.

3.8. LSGAN:

Mao et al. propose another solution for the instability of GAN, called Least Squares GAN (LSGAN) [13]. In this architecture, some parameters are added in loss function to avoid gradient vanishing. In this way, the fake data, which are discriminated as real but is far away from the dense distribution of real data,

will be penalized due to its distance from main mode of real data. Also, the gradient will become 0 only in the case that distribution of fake data perfectly matches the distribution of real data. Loss function for LSGAN is defined as follows:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [(D(x) - b)^2] + \mathbb{E}_{z \sim p_z(z)} [(D(G(z)) - a)^2] \quad (4)$$

4. Applications in Medical Image Processing

In this section, we summarize GAN-based methods which are proposed to solve medical imaging problems, in 6 application categories: de-noising, reconstruction, segmentation, detection, classification, and synthesis. In every subsection, a table summarizes the most important details of proposed methods and the medical image modalities they are designed for.

4.1. De-noising

Due to health hazards caused by excessive radiation, lowering the radiation dose has been adopted as an effective solution. However, dose reduction increases noise level in medical images which might lead to a possible loss of some diagnostic information. The main problem with state-of-the-art CNN-based de-noising methods is the limitation of using the mean squared error in optimization, which leads to blurred predicted images that do not provide texture quality of routine-dose images. Another problem is the shortage of well aligned images of low-dose and routine-dose [14, 15, 16]. GANs can eliminate this problem by detecting the mapping between noisy and de-noised images and generate de-noised images. Here some GAN-based de-noising methods are reviewed.

Wolterink et al. [14] propose a GAN based de-noising method that can learn texture information of images from a small amount of paired data. In this paper three combinations of two loss functions for the generator optimization are explored: 1) voxel-wise MSE between generated image and routine-dose CT image, and 2) adversarial loss. The performance of this architecture is investigated on different metrics. Results show that using just the adversarial loss reduces the noise level while it saves statistics of image better than other SOTA-methods relying on a pixel wise loss. Moreover, the runtime is less than 10 seconds.

On the other hand, Yang et al. [15] propose another method utilizing two perceptual losses for training the generator: 1) the loss calculated by comparing deep features (extracted by VGG[17]) of generated image and ground-truth, and 2) WGAN loss (Fig.7). In this way, benefiting stability of WGAN model, the noise level is decreased and critical structures of an image are not damaged. Authors believe that SSIM and PSNR are not adequate metrics to evaluate the performance of such a de-noising method, because they are not able to evaluate the feature preservation power of the methods. They suggest to estimate the distance of standard deviation (SD) of generated images to the SD of ground.

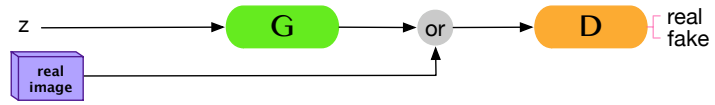


Figure 2: GAN

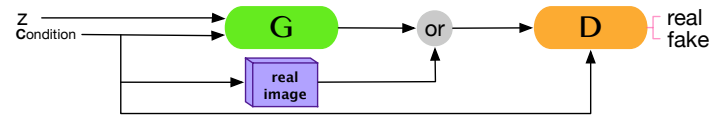


Figure 3: cGAN

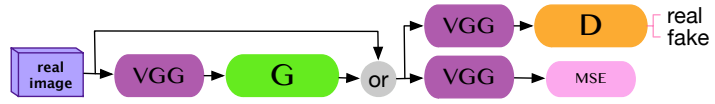


Figure 4: MGAN

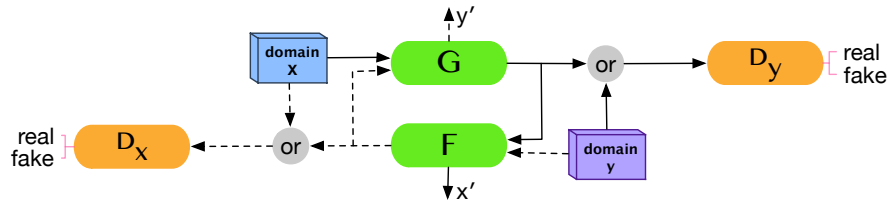


Figure 5: cycleGAN

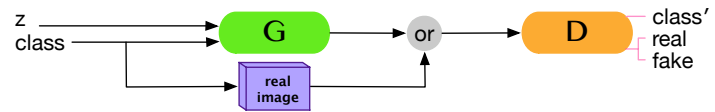


Figure 6: ACGAN

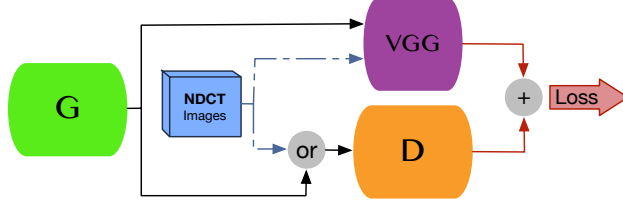


Figure 7: Proposed architecture in [15]

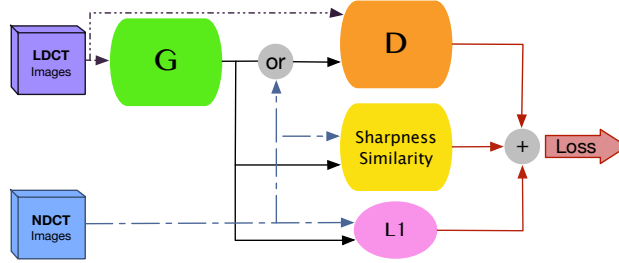


Figure 8: SAGAN architecture [16]

Evaluated results using this metric shows that the proposed method achieves the best performance in comparison with other methods.

To address the blurring problem of CNNs, Yi et al [16] propose the sharpness aware generative adversarial network (SAGAN) uses three losses in training: 1) A traditional pixel-wise loss to encourage data fidelity, 2) patch-GANs adversarial loss and 3) a sharpness mapping loss (Fig. 8). Presented results in the paper show that texture preservation, computational latency, generalizability and stability are advantages of the proposed method while in high level of noise it does not present a good performance. Moreover, small low-contrast data may be lost through sharp area detection.

Table 1 summarizes major GAN-based de-noising methods. It seems that an adequate objective metric to evaluate methods in preserving important medical information of the image is not available yet. As PSNR, MSE, SSIM, SD and mean - the most commonly used metrics in the evaluation of de-noising methods - are not sensitive enough to recognize texture details, the RoI area of any image should be segmented to be measured by metrics, which is an expensive procedure. So presenting a new metric to this goal can be a subject of future works. Despite this limitation, reviewed papers benefiting from the ability of GANs to learn main general features of medical images. Also manipulating the loss function to consider more textural features, good performance in medical image de-noising is achieved. However, finding a fast, accurate and more stable architecture is an open direction to be worked in future. Specially, if experts

evaluation on the de-noised images become provided.

Table 1: De-noising GAN-based methods in medical image processing

Method	Modality	Dataset	Performance
[14]			
Architecture: CNN, GAN Loss: CNN, GAN	CT (phantom) and (cardiac)	Unknown	Agatston Score: median=20.7 Min=6.1 Max=145.1
[15]			
Architecture: WGAN, VGG Loss: features distance, WGAN	CT	Unknown	Subjective [1 5]: Noise Suppression=3.20±0.25 Artifact Reduction=3.45±0.25 Overall Quality=3.70±0.15
SAGAN [16]			
Architecture: MGAN, Sharpness detector Loss: Pixel-wise, MGAN, Sharpness aware	CT	CT phantom (Catphan 600)	N=10 ⁴ PSNR=26.77 SSIM=0.8454 N=10 ⁵ PSNR=28.25 SSIM=0.87

4.2. Reconstruction

Reconstruction of lost image data (e.g. losing some frequencies through slow sampling) can play an effective role in the diagnosis procedure. Due to the good performance of GANs in the synthesis of unpaired data, they have considerable potential for this task. Here we overview GAN-based reconstruction methods.

In some medical imaging modalities such as MRI, which incurs a long acquisition time, involuntary (i.e. resulted by breathing) and voluntary (i.e. because of not comfortable situation) movement of the patient is very common. These motions lead to loss of some key parts of organ in the image. To address this problem, imaging time reduction is proposed. However, in MR imaging, scan time reduction leads to problems like spatial resolution loss along the z-axis and aliasing in x-y axes. Compressive Sensing (CS) for MRI is the theory that describes how much of these lost data can be reconstructed. While classic solutions directly use k-space information of images to reconstruct missing information [18], GAN based methods try to find a mapping between incomplete (zero-filled) and fully sampled MR Images.

Yu et al. [18] propose to use the U-Net architecture for the generator to extract better details from an input image. Also to consider both pixel wise and feature-based errors in optimization, a combination of loss functions is employed: 1) a pixel-wise MSE, 2) an adversarial loss and 3) a perceptual loss (by comparing VGG extracted features), which helps the network to perform more stable. In addition a refinement layer is added to force the generator to

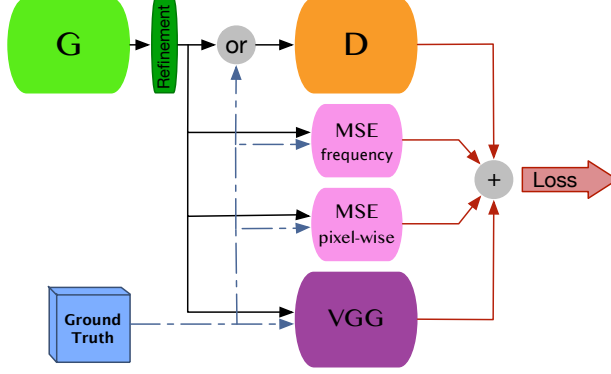


Figure 9: DGAN architecture [19]

generate only missing layers of image. This framework performs about 10 times faster than previous methods and is suitable for real-time reconstruction systems. However, in this paper, frequency domain information is not considered. To address this drawback, in their follow-up publication (DAGAN) [19], the authors added a frequency checking loss function (Fig.9), which is obtained by calculating the MSE in frequency domain. So the final loss for the generator optimization is adjusted as follows:

$$L_G = \alpha L_{image-MSE} + \beta L_{freq-MSE} + \gamma L_{VGG} + L_{GAN} \quad (5)$$

In this way preserving frequency information of the image enhance the performance of the network (2).

Since, a network like DAGAN with simultaneous optimizations with adversarial, MSE, and perceptual loss results low in PSNR, Seitzer et al. [20] proposed to add a refinement network to this architecture in order to separate the pixel-wise and perceptual training procedure. In the proposed architecture firstly, a reconstruction network is trained with MSE loss to learn the details of the image and then a refinement network is considered to fix the visual aspects of the reconstructed image (Fig. 10). To optimize the performance of refinement network 4 different optimization loss is considered as follows:

$$L_{ref} = \frac{1}{2} \left(\frac{L_{adv}}{M} + \frac{L_{feat}}{N} \right) + \frac{L_{VGG}}{O} + \alpha L_{pen} \quad (6)$$

Where L_{feat} is feature matching loss proposed in [21] and L_{pen} is a penalty to force the network to manipulate the result of MSE optimized network with the least changes, and L_{adv} and L_{VGG} are similar to the losses used in the DAGAN. To evaluate the performance of the network in addition to PSNR, mean opinion score (MOS) and semantic interpretability score (SIS) metrics are used. MOS is a subjective metric and SIS is mean Dice overlap between segmentation result on reconstructed image and a real HR image.

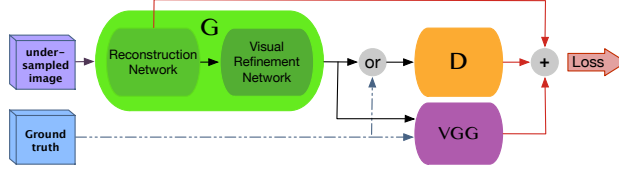


Figure 10: Proposed architecture in [20]

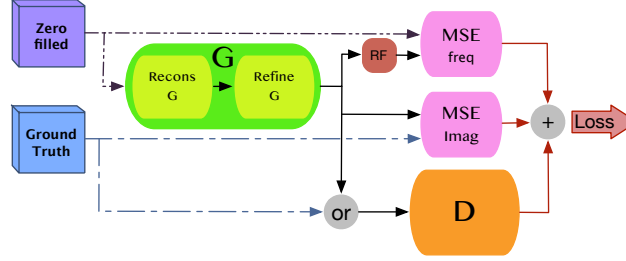


Figure 11: RefineGAN architecture [22]: The Generator G is a chain of two concatenated generators (first generator is for reconstruction and the second one is for refinement) cycle loss is calculated by MSE blocks

Quan et al. [22] propose a different framework (RefineGAN) to reconstruct MRI images, using a combination of convolutional auto-encoder, residual network and GAN architectures. In addition to the loss that the discriminator returns, in a cyclic strategy two other loss functions affect the generator. One of them compares the reconstructed image with the ground-truth (L_{imag}) and the other one compares damaged (zero-filled) reconstructed images with non-reconstructed versions (L_{freq}). The total loss function is defined as follows:

$$L_G = L_{adv} + \alpha L_{freq} + \beta L_{imag} \quad (7)$$

Moreover, they propose to use a chain of generators with similar architecture in which every generator address the ambiguities of the previous one (Fig.11). The results show that this framework not only performs fast enough for real time performance, but also it generates a high quality image even at low sampling rates like 10%.

Mardani et al. [23] also propose using LSGANs to reconstruct MRI Compressive Sensing (CS) images. In the proposed method, the generator is a ResNet with skip connections. To control the instability collapse of classic GAN, LSGAN $L1$ loss is added to general $L1$ loss. According to the results shown in the paper this method is superior in the speed, stability and diagnosis quality in comparison to CNN-based methods.

Shitrit et al. [24] propose an architecture that combine GAN training strategy with ResNet which makes the model able to reconstruct entire k-space grid

from under-sampled data better than other CNN based methods using only 52% of training data of them.

Li et al. [25] propose 3DSRGAN to reconstruct thin slice tomographic 3D images from thick ones. To train the generator four different losses are considered: 1) a pixel-wise loss (l_{MSE}), 2) an adversarial loss(l_{GAN}), 3) a 3D total variation loss (l_{tv}), which controls estimations for absent data using their neighbor slices information, and 4) a weight regularization loss (l_{wr}) to overcome over-fitting problem. The total loss for the generator is defined as follows:

$$L_G = l_{MSE} + \alpha l_{GAN} + \beta l_{tv} + \gamma l_{wr} \quad (8)$$

Another contribution of this method is employing fully 3D CNNs and residual blocks in the generator architecture to avoid gradient vanishing and to provide deep structural training. The results show that this method performs better than nearest neighbor and B-spline interpolation methods. Also 3DSRGAN provides less error in comparison with 2D/3DSRCNN.

Snchez et al. [26] adapted SRGAN [27] with 3D convolutional layers to deal with volumetric information in addition to manipulations which leads to enhance the stability of that. In the upsampling phase of image generation they explored three methods of nearest neighbor interpolation by convolutional layers: 1) resized convolution, 2) 3D adapted sub-pixel convolution method [28] (achieved the best performance in SSIM), and 3) convolutional nearest neighbor resize [29] (achieved the best performance in PSNR). Moreover, to stabilize the training procedure they used batch normalization in almost all layers of the generator in addition to LSGAN. Also, they used two other loss functions: 1) a pixel-wise loss to achieve high PSNR value, and 2) a gradient based loss (GDL) [30] to improve the quality of the generated image. Actually, the second loss is defined to remedy the blurring effect of pixel-wise loss.

To overcome huge memory and time usage of DNN for 3D images (SR) reconstruction, Chen, et al. [31] proposed multy-level densely connected super-resolution network (mDCSRN) which outperforms 6 times faster than other popular DNN methods to recover 4x resolution down-scaled MRI images. The architecture of this model is a combination of WGAN and DenseNet [32]. Although DenNet reduce the number of network parameters dramatically, it is not memory efficient enough for 3D image reconstruction. So in internal networks, authors manipulated the architecture of DenseNet to enhance skip connections significantly.

Ravi et al. [33] proposed to use GAN for unsupervised endomicroscopy super resolution (SR). To constrain the network to save main properties of the low resolution (LR) images, a cyclic consistency approach is considered in which a loss (l_{Vec}) is defined due to the distance between Voronoi vectorized form of SR and LR images. Also a l_{Reg} is defined to regularize the training procedure. In this way, the total training loss function is defined as:

$$loss = l_{adv} + l_{Vec} + L_{Reg} \quad (9)$$

Metrics defined to evaluate the performance are SSIM, $\Delta GCF_{\widehat{HR}}$ (improvement on the global contrast according to the high resolution image) , $\Delta GCF_{\widehat{LR}}$

(improvement on the global contrast according to the LR image and Tot_{cs} (the average value on them), in which the proposed method outperforms other methods in SSIM and Tot_{cs}).

So far, retinal images' resolution is not sufficient enough for small vessel segmentation. Mahapatra [34] address this problem by proposing a new GAN based network called super resolved generative adversarial network, which is able to reconstruct the high resolution retina image from a LR one. While previous methods can not save important local information of image for scales greater than 4, proposed method overcomes this limitation. The key point of proposed architecture is to consider two loss values to tune the generator: 1) an adversarial loss (L_{GAN}) and 2) a CNN loss, weighted by the saliency map of images to save important information of high frequency parts of that ($L_{CNN-sal}$). The final loss to train the generator is defined as follows:

$$L = L_{CNN-sal} + L_{GAN} \quad (10)$$

$$L_{CNN-sal} = \|w_{I_{HR}}I_{HR} - w_{I_{Gen}}I_{Gen}\|_2 \quad (11)$$

Where, $w_{I_{HR}}$ shows the saliency map of high resolution real images (or ground-truth) and $w_{I_{Gen}}$ shows the saliency map of generated SR images. Evaluated results in the paper indicates that the local saliency map played an effective role in preserving structural information. Table 2 and 3 summarizes properties of mentioned methods and their performance. It seems that GANs can provide good performance in reconstruction of medical images, by adding some manipulation in loss functions, which highlights texture details and special features.

Table 2: Reconstruction GAN-based methods in medical image processing - Brain & Chest

Method	Image Modality	Dataset	Performance
[18]			
Arch: cGAN, U-Net Loss: Adv, Pix-wise, Perceptual,Refinement	MRI	Brain: IXI, MICCAI (2013grand challenge)	mask 30%: NMSE=0.09±0.02 PSNR=39.53±4.12 (CPU, GPU) time=0.2±0.1, 5.4±0.1(ms)
DAGAN[19]			
Architecture: cGAN,U-Net Loss: Adv, Pix-wise, Frequency, Perceptual, Refinement	MRI	Brain: IXI, MICCAI (2013 grand challenge)	mask 30%: NMSE=0.08±0.02 PSNR=40.20±4.07 (CPU, GPU) time= 0.2±0.1, 5.4±0.1(ms)
RefineGAN[22]			
Architecture: Chain ofgenerator, ResNet Loss: Adv, Cyclic	MRI	Brain: IXI Chest: Data Science Bowl challenge	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
GANCs [23]			
Architecture: ResNet, LSGAN	MRI(Chest)	contrast-enhanced MRI abdomen dataset ofpediatric patients	SNR=20.48 SSIM=0.87 time=0.02

Table 3: Reconstruction GAN-based methods in medical image processing - Brain & Chest

Method	Image Modality	Dataset	Performance
[24]			
Architecture: ResNet, GAN Loss: Adv	MRI(Brain)	Unknown	PSNR=37.95
[25]			
Architecture: Res blocks, GAN Loss: Adv, Pixel-wise 3D total variation	MRI(Brain)	(glioma patients)	MSE=262.2 PSNR=24.2
[26]			
Architecture: SRGAN, subpixel-NN Loss: LSGAN, GDL, Pixel-wise	MRI(Brain)	(ANDI database)	scale 2: PSNR=39.28 SSIM=0.98 Scale 4: PSNR=33.58 SSIM=0.95
mDCSRN[31]			
Architecture: DensNet, WGAN Loss: MSE, WGAN	MRI(Brain)	Unknown	SSIM=0.94 PSNR=35.88 NRMSE=0.0852

4.3. Segmentation

Annotation of objects and organs in medical image processing plays an important role in anomaly detection and shape recognition. In addition, segmentation is defined as the preprocessing step of many other tasks like detection and classification. So automatic segmentation attracted the attention of a large number of researchers and in recent decades it was the most common subject of papers applied for deep learning in medical image processing [1].

In general, CNN-based segmentation methods utilize a pixel-wise loss which is not adequate to learn local and global relations between pixels. So they need statistical modeling methods e.g. conditional random fields [37] or statistical shape models [38] to correct their results. Although some patch-based CNN methods have been proposed to address this problem, these need to meet a trade-off between accuracy and patch size. Also U-Net based architectures using a weighted cross-entropy loss or the dice-loss are proposed as a solution, but these methods face weight optimization problems. So in addition to a weighted loss, a general loss is required to address this problem.

4.3.1. Brain:

Xue et al.[39] propose a U-Net GAN-based framework (SegAN) in which a multi-scale loss function is used to learn pixel dependencies. In contrast to the original GAN, this loss function is used to train both the generator and

Table 4: Reconstruction GAN-based methods in medical image processing - others

Method	Image Modality	Dataset	Performance
[20]			
Architecture: cGAN,U-Net	MRI Cardic	unknown	PSNR=31.82±2.28
Loss: Adv, feature matching, Perceptual, penalty			MOS=3.24±0.63 (max=3.78±0.45) SIS(max=1)=0.94
[34]			
Architecture: ResNet, GAN	Retinal Funduscopy	Unknown	(Scale 4, Scale 8) SSIM=0.89, 0.84
Loss: Adv, CNN (weighted by SL map)			RMSE=6.2, 7.5 PSNR=44.3, 39db
[33]			
Architecture: [35], GAN	Endomicroscopy	[36]	SSIM=0.8.7
Loss: Adversarial, Cyclic, regularization			$\Delta GCF_{\widehat{HR}}=0.66$ $\Delta GCF_{\widehat{LR}}=0.37$ $Tot_{cs}=0.66$

the discriminator. This framework is trainable without using patches or variant resolution input images and it does not need to use CRF as a correction. As an application, brain tumor segmentation in MRI 3D images is investigated. The loss function is defined as follows:

$$\min_{\theta_G} \max_{\theta_D} L(\theta_G, \theta_D) = \frac{1}{N} \sum_{n=1}^N l_{mae}(f_C(x_n \circ S(x_n)), f_C(x_n \circ y_n)) \quad (12)$$

Where l_{mae} is Mean Absolute Error (MAE) or L_1 distance, $(x_n \circ S(x_n))$ is the input image masked with a generated segmentation mask, $(x_n \circ y_n)$ is an input image masked by the ground-truth segmentation mask, $f_C(x)$ shows features extracted from the input image x and the l_{mae} is the Mean Absolute Error.

Rezaei et al. [40] also focus on the same application and propose a multi-class approach, in using a combination of cGAN and MGAN models. To overcome the well-known mode collapse phenomenon seen in GANs, Virtual-BatchNorm and Reference-BatchNorm [41] are proposed to train the generator and discriminator, respectively.

Moeskops et al.[42] demonstrate that using GANs training strategy in addition to DCNN methods not only can enhance the performance of deep semantic segmentation methods, but also can bring the functionality of non-semantic segmentation methods closer to semantic ones.

For brain tumor segmentation Zeju et al. [43] proposed a pipeline of pre-processing, GAN, and post-processing steps. The first step contains intensity normalization and mean/distribution equalization. Then the GA segments tumor in patches of preprocessed image and finally, in the last step patches are concatenated to specify the whole area of the tumor.

Since the performance of most of the supervised segmentation methods degrades on unseen images, Kamnitsas et al. [44] proposed unsupervised domain adaption for brain lesion segmentation. In this method the generator extracts invariant features of inputs from different domains and then generates the segmentation mask. In this way, having data of a target domain corresponds to one of the input domains can lead the mapping procedure from other inputs (from different domains) to their corresponding targets.

4.3.2. *Chest:*

Bad quality, local artifacts and the overlap of lung and heart area are the main obstacles for the segmentation procedure in chest X-Ray images. Existing approaches on this field do not provide a balance on global and local features. So they are not realistic segmentation methods for diagnosis tasks. Dai et al.[45] propose a GAN based solution (SCAN) to enhance global consistency of segmentation and extract contours of the heart and left/right lungs. The main contribution of this work is to use a fully connected network with a VGG down-sampling path using much fewer feature maps in the generator. In addition, residual blocks are employed to aid the optimization. This framework segments the RoI with human level performance, while using a limited amount of training data. To address instability drawbacks of GANs through the training procedure, the generator is pre-trained by pixel-wise loss.

4.3.3. *Eye:*

In retinal vessel segmentation, many CNN-based approaches performed even better than human experts. But segmented vessels can be blurred or contain false positive areas near minuscule or faint branches. Son et al.[46] replace the CNN with a GAN following the U-Net architecture for the generator. The experimental results on two datasets show that leveraging a traditional full-image discriminator leads to the best performance, even better than human expert’s annotation.

Lahiri et al.[47] propose a DC-GAN-based segmentation method which segments RoI patches from the background. While a similar CNN needs a huge amount of training data to perform well, the proposed structure achieves comparable performance using 9 times less training data.

Shankaranarayana et al. [48] proposed to use cGAN network to segment optic disc and cup in 2D color fundus images. The generator is a ResU-net network which is trained by adversarial and $L1$ losses. Results of the paper show that in such a network using cGAN enhances the segmentation of small, challenging ROI parts (cup), while GAN performs better in segmenting larger ROI parts (optic disc).

4.3.4. Abdomen:

Varying size and shape of the spleen in abdomen MRI images lead to false labeling in deep CNN segmentation methods. Huo et al.[49] employ a new GAN-based method (SSNet: splenomegaly segmentation network) to address this problem. In the proposed model the Generator is a novel deep network architecture inspired by the global convolutional network, which uses larger convolutional kernels to have better segmentation on objects with large variations. On the other hand, the discriminator follows the cGAN architecture to alleviate the false positive rate. Presented results in the paper show that this method achieves higher robustness and accuracy in comparison to benchmark methods (U-Net and GCN), reducing the false negative rate. Also it is shown that using two or three views of abdomen images in both training and testing enhance the performance of the network.

Yang et al.[50] propose a liver segmentation method in 3D abdomen CT images. The generator is a convolutional encoder-decoder inspired by the U-Net architecture. In practice this method enhances the accuracy of segmentation benefiting adversarial loss in addition to multi-class entropy loss.

Also, Kim et al. [51] proposed to use cycleGAN for liver and tumor segmentation. In this architecture one generator generates a segmentation mask from input image and the other one generates CT image from the segmentation mask. In order to enhance the performance of the model in segmenting tiny tumors, polyphase U-Net architecture is proposed to be used as the generator, because it retains the high frequency information and does not change the polarity of the input.

4.3.5. Microscopic images:

Automatic segmentation of this kind of images face some challenges due to the variety of size, shape, and texture of them [52, 53]. Kecheril et al. [52] proposed to use GAN with different training loss function, which considers a weight to specify which pixels if foreground/background are more important. The proposed architecture is a combination of U-net with long/short skip connections, ResNet, and multi-scale CNN. In addition, a post-processing procedure is proposed to correct the segmented area.

Also, Arbelle et al.[53] used GAN for cell segmentation. They proposed a GAN architecture in which rib cages - CNN blocks followed by a batch normalization are used in discriminator network. Results show that not only this architecture outperforms single CNN architectures, but also the number of training images does not affect the performance of the model strongly.

Moreover, Zhang et al. [54] proposed an adversarial network for biomedical image segmentation called DAN. The architecture of this segmentation network is a combination of DCAN [55] and VGG16. The training dataset consists of annotated (M images X_m and their paired ground-truth Y_m) and un-annotated images (N images U_n). The network in the supervised training learns how to segment the ROI and in the unsupervised training learns how to generate same quality segmentation map for unseen data. Two loss functions are considered to train the network: adversarial loss (for un-supervised training) which is defined

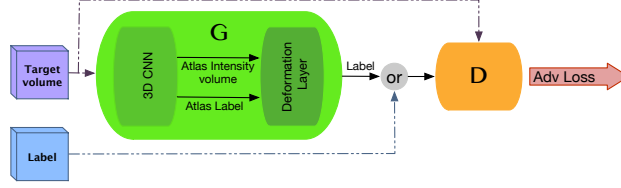


Figure 12: Proposed architecture in [56]

as binary loss l_{bce} to evaluate the quality of segmentation and multi-class cross-entropy loss L_{mce} to train the network to generate the segmentation map. The loss function is defined as follows:

$$L = \sum_{m=1}^M l_{mce}(G(X_m), Y_m) - \lambda \left[\sum_{m=1}^M l_{bce}(D(G(X_m), X_m), 1) + \sum_{n=1}^N l_{bce}(D(G(U_n), U_n), 0) \right] \quad (13)$$

4.3.6. Cardiology:

Left ventricle (LV) segmentation in 3D echocardiography as a real-time medical imaging provides a large volume of information about the patient situation. However, low contrast, high level of noise and automatism movement of data in echocardiography images challenge this procedure. Dong et al. [56] proposed VoxelAtlasGAN which combines an atlas-based segmentation method with cGAN architecture to segment LV in low-contrast cardiography images. In this method first the shape and intensity of the atlas is estimated by a CNN (V-Net [57]) and then a deformation network outputs the segmented image (Fig. 12). Both of the mentioned networks are placed in generator, which uses three loss functions for training: 1) Adversarial loss, 2) intensity loss, and 3) label loss, which compare the intensity and shape of the segmented real image with the generated one respectively. Using atlas-based segmentation prior to cGAN enhances the segmentation performance and interpretability of the model. In the paper it is shown that using cGAN decreases the complexity and time in comparison with other atlas-based methods while it needs less training data to be learned.

4.3.7. Spine:

Vertebrae segmentation and localization is the first step for diagnosis of the vertebrae disease and surgery planning. Although machine learning based approaches achieved some success in this field, they suffer from not learning the anatomy of the region of interest. To overcome this problem a solution is to deepen the network to increase the receptive field, which faces the memory limitation. To address this problem, Sekuboyina et al. [58] proposed a butterfly shape model benefiting adversarial training to segment and localize discs in vertebra CT images. The main idea behind the architecture of the generator

is using two views of CT images to capture both spine curve the and the rib-vertebrae joints. Firs, in a pre-processing step, the region of the spine is selected using the single-shot object detection (SSD) [59] method. Then the proposed model segments discs in two views of vertebrae and finally, in a post-processing step, these results are combined for disc localization.

Tables 5 to 10 summarizes GAN-based segmentation methods. GAN-based segmentation methods mainly worked on architectural subjects to address previous methods and GANs drawbacks. It seems that from the known DNN architectures, U-Net and ResNet - due to providing general identification features - are the most popular networks to be used as the generator in segmentation GAN-based models.

Table 5: Segmentation GAN-based methods in medical image processing-Brain

Method	Image Modality	Dataset	Performance
SeGAN[39]			(whole, Core, Enhanced)
Architecture:		BRATS 2013	Dice = 0.84, 0.70, 0.65
U-Net, GAN		(Leadboard)	Precision = 0.87, 0.80, 0.68
Loss:			Sensitivity = 0.83, 0.74, 0.72
Adv, weighted on	MRI	BRATS 2015	Dice = 0.85, 0.70, 0.66
multiScale features		(Test)	Precision = 0.92, 0.80, 0.69
			Sensitivity = 0.80, 0.65, 0.62
[40]			(Whole, core, Enhanced)
Architecture:		BRATS 2017	Dice = 0.70, 0.55, 0.40
c-GAN, MGAN	MRI		Sensitivity = 0.68, 0.52, 0.99
			Specificity = 0.99, 0.99, 0.99
[42]			
Architecture:		MICCAI 2012	Dice = 0.92±0.03
GAN	MRI	Challenge (adult)	
Loss:		MRBrainS13	Dice = 0.85±0.01
Adv, cross entropy		challenge(elderly)	
[43]			
Architecture:			(Whole, Core, Enhancing)
GAN	MRI	BRATS 2017	Dice = 0.87, 0.72, 0.68
Loss:			sensitivity = 0.87, 0.72, 0.68
Adv			
[44]			
Architecture:			Dice = 0.62
GAN, 3D-CNN	MRI (TBI)	unknown	Recall = 0.58
Loss:			Precision = 0.71
Adv, SGD			

4.4. Detection

In medical diagnosis many disease markers are known as anomalies. However, computational detection of anomalies from images requires a large amount of supervised training data. Even if such a huge database is available, there is no guarantee that a learned network is able to detect unseen cases.

Table 6: Segmentation GAN-based methods in medical image processing-Chest

Method	Image Modality	Dataset	Performance
SCAN[45]			
Architecture: VGG, ResNet	X-Ray	JSRT(247) Montgomery(135)	(Lungs, Heart)
Loss: Adv, pre-trained by Pixel-wise loss			Dice = 0.973, 0.927
			IoU = 0.947, 0.866

Table 7: Segmentation GAN-based methods in medical image processing-eye

Method	Image Modality	Dataset	Performance
[46] Architecture: U-Net, GAN Loss: Adv, Cross entropy	Funduscopy (Retina)	DERIVE STARE	Dice= 0.829 ROC=0.9803 PR=0.9149 Dice= 0.834 ROC=0.9838 PR=0.9167
[47] Architecture: DCGAN Loss: Adv, L-classification	Funduscopy (Retina)	DERIVE (blood vessels)	AUC= 0.945
[48] Architecture: c-GAN, ResU-net Loss: Adv, $L1$	Funduscopy (Retina)	RIM-ONE	(Optic disc, Optic cup) F-score= 0.97, 0.94 IOU=0.89, 0.76

Schlegl et al. [61] show that an unsupervised GAN-based architecture (AnoGAN) can detect anomalies in optical coherence tomography images of the retina. In this method, during training on healthy images, a GAN learns a mapping from the latent space to 2D healthy images. During testing, the GANs' latent code is optimized for the reconstruction of a new unseen input image and generates the corresponding healthy image version. Anomalies cannot be reconstructed from the GAN. Then, the generated image and test input are compared and differences are considered as anomalies. To capture the nearest latent value to the input image, a loss function based on visual (pixel wise cross entropy) and feature based similarity compares the generated and input real images. This loss function is used in both the training and detection (testing), which provides more stability for the model.

Chen et al. [62] employed a manipulated version of [61] for brain lesion detection in MRI images. They proposed to use WGAN with gradient penalty to have stable training and also enhance the coverage in the latent space. Due to

Table 8: Segmentation GAN-based methods in medical image processing-Abdomen

Method	Image Modality	Dataset	Performance
SSNet[49]			
Architecture: GCN, cGAN	MRI (Splenomegaly)	Unknown	Dice=0.9260
Loss: Adv, Dice			
[50]			
Architecture: U-Net, encoder-decoder	CT 3D (Liver)	MICCAI-SLiver07	Dice=0.95 ASD=1.90
Loss: Adv, multi-class entropy			
[51]			
Architecture: U-Net, cycleGAN	CT 3D (Liver)	LiTS2017	(liver, lesion) Dice= 0.89, 0.46 Recall=0.94, 0.5 Precision=0.86, 0.48
Loss: cycleGAN, cross entropy, L2			

Table 9: Segmentation GAN-based methods in medical image processing-Microscopic

Method	Image Modality	Dataset	Performance
[52]			
Architecture: GAN, U-net, res-Net, Multi scale CNN	Bright-fieled cell 2D	Columbus	F-score = 0.77 Precision = 0.82 Recall = 0.73
Loss: Adv, weighted loss		MetaXpress	F-score = 0.64 Precision = 0.66 Recall = 0.66
[53]			
Architecture: GAN (with rib cage)	cell 2D	H1299	F-score = 0.89 Precision=0.82 Recall = 0.85
Loss: Adv			
DAN [54]			
Architecture: GAN, DCAN, VGG	fungus 3D	2015 MICCAI Gland Challenge	(mean of 2 part results) F-score = 0.88 Dice=0.865 ObjectHausdorff = 74.55
Loss: Adv, multi-scale cross entropy			
[56]			
Architecture: cGAN, V-Net	Echocardiography 3D	unknown	Dice=0.95 MSD=1.85 HSD=7.26 corr-of-EF=0.91 time=0.1
Loss: Adv, intensity, label			

Table 10: Segmentation GAN-based methods in medical image processing-Spine

Method	Image Modality	Dataset	Performance
Btrfly Net [58]			
Architecture: GAN, Btrfly-Net	CT 3D	[60]	Precision=0.84 Recall=0.83 F1-score= 0.84
Loss: Adv, Btrfly-Net			

high variability of brain MRI images it is likely that, the distance between the abnormal image and its corresponding healthy generated one be less than the distance between the healthy image and its corresponding generated one. To address this drawback they added a regularisation loss function which controls the similarity between the real and generated images and also between their latent values.

Similar to [61] and [62] Baur et al. propose [63] for anomaly detection and delineation in brain MR images, while they address the expensive procedure of iterative optimization of the latent space in [61] and [62]. The proposed method provides a stable reconstruction of entire brain MR slices at higher resolution. In this method a VAE-GAN, i.e. a combination of a generative Variational Autoencoder and GANs are trained on brain MR slices of healthy anatomy. Similarly, during inference, they try to reconstruct the input sample to measure the discrepancy between the input and the reconstructed image to detect anomalies.

Although CNNs show good performance in detection of bold lesions, detection of the lesions with lower attributes challenges their performance [64]. To address this problem Baumgartner et al. [65] propose a map generator based on WGAN and the U-Net architecture (VA-GAN) to detect changes of the brain related to Alzheimer disease. To achieve this goal the generator is trained to generate a map M which converts the class of the image x_i from healthy to sick (y_i) if be added to the image. On the other hand, the discriminator optimizes the generators performance through following loss equation:

$$L_{GAN}(M, D) = \mathbb{E}_{x \sim p_d(x|c=0)}[D(x)] - \mathbb{E}_{x \sim p_d(x|c=1)}[D(x + M(x))] \quad (14)$$

To avoid the generator from being optimization by changing the identity of the subject, another loss function is considered. This loss encourages M to do mapping with smallest manipulations on the healthy image. The final optimization loss function is defined as:

$$L = L_{GAN}(M, D) + \lambda \|M(x)\|_1 \quad (15)$$

Due to this trained map, a sick brain can be mapped to a healthy representation and changes discovered by the mapping detect anomalies of Alzheimer and also the class of them.

To demonstrate that using the training strategy of GAN enhances the performance of cross-entropy U-Net detection, Kohl et al. [66] implement it on aggressive prostate cancer detection. They show that GANs provides better detection on every amounts of training samples in comparison with a single U-Net model.

Similarly, in skin lesion detection, Udrea et al. [67] show that using the combination of U-Net and c-GAN enhances the accuracy of the performance of the model to more than 90%.

Tuysuzoglu et al. [68] benefited adversarial training to detect the whole contour of gland in prostate ultrasound images from detected landmarks. In the first step a CNN model detects the landmarks on the boundary of the

gland. Then the proposed models maps these landmarks (in a pixel level) to the whole contour. Since the contrast of gland tissue is not high enough to be used for boundary detection an adversarial training is proposed to consider contour general features in addition to pixel level information.

Table 11 summarize these papers. Papers proposed in anomaly detection by GANs have more structural complexity in comparison with previous applications because they benefit from different aspects of GANs. In fact, the role of the discriminator is more highlighted in practice. Also, the extracted map, which defines the latent aspect of recognizing the healthy and anomaly images is used in a more perceptual way.

Table 11: Detection GAN-based methods in medical image processing.

Method	Image Modality	Dataset	Performance
AnoGAN[61] Architecture: DCGAN	SD-OCT scans	Unknown	Precision= 0.8834 Recall= 0.7277 Sensitivity=0.7279 Specificity=0.8928 AUC=0.89
[62] Architecture: AnoGAN, WGAN-GP Loss: WGAN-GP, regularization MRI(brain)			
VA-GAN[65] Architecture: WGAN, U-Net	MRI (brain)	ADNI	AUC = 0.92
[66] Architecture: U-Net, GAN Loss: MSE, GAN			NCC = 0.27
[67] Architecture: cGAN, U-net	MRI (prostate)	(NCT) Heidelberg	specificity=0.98±0.14 Dice=0.41±0.28 Sensitivity=0.55±0.36
[68] Architecture: GAN Loss: Adv, landmark location, contour association			
	Skin lesion (natural image)	Unknown	correct lesion detection= 0.914
	US (prostate)	unknown	Dice = 0.92±0.3

4.5. Classification

Due to cardiac and respiratory motions occurring during cardiac ultra-sound (US) imaging, resulting images might display incomplete information, like basal and apical slices of the heart which are key specifics to recognize Left Ventricular (LV) anatomy. Thus, an automatic system is needed to complete the missing

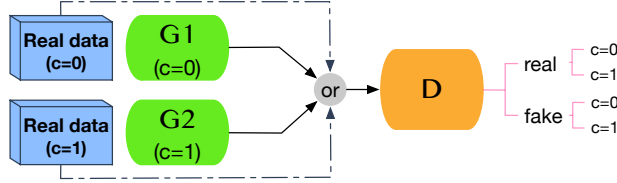


Figure 13: SCGAN architecture [69]

parts or to discard images with incomplete information, which can mislead the classification process.

As a solution for discarding unsuitable images, Zhang et al. [69] propose the Semi-coupled GAN (SCGAN) to classify useful cardiac images from ones with missing basal slices. The framework consists of two generators and one discriminator. Initially, the generators produce new cardiac samples (with and without the basal slice) using learned high level features from both categories. Then, the multi-class multi-label discriminator not only distinguishes between generated and real images but also classifies images into two classes: those being basal slice and not being basal slice (Fig. 13). Results show that this method achieves higher accuracy and reduce computation cost in comparison with CNN methods. In addition, SCGAN improves the robustness of adversarial training.

4.6. Synthesis

Originally, GANs have been proposed as an entirely unsupervised generative framework, with the goal to map from random noise to synthetic, realistically looking images following the training data distribution. With the conditional GAN, the framework has also been successfully turned into a supervised generative framework by conditioning both the generator and the discriminator 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. We want to emphasize that it is very important to make a distinction between these different concepts and consequently categorize the literature accordingly.

The generative property of both frameworks has been exploited in various ways for synthesizing certain types of medical images either from noise alone (see *Unconditional Image Synthesis*), or from prior knowledge (see *Conditional Image Synthesis*) such as metadata or even image data for mapping images from one modality to another. In the following, a broad overview on works from unconditional and conditional image synthesis will be given. In the particular case for conditional approaches, we further classify the contributions based on the image modality. For the literature on unconditional image synthesis we do not make this distinction due to the small amount of papers.

4.6.1. Unsupervised Image Synthesis:

A great variety of works has recently appeared in the field of unsupervised medical image generation using GANs. The synthesis of realistically looking medical images opens up many new opportunities to tackle well-known deep learning problems such as class imbalance, data augmentation [70] or the lack of labeled data. Further, it facilitates data simulation [71] and aids to gain deeper insights into the nature of data distributions and their latent structure.

Initial results have shown that GANs can be used to synthesize realistically looking patches of prostate lesions [72] or retinal images [61]. Both approaches rely on the DCGAN architecture to synthesize patches at a resolution of 16×16 px and 64×64 px, respectively. In [71], the authors successfully utilize DCGANs for generating 56×56 px patches of lung cancer nodules which could hardly be distinguished from real patches in a visual turing test involving two radiologists.

Frid-Adar et al. [70] make use of the DCGAN for the synthesis of focal CT liver lesion patches from different classes at a resolution of 64×64 pixels. For each class, i.e. cysts, metastases and hemangiomas, they train a separate generative model. As the training dataset is originally quite small, they use heavily augmented data for training 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 Convolutional Neural Network classifier.

The work in [73] has shown that the DCGAN with vanilla training is in fact also able to learn to mimic the distribution of MR data at considerably high resolution, even from a surprisingly small amount of samples. The real data distribution consisted of only 528 midline T1-weighted axial MR slices at a resolution of 220×172 px. After training for 1500 epochs, the authors obtained visually compelling results which human observers could not reliably distinguish from real MR midline slices.

In [74], the authors utilize and compare both DCGAN, LAPGAN and modifications of the latter for the task of skin lesion synthesis at a resolution of 256×256 px. Similar to [73], the training dataset was quite small, consisting of only 1,600 images. Probably 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 successfully been used for data augmentation when training a skin lesion classifier. In [75], the same authors employed the recently proposed concept of progressive GAN growing for synthesizing images of skin lesions and showed stunning, highly realistic synthetic images which even expert dermatologists could not reliably tell apart from real samples.

4.6.2. Conditional Image Synthesis:

CT from MR In many clinical settings, the acquisition of CT images is required. This, however, puts the patient at risk of cell damage and cancer because of the radiation exposure, which motivates the synthesis of CT images

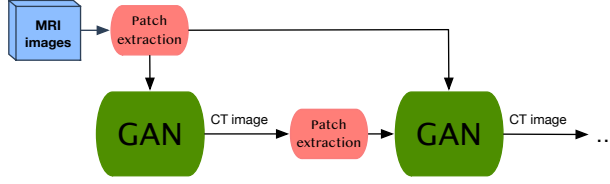


Figure 14: Proposed architecture in [22]

from MR acquisitions. Nie et al. [76] synthesize CT images from corresponding MR images with the help of a cascade of 3D Fully Convolutional Networks which they train with a normal reconstruction loss, an image gradient loss and additionally with an adversarial network in order to improve realism of the synthetic CT images. The idea of utilizing a cascade of generator networks originates from the so-called Auto-Context Model, in which a network provides its output as additional input to a succeeding network in order to provide context information and allow for refinements (Fig. 14). While Nie et al. require corresponding pairs of CT and MR images for training, Wolterink et al.[77] successfully utilize Cycle-GANs to transform 2D MR images to CT images without the need for paired, co-registered training data. Interestingly, in contrast to training from paired, co-registered data, their training led to even better results as the model avoids to learn mappings in the presence of registration artifacts.

MR from CT Similar to Wolterink et al., Chatsias et al. [78] successfully leverage CycleGANs for unpaired image-to-image translation, however for synthesizing pairs of cardiac MR images and a segmentation mask from pairs of cardiac CT slices with the ground-truth segmentation mask. The authors have shown that the performance of a segmentation model can be improved by 16% when additionally trained with the synthetic data, and that synthetic data alone is sufficient for training a model which performs only 5% worse than a model trained on real data.

Retinal Image Synthesis In [79] the authors utilize a slight modification of the adversarial training concept proposed in [4] for the challenging task of eye fundus image generation. They learn a mapping from binary images of vessel trees to new retinal images at a resolution of 512x512px, which look extremely realistic and rate very well in common scores for retinal image quality judgement. In follow-up work [80], the authors further introduce an adversarial autoencoder which is trained to compress vessel tree images into a multivariate normal distribution and to consecutively reconstruct them. The resulting generative autoencoder allows to synthesize arbitrary high resolution vessel tree images by sampling from the multivariate normal distribution. The synthetic images in turn are fed into the image-to-image translation model, ultimately leading to an end-to-end framework for realistic, high resolution retinal image synthesis. Very similarly, Guibas et al. [81] propose a two-stage approach, consisting of a GAN which is trained to synthesize vessel tree images from noise, and a second conditional GAN as seen in Pix2Pix [4] to generate realistic, high

resolution pairs of groundtruth vessel segmentation and the corresponding eye fundus image. In succession, they investigate the performance of a U-Net trained for segmentation using real data pairs and another model trained only on the synthetic samples, and find that training from only the synthetic data leads to an only slightly inferior model.

In [82], the authors also leverage the Pix2Pix framework for the tasks of synthesizing filamentary structured images, i.e. eye fundus images and neurons from binary segmentation masks. Compared against [79, 80], the authors also provide their framework with a reference image for style and train the generator also with the feedback from an additional VGG-network leveraged for style transfer. and show that only 10 training examples are sufficient for training such an image-to-image translation model. Opposed to Pix2Pix, they do not introduce noise with the help of dropout, but by augmenting noise to the bottleneck of the encoder-decoder network. In a set of use-case experiments on retinal image segmentation it is demonstrated that the introduction of additional synthetic images, i.e. training from both real and synthetic images, slightly improves the segmentation performance.

PET from CT PET images are frequently used for diagnosis and staging in oncology, and the combined acquisition of PET and CT images is a standard procedure in clinical routine. Furthermore, PET/CT imaging is becoming an important evaluation tool for new drug therapies. However, PET devices involve radioactivity and thus put patients at risk, and are expensive in general. Consequently, the medical image analysis community has been working on synthesizing PET images directly from CT data. In this context, GANs have also shown outstanding performance. Initial promising results for synthesizing liver PET images from CT data with conditional GANs have been obtained in [83]. The conditional GAN, again inspired by [4], is able to synthesize very realistic looking PET images, however at the cost of low response to underrepresented tumor regions, which leads to poor tumor detection performance in a set of use-case experiments. In contrast, the authors find that 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, though. Similarly, in [84] the authors utilize a conditional GAN for synthesizing 200×200 px sized PET images from pairs of CT images and binary labelmaps. While CT images alone would be sufficient as input, they note that by adding a labelmap which marks the location of a tumor, they obtain 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, showing that synthetic data can in fact be beneficial when there is a lack of labeled data.

PET from MRI For monitoring disease progression, understanding physiopathology and evaluate treatment efficacy of Multiple Sclerosis (MS), measuring the myelination content in PET images of the human brain has recently

shown to be very valuable. Unfortunately, PET imaging for MS is costly and invasive as it requires the injection of a radioactive tracer. In [85], the authors successfully utilize a cascade of two conditional GANs for synthesizing such PET images from a set of different MR modalities. Their approach operates directly on volumetric data, leveraging a 3D U-Net for the generator networks and discriminator networks with 3D convolutions. Interestingly, the authors noted that a single conditional GAN was insufficient for the task at hand as it produced blurry images. Splitting the synthesis task into smaller, more stable subproblems, seemed to drastically improve the results.

Ultrasound Hu et al. [86] propose a conditional GAN architecture for synthesizing 2D ultrasound images of a fetus phantom, as produced by a freehand US probe, given 3D spatial pixel locations within the anatomy. In contrast to the standard conditional GAN, the authors find it necessary to transform the pixel locations into featuremaps and to concatenate them with the produced featuremaps at each level of the generator to facilitate training. In their experiments they demonstrate the capability of simulating US images at locations unseen to the network, quantify the generation of sound images by comparing the location of clinically relevant anatomical landmarks in synthetic and real images, and verify the realism of the generated images in a usability study. The quantitative results show that anatomical landmarks are roughly synthesized at the right locations with a mean error of 6.1mm. In their usability study, the sonographer was able to mostly correctly distinguish between real and generated samples, which is due to checkerboard artifacts in the synthetic images. After blurring the images using a gaussian kernel with $\sigma = 1.5$, the sonographer was not able to reliably tell the difference anymore. The interested reader is also referred to the NiftyNet framework [87], in which this conditional GAN is contained. Tom et al. [88] apply GANs for intravascular ultrasound (IVUS) simulation in a multi-stage setup. A first generator conditioned on physically simulated tissue maps produces speckle images, which in turn act as the conditioning input to a second residual network based generator. The second generator maps the speckle images to low resolution, synthetic 64×64 px sized US images. A third generator transforms these low resolution images into high resolution samples at a resolution of 256×256 px. In a visual turing test, the synthetic images could not reliably be distinguished from real ones.

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. Cho et al. [89] point out that a tumor classifier generalizes poorly on both data with staining properties different from the training set, as well as on images that have been stain-normalized with state-of-the-art methods. To overcome these issues, they propose a feature-preserving conditional GAN for stain style transfer with the particular goal to prevent a degradation in performance of CAD systems on synthetic images. First, they map histological images to a canonical gray-scale representation. In succession, they leverage a conditional GAN to transform gray-scale images into RGB images with the desired staining. By employing an additional feature-preserving loss on the hidden layers of the discriminator, they demonstrate that a tumor

classifier model trained on data stemming from a certain distribution performs better on the stain-transferred images than on the original ones, and that their conditional GAN shows the smallest degradation in performance compared to other state-of-the-art stain transfer methods.

Bayramoglu et al. [90] leverage the Pix2Pix framework for virtual H&E staining on unstained hyperspectral microscopy images using 64×64 px sized patches. The authors report the SSIM and MSE between synthetically stained images and the ground-truth and point out to have obtained promising result, but require expert feedback in order to draw a valid conclusion.

BenTaieb et al. [91] try to tackle the stain transfer problem with the help of a so-called Auxiliary Classifier 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 preserved for the task-specific model and overall leads to superior results in stain-normalization compared to other state-of-the-art methods.

Aformentioned methods rely on paired training data to map from a source to target staining, which is often hardly available and requires preprocessing such as co-registration. However, co-registration itself is not perfect and is prone to artifacts. Shaban et al.[92] alleviate 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 show visually much more compelling stain transfer results than previous deep-learning and non-deep learning based methods. In addition, they also show quantitatively how their approach significantly reduces domain shift which usually hampers deep learning models: A classifier trained for mitosis detection provides much better classification results on images stain-transferred with the proposed approach than on original data, and again also other stain transfer methods.

Microscopy Han et al. [93] propose a conditional GAN framework similar to Pix2Pix for transferring between Phase Contrast and Differential Interference Contrast (DIC) Microscopy images, however with two discriminator networks rather than one. A U-net like generator is trained to synthesize the image of a certain modality from an image of the source modality and a cell mask. Two different discriminators then either discriminate between pairs of real source and target modality images versus pairs of real source and synthesized target modality image, or pairs of cell mask and real source versus cell mask and synthesized target images. In a set of qualitative and quantitative evaluations they rank their two-discriminator approach against the Pix2Pix framework which uses only a single discriminator. They report improved results in the metrics of SSIM and normalized RMSD when transferring from DIC image to Phase Contrast, and comparable results when trying to map from Phase Contrast to DIC. Noteworthy, the authors amount the comparable performance of the latter mapping to the details already present in Phase Contrast images, which leaves the cell mask with very little impact on the synthesis outcome.

Blood Vessels Machine Learning driven analysis methods for detecting atherosclerotic plaque or stenosis in coronary CT angiography (CCTA) are pow-

erful, but data-hungry. To deal with the lack of labeled data, Wolterink et al.[94] propose to synthesize plausible 3D blood vessel shapes with the help of a Wasserstein GAN from 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 and act as a proxy for the final vessel rendering. Magnetic Resonance Angiography (MRA) has also evolved into an important tool for visualizing vascular structures, but often times it is not acquired alongside the standard protocols. In [95], the authors propose the so-called steerable GAN for synthesizing 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, an ℓ_1 -loss between real and synthesized image as well as a steerable filter loss to promote faithful reconstructions of vascular structures.

Tables 12, 13 and 14 give an overview of all the presented image synthesis methods. The unconditional synthesis methods are summarized in Table 12, whereas the conditional GAN variants are summarized in Table 13 and 14. In particular, we report the method, i.e. the underlying GAN architecture, the image modalities on which the particular method operates, the datasets which have been used and the resolution of the synthesized images. 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 case unavailable, ii) hardly interpretable and iii) overall hardly comparable. In general, many interesting GAN-based approaches have been made for both unsupervised and conditional image synthesis. However, often the validity of the method at hand is questionable and requires more elaboration. For instance, in many visual turing tests it is fairly easy to distinguish between real and generated images [86, 71, 70] due to artifacts in synthetic samples, such as the well known checkerboard pattern. In [86, 71], the authors tackle this problem by applying anisotropic or gaussian filtering to both real and fake samples before presenting them to the raters [86, 71], which is only valid as long as blurry images still contain the required amount of information for the task at hand. Another problem is that GANs are prone to the phenomenon of mode collapse, in which the model is only able to generate samples stemming from one or a few modes of the real data distribution, resulting in very similar looking synthetic samples. Particularly in the works of [72] and [61], where samples look fairly similar, a thorough elaboration on whether mode collapse has occurred or not would have been very interesting. In general, the community still lacks a meaningful, universal quantitative measure for judging realism of synthetic images. Regardless of the realism, aforementioned works have shown that GANs can be used successfully for data simulation and augmentation in classification and segmentation tasks. How realism, artifacts in and specific properties of generated samples affect a machine learning model when used for data augmentation also remains an open question.

Table 12: Unconditional GANs for Medical Image Synthesis

Method	Image Modality	Dataset	Resolution
[72] Architecture: DCGAN	MRI Prostate Lesions	SPIE ProstateX Challenge 2016	16×16
[71] Architecture: DCGAN	CT Lung Cancer Nodules	LIDC-IDRI	56×56
[70] Architecture: DCGAN	focal CT liver lesion patches	non-public	64×64
[73] Architecture: DCGAN	2D axial brain MR slices	Baltimore Longitudinal Study of Aging (BLSA)	220×172
[74, 75] Architecture: DCGAN,LAPGAN,PGAN	Dermoscopic Images of Skin Lesions	ISIC2017 & ISIC2018	256×256

Table 13: Conditional GANs for Medical Image Synthesis

Method	Image Modality	Dataset	Resolution
[76] Architecture: 3D Autocontext FCN with adversarial loss, image gradient loss and ℓ_2 -loss	MR to CT	ADNI and 22 non-public pelvic image pairs	n/a
[77] Architecture: CycleGAN	2D sagittal brain MR and CT slices	non-public	256×256
[78] Architecture: CycleGAN	2D cardiac MR w. segmentation mask to cardiac CT w. segmentation mask	non-public	232×232
[79, 80] Architecture: AAE and Pix2Pix	2D binary vessel tree images to retinal images	DRIVE, MESSIDOR	512×512
[96] Architecture: 3D cond. GAN	3D volumes of lung nodules	LIDC	64×64×64
[81] Architecture: GAN and Pix2Pix	2D binary vessel tree images to retinal	DRIVE, MESSIDOR	512×512
[82] Architecture: Pix2Pix w. Style Transfer	eye fundus images, microscopic neuronal images	DRIVE, STARE, HRF, NeuB1	512×512 and higher
[83] Architecture: Pix2Pix and FCN	2D liver tumor CT to PET images	non-public	n/a

5. Discussion

5.1. Overview

GANs is receiving significant attention from the medical imaging community - this is evident by the sudden spike in the number of papers published using GANs. We found a total of 63 papers by searching Google Scholar and PubMed with 'GAN' or 'Generative Adversarial Networks' in title or keywords. Of these,

Table 14: Conditional GANs for Medical Image Synthesis

Method	Image Modality	Dataset	Resolution
[84] Architecture: conditional multi-channel GAN	CT and binary segmentation pairs to PET images	non-public	200×200
[86] Architecture: spatially cond. GAN	2D US	non-public fetus phantom	160×120
[88] Architecture: multi-stage cond. GAN	simulated tissue maps to 2D Intravascular US	IVUS challenge	256×256
[89] Architecture: feature-preserving conditional style- transfer GAN	Digital Histopathology	CAMELYON16	n/a
[90] Architecture: Pix2Pix	Hyperspectral microscopic images to H&E stained images	non-public	64×64
[91] Architecture: ACGAN	Digital Histopathology	MITOS-ATYPIA14 MICCAI16 GlaS challenge non-public ovarian carcinoma	250×250
[92] Architecture: CycleGAN	Digital Histopathology	MITOS-ATYPIA14 Camelyon16	256×256
[93] Architecture: cond. GAN with two Discriminators	DIC & Phase Contrast Microscopy	non-public	256×256
[94] Architecture: WassersteinGAN	Geometric parameters extracted from CCTA	non-public	n/a
[95] Architecture: cond. steerable GAN	MRA from T1 & T2w MRI axial slices	IXI Dataset	n/a

we shortlisted 63 papers for review based on the innovation in key aspects of the GAN architecture. Of these 63, 28 are proposed in synthetic applications. However, the application fields are quite diverse ranging from segmentation, reconstruction all the way to de-noising - showing possible applications of GANs across many medical tasks.

5.2. Benefits of GANs in the medical field

Deep generative models based on GANs capable of producing realistic looking images, provide major advantages over the more established discriminative frameworks in two challenges that are unique to medical settings:

Scarcity of annotations: Often times, annotations are expensive and hard to come-by in medical imaging. Supervised learning based deep neural networks for such problems is challenging - leading to the possibility of deploying semi- or un-supervised learning. GANs can benefit both of these upcoming frameworks, as demonstrated by multiple studies in synthesis and transformation (sec. 4.6).

Unpaired data: The idea of multi-modal image fusion for better diagnostic decision making is very well grounded. However, finding properly registered data (pixel-wise or area-wise) is supremely challenging. The ability of modern

GAN frameworks e.g. cycle GAN to learn distinctive patterns from unpaired training images and generating realistic outputs is certainly inspiring. The reconstruction quality of GANs can be considered as a significant benefit in its own right - which might pull out the medical image reconstruction quality from blurring effect.

5.3. Drawbacks

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

Trustability of Synthesized Data: In healthcare, where trustability of the clinicians is the biggest challenge for any technology, images synthesized by GANs provide little comfort. The basic networks - generator and discriminator - are still deep neural networks, the mechanism of which is not well studied. In computer vision, where the overall perception is the main concern, these results are adequate. In medical images, however, intensities are typically associated with some meanings e.g. tissue types can be broadly categorized based on HU of CT data. Such an association and mapping is currently missing from the GAN reconstruction - a shortcoming severe enough for clinicians to distrust images synthesized by GAN.

Unstable Training: The typical GANs training is unstable because of numerical reasons pointed out in learning literature [3]. This results in situations such as mode collapse. State-of-the-art learning theory focuses on solving such numerical problems in GANs training for real images. However, in medical imaging, where the modes of images are unclear, how to identify such a problem is unclear. This leads to the question of what sort of numerical singularities might arise in medical imaging and how to address those.

Evaluation Metric: This is a problem, in tandem with the general computer vision community. The best possible way to evaluate reconstruction result is still unclear. In medical imaging, researchers rely mostly on traditional metrics such as PSNR or MSE to evaluate GAN reconstruction quality. This is a tricky situation in the sense that the disadvantages of such metrics were the main reason to move toward GAN. So how can we evaluate potentially better results with metrics which are not capable to understand it?

5.4. Future Works

We believe GANs need to address the significant drawbacks discussed in section 5.3 before being a technology that is trusted in healthcare. 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.

The training instability issue needs to be addressed as well - which means rigorous experimentation to understand the convergence and saddle points of GAN in the medical imaging context. The question regarding metric is far trickier, going about with understanding the performance of GANs synthesized images in CAD by clinicians is a necessary first step.

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 successful GANs deployment in the real clinical scenario.

References

- [1] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, C. I. Sánchez, A survey on deep learning in medical image analysis, *Medical image analysis* 42 (2017) 60–88.
- [2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: *Advances in neural information processing systems*, 2014, pp. 2672–2680.
- [3] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A. A. Bharath, Generative adversarial networks: An overview, *IEEE Signal Processing Magazine* 35 (1) (2018) 53–65.
- [4] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017) 5967–5976.
- [5] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. J. Goodfellow, R. Fergus, Intriguing properties of neural networks, *CoRR* abs/1312.6199.
- [6] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, *CoRR* abs/1511.06434.
- [7] M. Mirza, S. Osindero, Conditional generative adversarial nets, *CoRR* abs/1411.1784.
- [8] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [9] C. Li, M. Wand, Precomputed real-time texture synthesis with markovian generative adversarial networks, in: *European Conference on Computer Vision*, Springer, 2016, pp. 702–716.
- [10] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, *2017 IEEE International Conference on Computer Vision (ICCV)* (2017) 2242–2251.
- [11] A. Odena, C. Olah, J. Shlens, Conditional image synthesis with auxiliary classifier gans, in: *ICML*, 2017.

- [12] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan, CoRR abs/1701.07875.
- [13] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, S. P. Smolley, Least squares generative adversarial networks, 2017 IEEE International Conference on Computer Vision (ICCV) (2017) 2813–2821.
- [14] J. M. Wolterink, T. Leiner, M. A. Viergever, I. Išgum, Generative adversarial networks for noise reduction in low-dose ct, IEEE transactions on medical imaging 36 (12) (2017) 2536–2545.
- [15] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, G. Wang, Low dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss, IEEE Transactions on Medical Imaging.
- [16] X. Yi, P. Babyn, Sharpness-aware low-dose ct denoising using conditional generative adversarial network, Journal of Digital Imaging (2018) 1–15.
- [17] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, CoRR abs/1409.1556.
- [18] S. Yu, H. Dong, G. Yang, G. G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. R. Arridge, J. Keegan, D. N. Firmin, Y. Guo, Deep de-aliasing for fast compressive sensing mri, CoRR abs/1705.07137.
- [19] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, et al., Dagan: Deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction, IEEE Transactions on Medical Imaging.
- [20] M. Seitzer, G. Yang, J. Schlemper, O. Oktay, T. Würfl, V. Christlein, T. Wong, R. Mohiaddin, D. N. Firmin, J. Keegan, D. Rueckert, A. Maier, Adversarial and perceptual refinement for compressed sensing mri reconstruction, CoRR abs/1806.11216.
- [21] T. Salimans, I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, in: NIPS, 2016.
- [22] T. M. Quan, T. Nguyen-Duc, W.-K. Jeong, Compressed sensing mri reconstruction with cyclic loss in generative adversarial networks, CoRR abs/1709.00753.
- [23] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, M. T. Alley, N. Thakur, S. Han, W. J. Dally, J. M. Pauly, L. Xing, Deep generative adversarial networks for compressed sensing automates mri, CoRR abs/1706.00051.
- [24] O. Shitrit, T. Riklin-Raviv, Accelerated magnetic resonance imaging by adversarial neural network, in: DLMIA/ML-CDS@MICCAI, 2017.

- [25] Z. Li, Y. Wang, J. Yu, Reconstruction of thin-slice medical images using generative adversarial network, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2017, pp. 325–333.
- [26] I. Sánchez, V. Vilaplana, Brain mri super-resolution using 3d generative adversarial networks, 2018.
- [27] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017) 105–114.
- [28] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 1874–1883.
- [29] A. P. Aitken, C. Ledig, L. Theis, J. Caballero, Z. Wang, W. Shi, Checkerboard artifact free sub-pixel convolution: A note on sub-pixel convolution, resize convolution and convolution resize, CoRR abs/1707.02937.
- [30] M. Mathieu, C. Couprie, Y. LeCun, Deep multi-scale video prediction beyond mean square error, CoRR abs/1511.05440.
- [31] Y. Chen, F. Shi, A. G. Christodoulou, Z. Zhou, Y. Xie, D. Li, Efficient and accurate mri super-resolution using a generative adversarial network and 3d multi-level densely connected network, CoRR abs/1803.01417.
- [32] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (2018) 834–848.
- [33] D. Ravì, A. B. Szczotka, D. I. Shakir, S. P. Pereira, T. Vercauteren, Adversarial training with cycle consistency for unsupervised super-resolution in endomicroscopy, 2018.
- [34] D. Mahapatra, B. Bozorgtabar, S. Hewavitharanage, R. Garnavi, Image super resolution using generative adversarial networks and local saliency maps for retinal image analysis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 382–390.
- [35] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network., in: CVPR, Vol. 2, 2017, p. 4.

- [36] B. André, T. Vercauteren, A. M. Buchner, M. B. Wallace, N. Ayache, A smart atlas for endomicroscopy using automated video retrieval, *Medical image analysis* 15 (4) (2011) 460–476.
- [37] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (2018) 834–848.
- [38] A. Tack, A. Mukhopadhyay, S. Zachow, Knee menisci segmentation using convolutional neural networks: data from the osteoarthritis initiative., *Osteoarthritis and cartilage* 26 5 (2018) 680–688.
- [39] Y. Xue, T. Xu, H. Zhang, L. R. Long, X. Huang, Segan: Adversarial network with multi-scale l1 loss for medical image segmentation, *Neuroinformatics* (2018) 1–10.
- [40] M. Rezaei, K. Harmuth, W. Gierke, T. Kellermeier, M. Fischer, H. Yang, C. Meinel, A conditional adversarial network for semantic segmentation of brain tumor, in: *BrainLes@MICCAI*, 2017.
- [41] I. J. Goodfellow, Nips 2016 tutorial: Generative adversarial networks, *CoRR* abs/1701.00160.
- [42] P. Moeskops, M. Veta, M. W. Lafarge, K. A. Eppenhof, J. P. Pluim, Adversarial training and dilated convolutions for brain mri segmentation, in: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2017, pp. 56–64.
- [43] Z. Li, Y. Wang, J. Yu, Brain tumor segmentation using an adversarial network, in: *International MICCAI Brainlesion Workshop*, Springer, 2017, pp. 123–132.
- [44] K. Kamnitsas, C. Baumgartner, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, A. Nori, A. Criminisi, D. Rueckert, et al., Unsupervised domain adaptation in brain lesion segmentation with adversarial networks, in: *International Conference on Information Processing in Medical Imaging*, Springer, 2017, pp. 597–609.
- [45] W. Dai, J. Doyle, X. Liang, H. Zhang, N. Dong, Y. Li, E. P. Xing, Scan: Structure correcting adversarial network for organ segmentation in chest x-rays, 2017.
- [46] J. Son, S. J. Park, K.-H. Jung, Retinal vessel segmentation in fundoscopic images with generative adversarial networks, *CoRR* abs/1706.09318.
- [47] A. Lahiri, K. Ayush, P. K. Biswas, P. Mitra, Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale microscopy 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.

- [48] S. M. Shankaranarayana, K. Ram, K. Mitra, M. Sivaprakasam, Joint optic disc and cup segmentation using fully convolutional and adversarial networks, in: *Fetal, Infant and Ophthalmic Medical Image Analysis*, Springer, 2017, pp. 168–176.
- [49] Y. Huo, Z. Xu, S. Bao, C. Bermudez, A. J. Plassard, J. Liu, Y. Yao, A. Assad, R. G. Abramson, B. A. Landman, Splenomegaly segmentation using global convolutional kernels and conditional generative adversarial networks, in: *Medical Imaging 2018: Image Processing*, Vol. 10574, International Society for Optics and Photonics, 2018, p. 1057409.
- [50] D. Yang, D. Xu, S. K. Zhou, B. Georgescu, M. Chen, S. Grbic, D. Metaxas, D. Comaniciu, Automatic liver segmentation using an adversarial image-to-image network, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 507–515.
- [51] B. Kim, J. C. Ye, Cycle-consistent adversarial network with polyphase u-nets for liver lesion segmentation.
- [52] S. K. Sadanandan, J. Karlsson, C. Wählby, Spheroid segmentation using multiscale deep adversarial networks, 2017 IEEE International Conference on Computer Vision Workshops (ICCVW) (2017) 36–41.
- [53] A. Arbellé, T. Riklin-Raviv, Microscopy cell segmentation via adversarial neural networks, 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (2018) 645–648.
- [54] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D. P. Hughes, D. Z. Chen, Deep adversarial networks for biomedical image segmentation utilizing unannotated images, in: *MICCAI*, 2017.
- [55] H. Chen, X. Qi, L. Yu, P.-A. Heng, Dcan: Deep contour-aware networks for accurate gland segmentation, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 2487–2496.
- [56] S. Dong, G. Luo, K. Wang, S. Cao, A. Mercado, O. Shmuelovich, H. Zhang, S. Li, Voxelatlasgan: 3d left ventricle segmentation on echocardiography with atlas guided generation and voxel-to-voxel discrimination, *CoRR* abs/1806.03619.
- [57] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, 2016 Fourth International Conference on 3D Vision (3DV) (2016) 565–571.
- [58] A. Sekuboyina, M. Rempfler, J. Kukacka, G. Tetteh, A. Valentinitzsch, J. S. Kirschke, B. H. Menze, Btrfly net: Vertebrae labelling with energy-based adversarial learning of local spine prior, *CoRR* abs/1804.01307.

- [59] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, Ssd: Single shot multibox detector, in: European conference on computer vision, Springer, 2016, pp. 21–37.
- [60] B. Glocker, D. Zikic, E. Konukoglu, D. R. Haynor, A. Criminisi, Vertebrae localization in pathological spine ct via dense classification from sparse annotations, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, pp. 262–270.
- [61] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, G. Langs, Un-supervised anomaly detection with generative adversarial networks to guide marker discovery, in: International Conference on Information Processing in Medical Imaging, Springer, 2017, pp. 146–157.
- [62] X. Chen, E. Konukoglu, Unsupervised detection of lesions in brain mri using constrained adversarial auto-encoders, CoRR abs/1806.04972.
- [63] C. Baur, B. Wiestler, S. Albarqouni, N. Navab, Deep autoencoding models for unsupervised anomaly segmentation in brain mr images, arXiv preprint arXiv:1804.04488.
- [64] R. Shwartz-Ziv, N. Tishby, Opening the black box of deep neural networks via information, CoRR abs/1703.00810.
- [65] C. F. Baumgartner, L. M. Koch, K. C. Tezcan, J. X. Ang, E. Konukoglu, Visual feature attribution using wasserstein gans, CoRR abs/1711.08998.
- [66] S. Kohl, D. Bonekamp, H.-P. Schlemmer, K. Yaqubi, M. Hohenfellner, B. Hadaschik, J.-P. Radtke, K. H. Maier-Hein, Adversarial networks for the detection of aggressive prostate cancer, CoRR abs/1702.08014.
- [67] A. Udrea, G. D. Mitra, Generative adversarial neural networks for pigmented and non-pigmented skin lesions detection in clinical images, in: Control Systems and Computer Science (CSCS), 2017 21st International Conference on, IEEE, 2017, pp. 364–368.
- [68] A. Tuysuzoglu, J. Tan, K. Eissa, A. P. Kiraly, M. Diallo, A. Kamen, Deep adversarial context-aware landmark detection for ultrasound imaging, CoRR abs/1805.10737.
- [69] L. Zhang, A. Gooya, A. F. Frangi, Semi-supervised assessment of incomplete lv coverage in cardiac mri using generative adversarial nets, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2017, pp. 61–68.
- [70] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, Synthetic data augmentation using gan for improved liver lesion classification, 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (2018) 289–293.

- [71] M. J. M. Chuquicuma, S. Hussein, J. R. Burt, U. Bagci, 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) 240–244.
- [72] A. Kitchen, J. Seah, Deep generative adversarial neural networks for realistic prostate lesion mri synthesis, CoRR abs/1708.00129.
- [73] C. Bermudez, A. J. Plassard, L. T. Davis, A. T. Newton, S. M. Resnick, B. A. Landman, Learning implicit brain mri manifolds with deep learning, arXiv preprint arXiv:1801.01847.
- [74] C. Baur, S. Albarqouni, N. Navab, Melanogans: High resolution skin lesion synthesis with gans, CoRR abs/1804.04338.
- [75] C. Baur, S. Albarqouni, N. Navab, Generating highly realistic images of skin lesions with gans, arXiv preprint arXiv:1809.01410.
- [76] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, D. Shen, Medical image synthesis with context-aware generative adversarial networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 417–425.
- [77] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, I. Išgum, Deep mr to ct synthesis using unpaired data, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2017, pp. 14–23.
- [78] A. Chatsias, T. Joyce, R. Dharmakumar, S. A. Tsiftaris, Adversarial image synthesis for unpaired multi-modal cardiac data, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2017, pp. 3–13.
- [79] P. Costa, A. Galdran, M. I. Meyer, M. D. Abràmoff, M. Niemeijer, A. M. Mendona, A. Campilho, Towards adversarial retinal image synthesis, CoRR abs/1701.08974.
- [80] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abràmoff, A. M. Mendonça, A. Campilho, End-to-end adversarial retinal image synthesis, IEEE transactions on medical imaging.
- [81] J. T. Guibas, T. S. Virdi, P. S. Li, Synthetic medical images from dual generative adversarial networks, CoRR abs/1709.01872.
- [82] H. Zhao, H. Li, L. Cheng, Synthesizing filamentary structured images with gans, CoRR abs/1706.02185.
- [83] A. Ben-Cohen, E. Klang, S. P. Raskin, M. M. Amitai, H. Greenspan, Virtual pet images from ct data using deep convolutional networks: Initial results, in: International Workshop on Simulation and Synthesis in Medical Imaging, Springer, 2017, pp. 49–57.

- [84] L. Bi, J. Kim, A. Kumar, D. Feng, M. Fulham, Synthesis of positron emission tomography (pet) images via multi-channel generative adversarial networks (gans), in: *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, Springer, 2017, pp. 43–51.
- [85] W. Wei, E. Poirion, B. Bodini, S. Durrleman, N. Ayache, B. Stankoff, O. Colliot, Learning myelin content in multiple sclerosis from multimodal mri through adversarial training, *CoRR* abs/1804.08039.
- [86] Y. Hu, E. Gibson, L.-L. Lee, W. Xie, D. C. Barratt, T. Vercauteren, J. A. Noble, Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks, in: *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, Springer, 2017, pp. 105–115.
- [87] E. Gibson, W. Li, C. H. Sudre, L. Fidon, D. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, T. Whyntie, P. Nachev, D. C. Barratt, S. Ourselin, M. J. Cardoso, T. Vercauteren, Niftynet: a deep-learning platform for medical imaging, in: *Computer Methods and Programs in Biomedicine*, 2018.
- [88] F. Tom, D. Sheet, Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning, 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (2018) 1174–1177.
- [89] H. Cho, S. Lim, G. Choi, H. Min, Neural stain-style transfer learning using gan for histopathological images, *CoRR* abs/1710.08543.
- [90] N. Bayramoglu, M. Kaakinen, L. Eklund, J. Heikkilä, 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.
- [91] A. Bentaieb, G. Hamarneh, Adversarial stain transfer for histopathology image analysis, *IEEE Transactions on Medical Imaging* 37 (3) (2018) 792–802.
- [92] M. T. Shaban, C. Baur, N. Navab, S. Albarqouni, Staingan: Stain style transfer for digital histological images, *CoRR* abs/1804.01601.
- [93] L. Han, Z. Yin, Transferring microscopy image modalities with conditional generative adversarial networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 99–107.
- [94] J. M. Wolterink, T. Leiner, I. Isgum, Blood vessel geometry synthesis using generative adversarial networks, *arXiv preprint arXiv:1804.04381*.

- [95] S. Olut, Y. H. Sahin, U. Demir, G. Unal, Generative adversarial training for mra image synthesis using multi-contrast mri, arXiv preprint arXiv:1804.04366.
- [96] D. Jin, Z. Xu, Y. Tang, A. P. Harrison, D. J. Mollura, Ct-realistic lung nodule simulation from 3d conditional generative adversarial networks for robust lung segmentation, arXiv preprint arXiv:1806.04051.