



Automation, Surgery Support and Intuitive 3D visualization  
to optimize workflow in image guided therapy SysTems

## DELIVERABLE D2.2

State-of-the-art report on synthetic image generation



Project number:	ITEA 20044
Document version no.:	v 1.1
Edited by:	Anders Eklund
Date:	2022.11.07

**ITEA Roadmap challenge:**  
Smart Health

This document and the information contained are the property of the ASSIST Consortium and shall not be copied in any form or disclosed to any party outside the Consortium without the written permission of the Project Coordination Committee, as regulated by the ASSIST Consortium Agreement and the ITEA3 Articles of Association and Internal Regulations.



## HISTORY

Document version #	Date	Remarks
V0.1	2022.09.21	Compilation of first input by partners
V0.2	2022.10.24	Reviewed by Quantib
V0.3	2022.11.04	Updated by LiU

### Deliverable review procedure:

- **2 weeks before due date:** deliverable owner sends deliverable –approved by WP leader– to Project Manager
- **Upfront** PM assigns a co-reviewer from the PMT group to cross check the deliverable
- **1 week before due date:** co-reviewer provides input to deliverable owner
- **Due date:** deliverable owner sends the final version of the deliverable to PM and co-reviewer



## TABLE OF CONTENTS

<b>1</b>	<b>ABBREVIATIONS.....</b>	<b>5</b>
<b>2</b>	<b>EXECUTIVE SUMMARY.....</b>	<b>6</b>
<b>3</b>	<b>INTRODUCTION TO SYNTHETIC IMAGES .....</b>	<b>7</b>
<b>4</b>	<b>GENERATING IMAGES THROUGH DEEP LEARNING.....</b>	<b>9</b>
<b>4.1</b>	<b>GANs .....</b>	<b>9</b>
4.1.1	Noise-to-image GANs (creating completely new images) .....	10
4.1.2	Image-to-image GANs (image translation) .....	13
4.1.3	3D aware generation of 2D images.....	14
4.1.4	3D GANs .....	15
<b>4.2</b>	<b>Ensembles of GANs .....</b>	<b>15</b>
<b>4.3</b>	<b>Diffusion Models.....</b>	<b>16</b>
<b>4.4</b>	<b>Training networks with synthetic images .....</b>	<b>17</b>
4.4.1	GANs for data augmentation.....	17
4.4.2	GANs for domain adaptation.....	18
4.4.3	Other applications.....	18
<b>5</b>	<b>SYNTHETIC IMAGES IN DIFFERENT USE CASES .....</b>	<b>19</b>
<b>5.1</b>	<b>Brain tumour segmentation .....</b>	<b>19</b>
<b>5.2</b>	<b>Liver + pancreatic tumours.....</b>	<b>22</b>
<b>5.3</b>	<b>Intracranial hemorrhage.....</b>	<b>26</b>
<b>5.4</b>	<b>Image guided lung interventions.....</b>	<b>28</b>
5.4.1	Generation of Synthetic 3D CBCTs based on Clinical CBCT Data.....	29
<b>6</b>	<b>QUALITY METRICS FOR SYNTHETIC MEDICAL IMAGES .....</b>	<b>31</b>
<b>6.1</b>	<b>Quality of synthetic images from a mathematical perspective .....</b>	<b>31</b>
<b>6.2</b>	<b>Quality of synthetic images for treatment planning.....</b>	<b>32</b>
<b>7</b>	<b>LEGAL ASPECTS OF SYNTHETIC IMAGES.....</b>	<b>33</b>
<b>8</b>	<b>CONCLUSION .....</b>	<b>35</b>



---

<b>9</b>	<b>REFERENCES.....</b>	<b>36</b>
----------	------------------------	-----------



## 1 Abbreviations

AI	Artificial intelligence
CNN	Convolutional neural network
CNNs	Convolutional neural networks
CBCT	Cone beam computed tomography
CT	Computed tomography
DICOM	Digital imaging and communications in medicine
FID	Frechet inception distance
GAN	Generative adversarial network
GDPR	General data protection regulation
ICH	Intracranial hemorrhage
IS	Inception score
MR	Magnetic resonance
MRI	Magnetic resonance imaging



## 2 Executive summary

This document presents state-of-the-art methods for synthetic image generation, focused on medical images. The document covers using GANs and diffusion models, as well as generating images using physics simulations. The document also covers different metrics that can be used for evaluating how realistic the synthetic images are, and briefly mentions the legal aspects related to the generation, and sharing of synthetic data. Finally, the document discusses how synthetic images can be used for different use cases (brain tumours, liver and pancreas tumours, intracranial hemorrhage, image guided lung interventions).



### 3 Introduction to synthetic images

Access to large (annotated) medical image datasets is a major hurdle for training deep learning models in medical imaging. There is not a lack of data per se, but regulations such as GDPR prevents researchers or medical device companies from obtaining large number of images from clinical care providers. This contrasts with the computer vision field, where the large open ImageNet database (Deng et al., 2009) with several million images has been extremely important for developing new deep learning-based methods. There are a number of openly available medical imaging datasets, but they are much smaller compared to ImageNet (for example, the Human Connectome project (HCP) (Van essen et al., 2013) has 1,100 subjects, OpenNeuro (Poldrack et al., 2013) has about 25,000, UK biobank will scan 100,000). In addition to their limited size, openly available medical image datasets are often anonymized through defacing, can represent selective populations around universities, and are often curated before distribution to eliminate bad quality data. Furthermore, the datasets often focus on healthy controls, and it is more difficult to find open datasets with a large number of subjects (> 1000) with a specific disease. For the development of segmentation algorithms, it is difficult to find images with high quality annotations. This limits the potential applicability of any model trained on such data in clinical settings.

Through the introduction of AI models such as generative adversarial networks (GANs) (Goodfellow, 2014), computers can nowadays generate very realistic synthetic images (see for example <https://thispersondoesnotexist.com>) by learning the high-dimensional distribution of real images. The first synthetic images had very low resolution (e.g. 32 x 32 or 64 x 64 pixels), while the most recent networks can generate high quality images of one or several megapixels. Generating synthetic medical images also has the potential to solve the data availability issue for medical imaging. However, given that the training of GANs requires large datasets, this creates a catch-22 situation. This could potentially be solved by starting from a pre-trained GAN. Another possible solution is to train GANs, or other models, in a federated setting with images stored at each hospital.

There are many possible applications of synthetic images in medical imaging. Synthetic images can make it easier to share medical data, as the synthetic images do not belong to a specific person. Another application is advanced data augmentation, allowing the creation of new images from the same distribution, as opposed to only modifying existing images through conventional augmentation strategies (e.g., rotation, translation). Finally, different types of medical images can be synthesized from one another, such as synthesizing a CT image from an MR image. Because this solely requires the collection of an MR image, this saves considerable time in the clinical workflow (5 – 20 minutes), and avoids the exposure of radiation to the patient.





## 4 Generating images through deep learning

This section will first introduce generative adversarial networks (GANs) (Goodfellow et al., 2014), and then introduce diffusion models. Both have been successful in generating realistic synthetic images.

### 4.1 GANs

GANs recently gained a lot of attention in the computer vision community. Generative modeling is an unsupervised machine learning task that involves discovering and learning the patterns automatically in the input data and then trains itself to further generate new examples. Generative modeling can further be subdivided into two different models, a generator, and a discriminator. Both models are trained together and the clever competition between the generator and the discriminator provides a way of incorporating unlabeled samples into training and imposing higher-order consistency. Figure 1 shows the basic architecture of GAN model. The combined loss function of both generator (G) and the discriminator (D) can be seen in Equation 1, which shows the adversarial loss, where we can see that generator works to minimize the loss and oppositely the discriminator works to maximize the loss. Only a proper balance between the generator and discriminator results in high quality output. I.e., the generator should generate examples for which the discriminator has a hard time determining whether it is a ground truth example, or a synthetic example. Synthetic labels can be created by providing them alongside the medical images themselves, in a separate input channel.

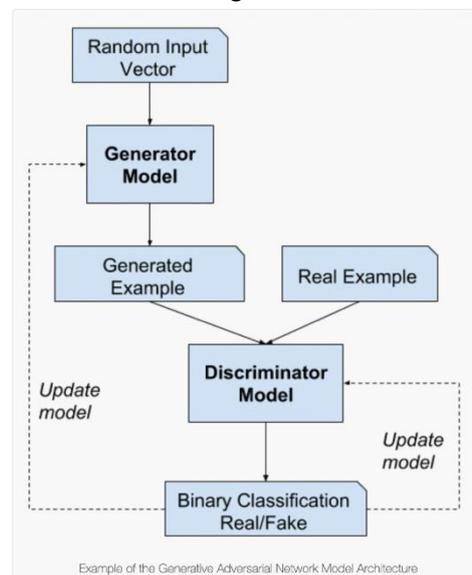


Figure 1: Basic architecture of a GAN (generative adversarial network). A GAN has a generator and a discriminator that respectively compete to generate better synthetic images, and to be better at distinguishing synthetic images from real images.

$$L = \min_G \max_D [\log(D(x)) + \log(1 - D(G(z)))]$$

Equation 1: Adversarial loss function, computed through minimization of the generator loss, and maximization of the discriminator loss.



GANs have proven themselves to be useful in scenarios such as data augmentation (Cha et al., 2019), image-to-image translation (Kang et al., 2021), domain adaptation, and filling in the missing data in many cases. Recently the success of GANs in natural imaging has gained the attention of researchers in the medical imaging community. GANs provide the opportunity to resolve the issue of requiring large annotated medical datasets by producing synthetic data which can further be used to train deep models. Many variants of the GAN framework have been proposed in the literature. We therefore take the opportunity to explain some of the commonly used GAN frameworks available in literature below.

#### 4.1.1 Noise-to-image GANs (creating completely new images)

A general problem with GANs is that the training is very unstable, for example the mode collapse problem where the generator only generates a few images that fool the discriminator. Many papers have therefore focused on improving the training stability. We will here briefly present some of the most recent GAN architectures.

##### **Progressive GAN**

Progressive growing of GANs was proposed by Karras et al. (2017) from Nvidia research Lab. The code that was developed was released open source at [https://github.com/tkarras/progressive\\_growing\\_of\\_gans](https://github.com/tkarras/progressive_growing_of_gans). In this framework they propose a training methodology for generative adversarial networks where the key idea is to grow both the generator and the discriminator progressively, starting from a low resolution and enhancing the resolution while the training progresses. This is done by progressively adding new layers to the network as shown in Figure 2. The generator and the discriminator are mirror images of each other and always grow in synchrony from a low resolution (4x4 pixels in the image below), to the final resolution (1024x1024 pixels in the image below). All the existing layers in the model remain trainable while the newly added layers are faded in to avoid a sudden impact on already trained low resolution preceding layers. Progressively increasing the image resolution was noted to improve training stability, preventing the occurrence of mode collapse.

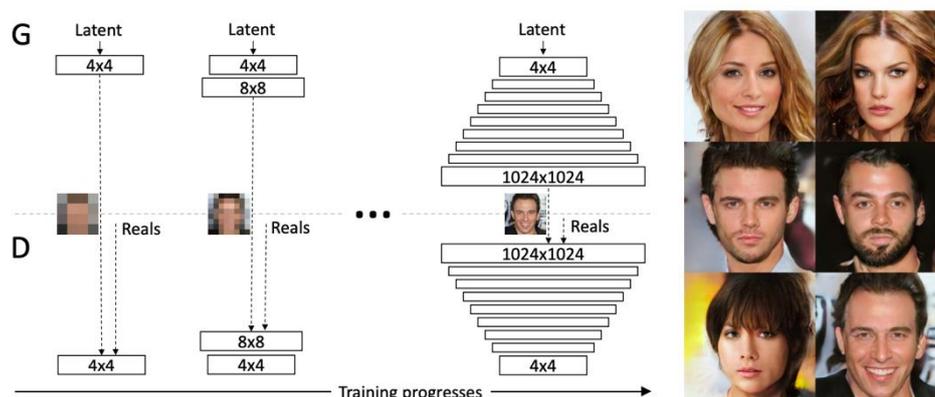


Figure 2: The training set starts with both generator (G) and discriminator (D) with a low spatial resolution of  $4 \times 4$  pixels. With the progression of training, both G and D get an incremental increase in layers providing higher spatial resolution respectively. On the right-hand side, six different images generated using Progressive growing GAN with a spatial resolution of  $1024 \times 1024$  can be seen.



## StyleGAN

A Style-Based Generator was proposed by Karras et al. (2020) where the authors have proposed an alternative generator architecture for generative adversarial networks, borrowing from style transfer literature. The code was released open source at <https://github.com/NVlabs/stylegan>. Using this architecture, the model learns unsupervised separation of high-level attributes and stochastic variation in the generated images and enables intuitive, scale-specific control of the synthesis. Unlike the progressive GAN, the proposed architecture does not take the input latent code using the input layer, but takes the input from a learned constant instead as shown in Figure 3. One of the downsides of StyleGAN is the presence of water droplet-like artifacts in the synthetic images, which triggered development of StyleGAN2.

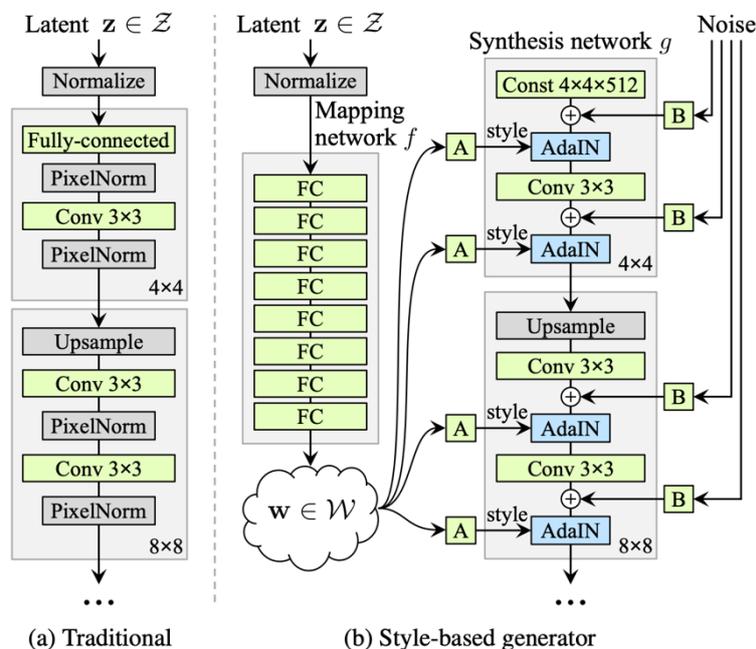


Figure 3: Latent code is fed to an intermediate latent space which then controls the generator by using adaptive instance normalization (AdaIN) at each convolutional layer. Here it can be seen that Gaussian noise is added after each convolution. “A” stands for a learned affine transform, and “B” applies learned per-channel scaling factors to the noise input. There are 8 and 18 layers in mapping network and synthesis network  $g$  respectively. There is a separate  $1 \times 1$  convolution layer at the end to allowing RGB images to be obtained.

## StyleGAN2

Although the aforementioned style-based GAN architecture yields state-of-the-art results, Karras et al. (2020) revisited and analyzed several artifacts and have proposed changes in both the training methods and the architecture of styleGAN as well. Again the code was released open source (<https://github.com/NVlabs/stylegan2>). The authors have redesigned the generator normalization, revisited the progressive growing, and focused on the generator to produce better results by providing better initial conditions. This adds another benefit by making it easier to invert the generator, making it possible for any generated image to reliably attribute to a particular network. The authors



removed the normalization artifacts by pinpointing the problem to AdaIN operation which destroys the information found in the magnitudes of the features relative to each other by computing the mean and variance of individual feature maps separately. They have proved the hypothesis that by removing the normalization step from the generator the droplet affect disappears completely. Figure 4 shows the detailed and the revised architecture of StyleGAN and serves as the starting point of redesigned normalization.

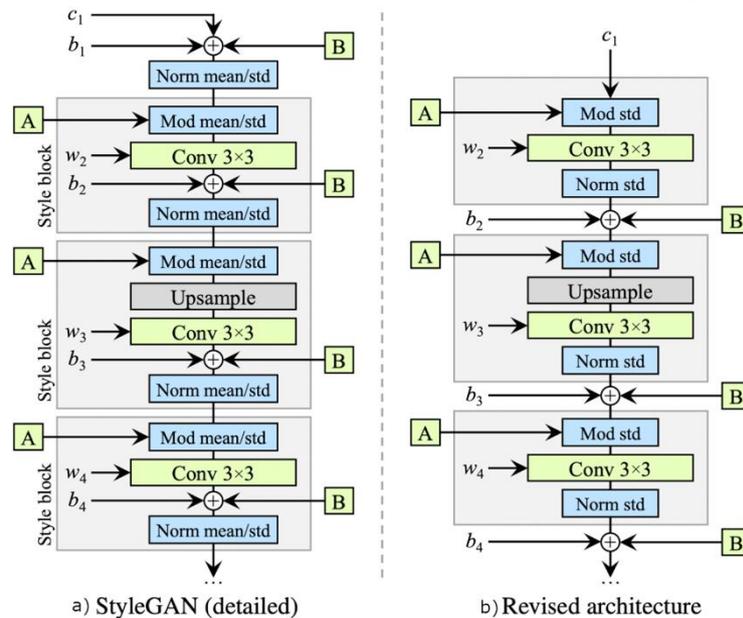


Figure 4: Fig.4 a) shows the detailed architecture of StyleGAN and StyleGAN2 where A denotes a learned affine transform from  $W$  that produces a style and B is a noise broadcast operation. ( $w$ ) are the learning weights, ( $b$ ) represents the bias, and ( $c$ ) represents the constant input. The grey box shows that the style is active per box. Fig.4 b) shows the several changes made to the original architecture where redundant operations at the beginning are removed and the addition of Band ( $b$ ) happens outside the active area of style and adjusts only the standard deviation per feature map.

### StyleGAN3

Karras et. Al (2021) further investigated Stylegan2 and exposed the root cause of careless signal processing which was causing the aliasing problem in the generator network. The leakage of information into the hierarchical synthesis process was eliminated by interpolation of all the signals in the network to continuous. The newly constructed network matches the FID (Fréchet inception distance, see Section 6.1) of StyleGAN2, but the internal representation is dramatically different and is equivariant to rotation and translation even at subpixel scales. This newly developed architecture (Figure 5) is suitable for the creation of both images and videos.

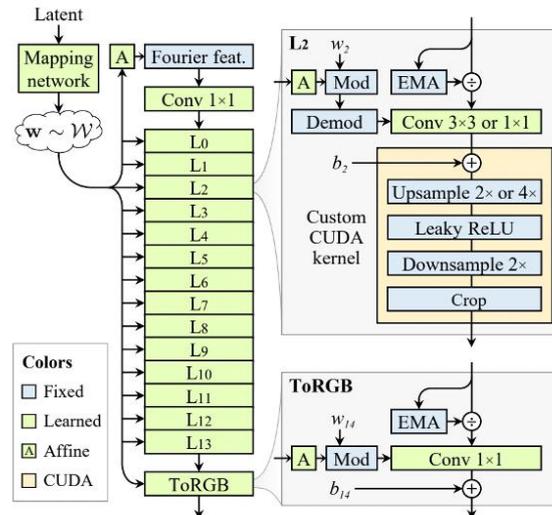


Figure 5: An alias-free generator used in StyleGAN3. The main data path consists of Fourier features and normalization, modulated convolutions, and filtered nonlinearities.

#### 4.1.2 Image-to-image GANs (image translation)

While section 4.1.1 covers GANs that can generate completely new images, this section covers GANs that can instead transform one image type into another (e.g. transforming a CT image into a MR image). Training of image-to-image GANs is typically more stable, as the GAN starts from an image (instead of starting from noise). For 2D images, the most popular architectures for image-to-image GANs are pix2pix (Isola et al. 2017) and CycleGAN (Zhu et al. 2017), both discussed below. While pix2pix requires paired and registered images, CycleGAN works for unpaired images which are not registered.

##### Pix2pix

A conditional GAN (cGAN) is an extension of the GAN architecture described previously where generated images can be controlled with specific conditions. Pix2Pix is an implementation of cGAN. For the training of noise-to-image GANs, the generator model is given input images while the discriminator model is given a random mix between synthetic and real examples to train. On the other hand, Pix2Pix is trained with both the input and the target images. Using this approach, the network learns to map the input to the output images by means of a specific loss function. In traditional GANs the generator usually calculates a loss which depends of the data while cGANs calculates a structured loss that focuses on a possible structure between the target image and the network output image. This makes the output image structurally alike to the target image. The discriminator still works in the traditional way and tries to classify each image as a real or synthetic sample.

##### CycleGAN

The image-to-image translation is done by learning the mapping of input images and translating them to the output image where both input and output images are aligned. However, in many cases, the availability of paired and aligned (registered) images are not available. Therefore Zhu et al. presented an approach capable of learning from the source domain and map to the target domain in the absence of source (X) and target (Y) domains.



The goal is to learn the mapping  $G: X \rightarrow Y$  in a way that the distribution of images among  $G(X)$  and  $Y$  is indistinguishable. This is done by using an adversarial loss. They have coupled it with inverse mapping  $F: Y \rightarrow X$  and introduce a cycle consistency loss to enforce  $F(G(X)) \approx X$  (and vice versa) to cope with the highly constrained mapping. The architecture of CycleGAN can be seen in Figure 6

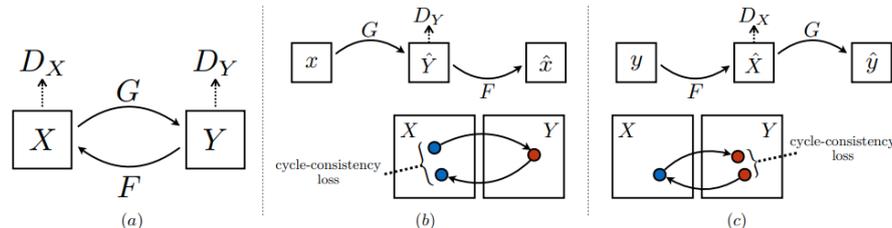


Figure 6: (a) There are two mapping functions  $G: X \rightarrow Y$  and  $F: Y \rightarrow X$  with their associated discriminators  $D_X$  and  $D_Y$ .  $D_X$  and  $D_Y$  encourage  $G$  and  $F$  to produce images to the corresponding domain. Furthermore, two cycle consistency losses have been introduced that work with the intuition that once the generated image is generated back to its original domain it should give back the original image from where it started. (b) forward cycle-consistency loss:  $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ , and (c) backward cycle-consistency loss:  $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$

A general challenge with training image-to-image GANs is that they often require more GPU memory, compared to training noise-to-image GANs, as several generators and discriminators are trained at the same time. This is generally not a big problem in 2D, but in 3D (described below) it becomes challenging to train 3D CycleGAN (Abramian & Eklund, 2019) or Vox2Vox (3D pix2pix) (Cirillo et al., 2020). It may therefore be necessary to instead train the 3D image-to-image GANs using sub-volumes. If the architecture is fully convolutional, it can then be applied to a volume of any size.

#### 4.1.3 3D aware generation of 2D images

The biggest advantage of using 3D models is that they can relate to different angles of the same synthetic object, given that 2D models lack the representation of information from different angles. Until now the best 2D models have attained higher stability far away from natural results but still lack in 3D because they do not store 3D information and are unable to keep the display stable from different viewing angles.

Recently Eric R., et al. (2022) from Nvidia and Stanford have introduced 3D aware GANs for the creation of better synthetic images, in their paper they have shown that 3D GANs can produce even better synthetic images, while also being able to perform reconstructions (i.e. filling in the gaps of missing data). This is demonstrated by a hybrid approach that combines implicit and explicit representation resulting in an efficient approach that scales effecticely with increasing resolutions. They rely on a three-plane 3D representation rather than a full voxel grid which is connected before a StyleGAN2 generator mesh and finally the output of the generator is stored (Figure 7). AI neural renderer further decodes the saved information and sends it to a super-resolution module which scales the 128 by 128 pixels up to 512 by 512 pixels. Their approach not only produces state of the art qualitative and quantitative results for view-consistent 3D-aware image synthesis, but at the same time can generates high quality 3D shapes of the synthesized scenes due to its strong 3D-structure-aware inductive bias. One of the prominent achievements is that EG3D can produce a matching 3D reconstruction from a single image resulting in 3D aware represented images.

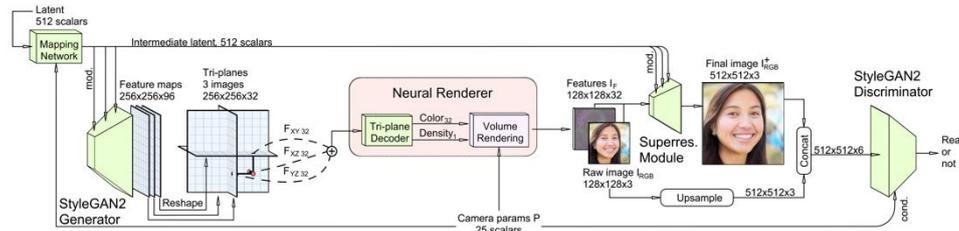


Figure 7: This method is a combination of several parts: a feature generator, a mapping network, a tri-plane 3D representation with lightweight feature decoder, a neural volume renderer, a super resolution module and a pose conditioned Stylegan2 discriminator with dual discrimination. Using neural rendering and decoupling feature generation the architecture allows Stylegan2 to generate 3D scene generalization.

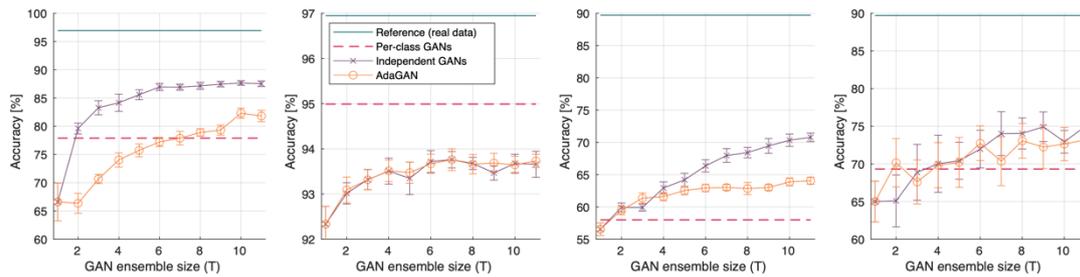
#### 4.1.4 3D GANs

Regarding 3D GANs, required to synthesize volumes, there is much less work compared to the aforementioned work on generating images using 2D GANs. There are several reasons for this; One reason is that 3D GANs require more GPU memory and processing time, for the training, compared to training 2D GANs (which are already computationally challenging). Another reason is that training 3D GANs requires a large number of 3D volumes, and large open 3D (medical) datasets are not as common compared to large open 2D datasets used in computer vision (e.g. ImageNet which contains several million images). Although the UK biobank aims to scan 100,000 subjects, there are very few openly available medical imaging datasets of this size.

Regarding noise-to-image GANs, Wu et al. (2016) used a 3D GAN to generate synthetic objects, but the resolution was limited to  $64 \times 64 \times 64$  voxels, while the major challenge for all GANs is to generate high resolution images or volumes. Kwon et al. (2019) generated brain volumes of  $64 \times 64 \times 64$  voxels, but they appear to lack detail. Jung et al. (2021) used a 2D GAN to generate a sequence of slices and then used a combination of 2D and 3D discriminators to enforce consistency in 3D. However, no 3D volumes are shown in the paper and the obtained resolution is not mentioned. Bu et al. (2021) used a 3D GAN to synthesize lung nodules, but the resulting volumes are limited to  $32 \times 32 \times 32$  voxels by synthesizing the lung nodules as opposed to the full lung volume.

## 4.2 Ensembles of GANs

A major difficulty in generating diverse high-resolution images is to fully capture the complex high-dimensional data distribution. Consider that a one-megapixel image has one million dimensions, while the (non-linear) subspace of real images has a much lower dimension. When training a deep neural network, the weights are typically initialized randomly. This means that training several GANs with the same architecture will, due to the random initialization, capture slightly different parts of the complex high-dimensional data distribution. Eilertsen et al. (2021) demonstrated, Figure 8, that training an ensemble of 2-10 GANs in general leads to better performance when using the synthetic images to train a classifier. The drawback of this approach is of course that the training time increases linearly with the number of GANs trained.



(a) SVHN-II – DCGAN (b) SVHN-II – PG-GAN (c) CIFAR-II – DCGAN (d) CIFAR-II – PG-GAN

Figure 8: Classification performance on synthetic datasets, comparing different ensemble sizes and approaches. Each datapoint has been estimated from the mean of 10 separate trainings, and standard deviations are reported with error bars. Using an ensemble of GANs clearly leads to better results, compared to using a single GAN, but the performance is still considerably worse compared to using real images. Image from Eilertsen et al. (2021).

### 4.3 Diffusion Models

Ho et al. (2020) introduced the diffusion model which is a generative model that can produce similar synthetic data on which it is trained. Diffusion models work in a forward destruction of data by adding noise, and then learns to generate back realistic images by reversing the noise process. Once the network has learned to perform denoising, any random noise image can be converted to a realistic image. Diffusion models were introduced through text-to-image models like DALL-E and DALL-E 2, but have proven themselves generating images from noise.

More specifically, the diffusion model is a latent variable model that uses a fixed Markov chain to map the latent space. Noise is gradually added to the data and once the approximate posterior with the same dimensionality as the input image is achieved then the model learns the reverse process, where an image earlier transformed to a pure Gaussian noise is converted back to new data as seen in Figure 9.

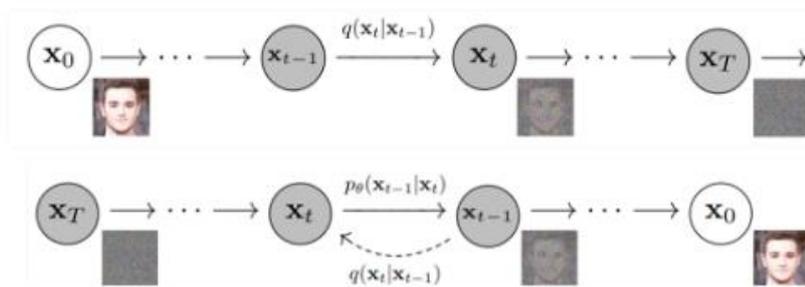


Figure 9: Diffusion Model where  $X_0$  is the original image ( $x_1, \dots, x_T$ ) are the latent variables with the same dimensionality as  $X_0$

Over recent years, research in diffusion models has exploded and is producing state-of-the-art image quality. It is of great benefit that diffusion models do not require adversarial training and have added the benefits of scalability and parallelizability increasing training efficiency. In general, diffusion models are faster to train than GANs, but suffer from longer inference times since generating a synthetic image requires an iterative process.



## 4.4 Training networks with synthetic images

Access to large annotated datasets is one of the major challenges in deep learning, especially for medical applications. Generating synthetic data using GANs, both images and annotations, can be a way to alleviate this data demand. There are several scenarios where GANs can be used to generate helpful training data. The most straightforward of which being training a GAN on the already available dataset, to then generate more samples that can be added to the training set. In this scenario the usage of the GAN can be viewed as an advanced form of data augmentation. Other scenarios where GANs can be useful include domain adaptation, privacy preserving data sharing and inclusion of prior knowledge into the training data.

### 4.4.1 GANs for data augmentation

Data augmentation is a commonly used technique during network training and can be viewed as a way to expand the dataset, e.g. by means of translation, rotation or intensity scaling. A downside of standard data augmentation is that it requires domain knowledge to perform well, to for example guarantee that the augmented images still represent anatomically correct structures. Using GANs for data augmentation removes this requirement and can hence be a good alternative for data augmentation. In short, a GAN is trained on the already existing training set and is then used to generate new training samples, with annotations, that are added to the training dataset. For classification, two examples where this has been done successfully are by Maayan et al. for liver lesion classification and by Li et al. for blood cell image classification. For image segmentation, Neff et al. managed to improve performance for lung segmentation in chest radiographs by using a GAN for data augmentation, see Figure 10. Kossen et al. did the same thing for brain vessel segmentation in TOF-MRA images showing a slight increase in performance when adding synthetic data compared to using conventional augmentation (Dice score from 0.879 to 0.887).

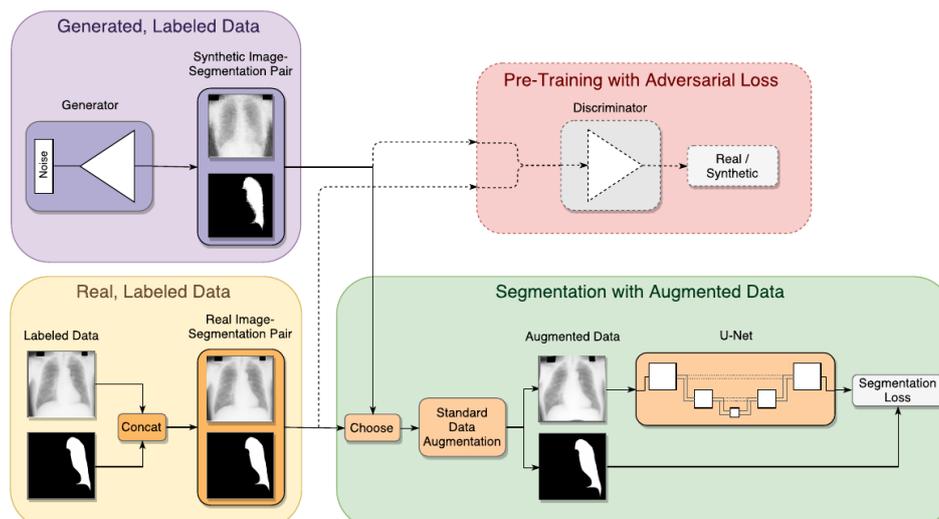


Figure 10: Training setup using GANs for data augmentation from “Generative adversarial networks to synthetically augment data for deep learning-based image segmentation.” by Neff et al.



#### **4.4.2 GANs for domain adaptation**

Domain adaptation aims to train a model that performs well on samples from a target domain that is significantly different from samples from the source domain where labels are available. An example using GANs for domain adaptation was done by Sandfort et al. (2019) where a CycleGAN was used to transform contrast CT images into non-contrast CT images. The synthetic non-contrast CT images were then added to the training set, together with the annotations previously only available for contrast CT images. Training using the extended dataset improved performance for kidney, liver and spleen segmentation in non-contrast CT images. A similar approach was used by Dou et al. but instead of actually creating synthetic images, Dou et al. used an adversarial training setup to align intermediate features during training of a cardiac segmentation network. A discriminator was trained to predict if the features were from a CT or an MR image, the segmentation network was then trained to produce intermediate features that confused the discriminator. This enabled the segmentation network to perform well on CT images although labels were only available on MRI images.

#### **4.4.3 Other applications**

To deal with the difficulty of sharing medical data due to privacy reasons, Yoon et al. (2020) proposed to train a GAN that, given a data sample and a noise vector, produces a new data sample. The GAN is specifically trained to produce samples that are hard to identify to enable sharing of the synthetic dataset while still preserving the privacy of patients in the original dataset. Amirrajab et al. (2020) proposes to use a human phantom and a GAN to generate labelled cardiac MR images. The human phantom captures new anatomical variations not present in the original dataset, and adding the synthetically created samples improved segmentation performance for the trained network.



## 5 Synthetic images in different use cases

There are generally two ways GANs can be used for medical imaging. First is the generative aspect where GANs are used to explore and discover the underlying structure of training data and learns to generate realistic synthetic images. The second focuses on the discriminative aspect where the discriminator can be used to regularize or detect abnormalities in the provided data (Yi et al. 2019) (not further discussed in this document). Figure 11 from Yi et al. provides examples of GAN related applications, with examples (a), (b), (c), (d), (e), (f) that focus on the generative aspect while example (g) exploits the discriminative aspect.

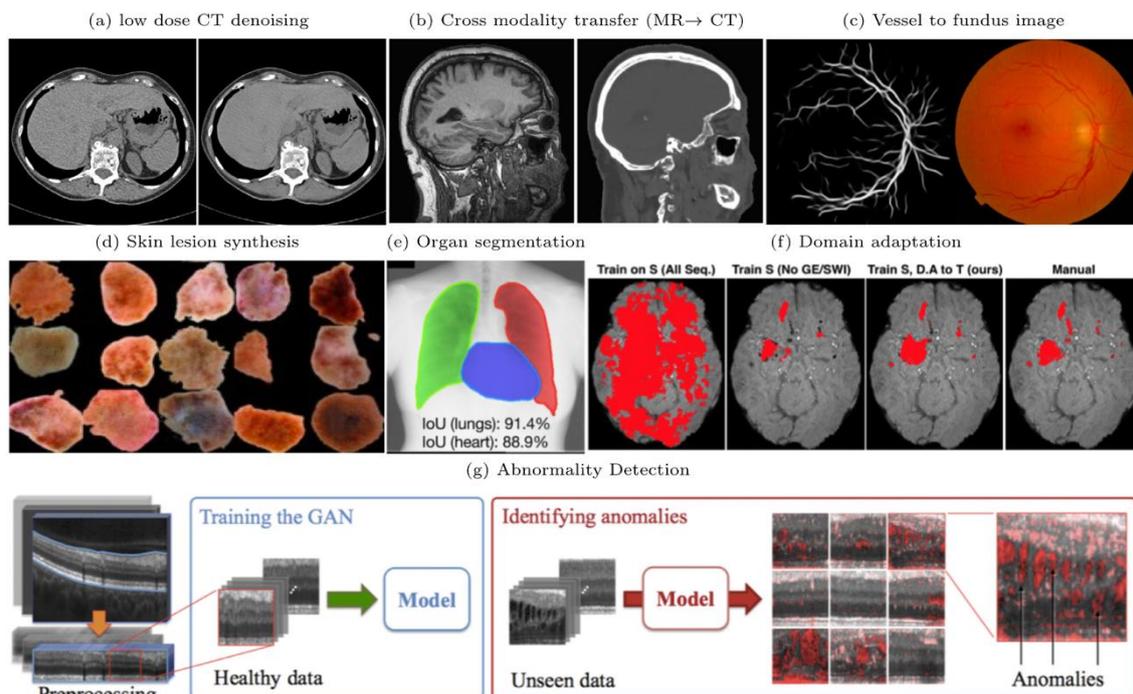


Figure 11: Example applications using GANs. (a) Left side shows the noise contaminated low dose CT and the right side shows the denoised CT that well preserved the low contrast regions in the liver (Yi and Babyn, 2018). (b) Left side shows the MR image and right side shows the synthesized corresponding CT. Bone structures were well delineated in the generated CT image (Wolterink et al., 2017a). (c) The generated retinal fundus image have the exact vessel structures as depicted in the left vessel map (Costa et al., 2017b). (d) Randomly generated skin lesion from random noise (a mixture of malignant and benign) (Yi et al., 2018). (e) An organ (lung and heart) segmentation example on adult chest X-ray. The shapes of lung and heart are regulated by adversarial loss (Dai et al., 2017b). (f) The third column shows the domain adapted brain lesion segmentation result on SWI sequence without training with the corresponding manual annotation (Kamnitsas et al., 2017). (g) Abnormality detection of optical coherence tomography images of the retina (Schlegl et al., 2017).

### 5.1 Brain tumour segmentation

Brain tumours compose about 2% of the cancer incidences, affecting some 300,000 subjects globally each year (Leece et al., 2017), with a low survival rate and a high morbidity for the patients. Though not being the most prevalent cancer type, brain tumours are prone to complicated and challenging treatment procedures that are often a combination of surgery, radiotherapy and chemotherapy, where treatment planning and follow up of the treatment is highly dependent on radiology images. The best treatment for a specific patient depends on the amount of tumours (e.g. one large



tumour, or many small metastases), and their size and location. Furthermore, the size of the tumour is important in calculating the radiotherapy dosage required to kill the cancer cells. MRI is normally used to obtain this information, and to plan the treatment, as MRI provides excellent contrast between soft tissue types, and different MR sequences can be used to expose different information or contrast. A CT scan is normally required to compute the dose fall-off. The CT image can with a GAN (such as CycleGAN) be generated from the MR image, but it will not be further discussed here. It is also necessary to segment important organs-at-risk (e.g. the optic nerve) which should not be damaged by the radiation, see Figure 12. The treatment plan, i.e. how much radiation to apply to different parts of the brain, can be generated manually, through mathematical optimization or through machine learning.

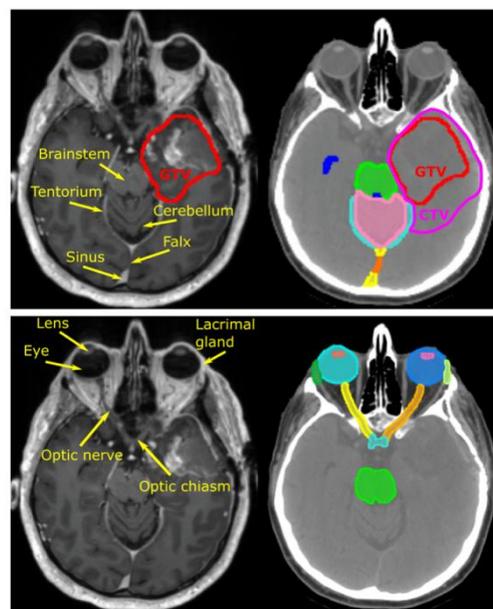


Figure 12: Illustration of brain tumour (red, to be targeted by radiation) and organs-at-risk (yellow, which should receive as little radiation as possible). GTV = gross tumour volume, CTV = clinical target volume (CTV). Deep learning can reduce the treatment planning time substantially, by performing automatic segmentation of tumour(s) and organs-at-risk (instead of the time consuming nature of creating segmentations manually). Image from an open dataset in the cancer imaging archive (see references).

Segmentation of the tumour(s) and organs-at-risk is currently often performed manually or semi-automatically by a neuro-radiologist, medical physicist, radiation oncologist or RTT (radiotherapy technician). Manual segmentations can be very time consuming, typically requiring between 10 – 60 minutes per patient, especially for cases containing many metastases and organs-at-risk. It is therefore desirable to use deep learning for automatic segmentation, but collecting and annotating brain images from a large number of subjects is time consuming and costly. To train image segmentation networks using synthetic images, it is necessary to synthesize images and annotations at the same time. This is more challenging compared to just generating new images. There are basically two approaches that can be used to achieve this. The first approach is to first use a noise-to-image GAN to generate synthetic label images, and to then use an image-to-image GAN to generate MR brain images from the label image. The second approach is to generate the brain image and the label image at the same time, using a single noise-to-image GAN.



The BraTS 2020 dataset, containing data from 369 tumour patients, can be used for training GANs to produce synthetic brain tumour images and annotations. This data has four MR modalities (T1, T1 contrast, T2, T2 FLAIR), and we can consider the ground truth annotation for the tumour to be the fifth modality. The data comes in NIFTI volumes, which for 2D GANs can be further separated into slices. The original size of the extracted slice is 240 x 240 which is zero-padded to 256 x 256. This was done because GANs typically require the input dimensionality of an image to be a power of 2.

Figure 13 shows the first approach mentioned previously, where two GANs together generate the annotations and the corresponding MR image (Foroozandeh & Eklund, 2020). Adding synthetic images gave a small improvement to the segmentation accuracy, while only using synthetic images resulted in much worse performance compared to using real images. Figure 14 shows the second approach (Nijhawan, 2021) where a single GAN is trained to generate MR images and annotations as separate channels in a single image. This approach seems to work better, but a direct comparison is required to draw any conclusion.

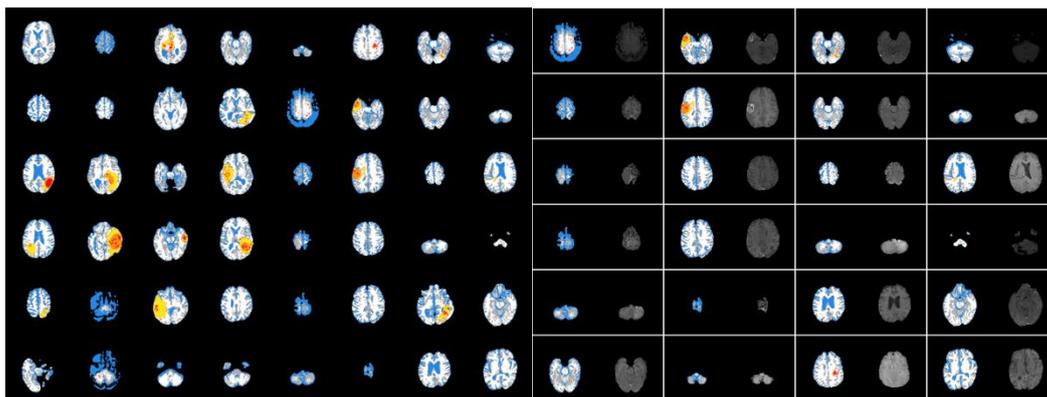


Figure 13. Generating tumour annotations and the corresponding MR image using a two-step procedure. First, synthetic annotations are generated with a noise-to-image GAN (left image). Second, a corresponding MR image is generated using an image-to-image GAN. Note that the annotations cover the full brain (automatically obtained from the function FAST in the FSL software), as it will be too difficult for a GAN to generate a full brain MR image from only the tumour annotation.

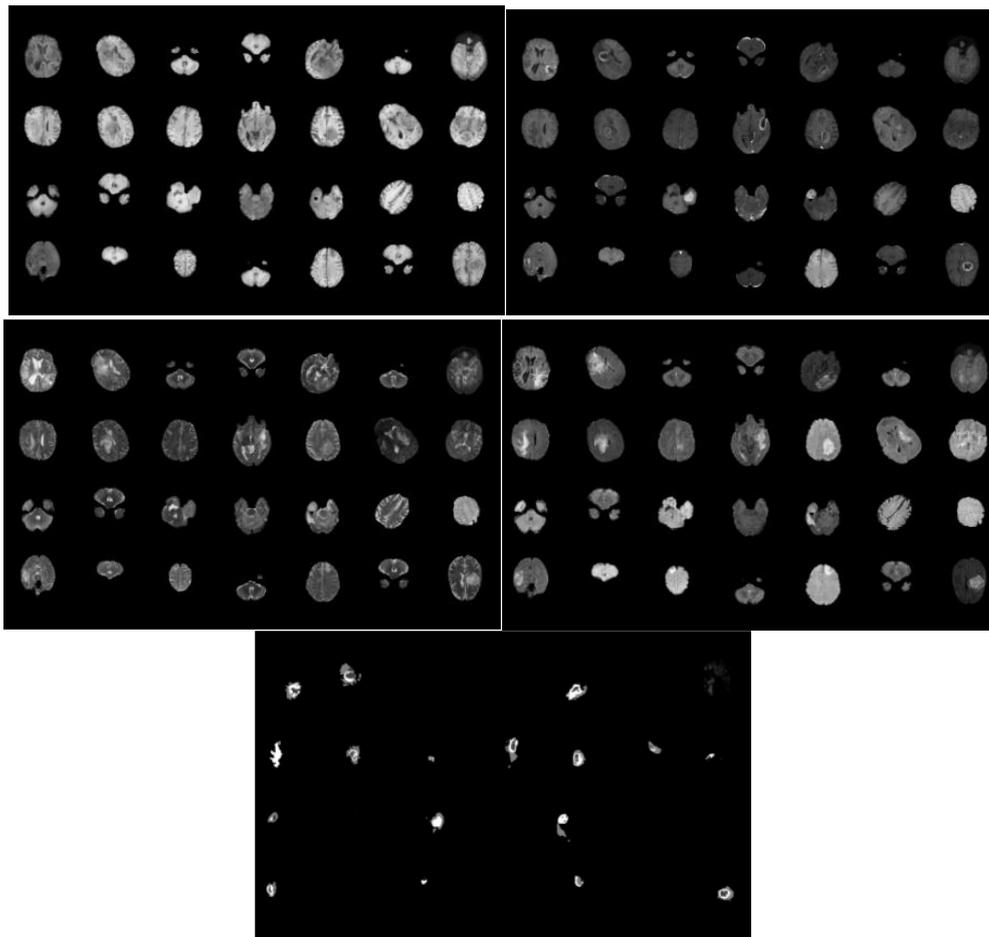


Figure 14. Generating 4 MR images and brain tumour annotations at the same time, using a single noise-to-image GAN to generate 5 channel images.

## 5.2 Liver + pancreatic tumours

Liver and pancreatic cancer are respectively the third and fourth leading causes of cancer death world-wide. Pancreatic cancer is projected to become the second leading cause of cancer death within a decade. Algorithms that facilitate the automatic segmentation of the liver, pancreas and tumour tissues in commonly used imaging modalities, e.g. CT and MRI, are important steps towards the early diagnosis of tumours in the liver and pancreas. Early and accurate detection allows for a complete resection of the tumour while sparing surrounding healthy tissues, resulting in increased chances of survival after the disease, in addition to an improved quality of life. Currently, the best results for liver, pancreas and tumour segmentations are all obtained using deep-learning techniques.

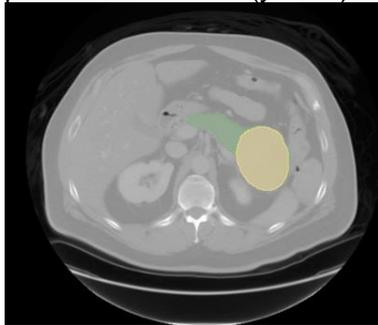
However, the development of high-quality deep-learning algorithms requires large and varied datasets for training and development. Novel algorithm developments are therefore hampered by the lack of available data. It is important to note that neural networks should ideally be trained using multi-center, multi-scanner data, to ensure algorithms work sufficiently well on data from various institutions and scanners. Although medical care providers commonly have lots of clinical data at their disposal, this data is often unlabelled (missing segmentations of the tumour and liver/pancreatic) and are frequently obtained using a single scanner. Additionally, strict



patient privacy regulations often prevent sharing of this data with other clinical care providers or medical device manufacturers.

To assist in the initial training and development of segmentation algorithms, a number of publicly available datasets can be found online. This data is often released for segmentation challenges, that are organized to incentivize researchers and medical device manufacturers to develop and publish high quality segmentation algorithms. Publicly available datasets for the liver and pancreas include:

- The Medical Decathlon dataset (<https://medicaldecathlon.com/>), including 281 3D CT images of the pancreas, accompanied by pancreas (green) and pancreatic tumour (yellow) segmentations.



The Medical Decathlon dataset also includes liver CT data for 131 (different) cases, including liver and liver tumour segmentations.

- The Pancreas-CT dataset of the Cancer Imaging Archive (<https://wiki.cancerimagingarchive.net/>), which includes 80 3D CT scans of the pancreas. Pancreas segmentations are available, but no pancreatic tumours were present in the scanned patients.
- The AMOS 2022 grand challenge dataset (<https://amos22.grand-challenge.org/>), providing 200 CT scans and 40 MRI scans with annotation of 15 different organs (including the liver and pancreas). No associated tumours were segmented or present in the scanned patients.

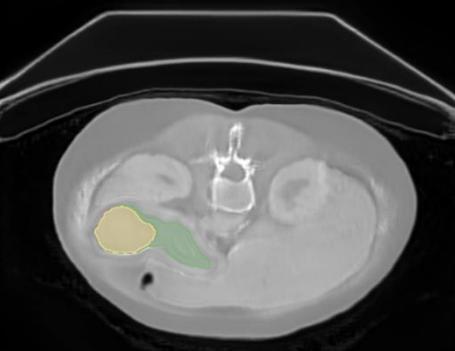
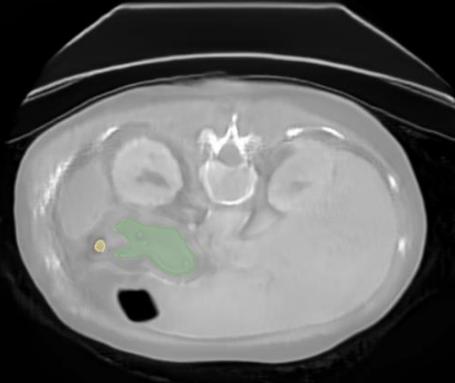
Despite the availability of this data, the total number of scans (~600 of the pancreas, ~370 of the liver) is still small compared to the number of examples that can be found on ImageNet for natural objects. In addition, it should be noted that only the Medical Decathlon dataset provides associated tumour segmentations, thereby limited the amount of training data for liver and pancreatic tumour segmentation.

In an attempt to artificially increase the amount of training data available, we have therefore used these datasets for the training of GANs, allowing the creation of synthetic CT scans and associated labels. Initial experiments were run using 2 of the aforementioned GAN architectures, including 'Progressive Growing of GANs (PGGAN)' and 'StyleGAN2'. Results are summarized in the table below:



Architecture	Synthesized image	Description
PGGAN		The CT image itself looks realistic. The shape of the pancreas segmentation looks correct, but it is situated at an incorrect position. The tumour segmentation is too small, and does not correspond to any hypo-intensities of the synthetic CT scan. Some artifacts, indicated by the red circles, can be noted.
PGGAN		See above, the exact same conclusions can be drawn.
StyleGAN2		Both the synthetic CT scan and synthesized pancreas and pancreatic tumour labels look realistic.



StyleGAN2		Although the tumour is rather large, the synthetic CT scan and pancreas look realistic.
StyleGAN2		Although the synthesized CT scan looks correct, the shape of the pancreas is unrealistic, and the tumour is situated outside of the pancreas.

The examples above illustrate that the PGGAN and StyleGAN2 architectures are capable of producing realistic synthetic CT images. Although mistakes are made occasionally, the pancreas segmentations generally look realistic. The tumour segmentations were harder to synthesize, with sometimes producing rather unrealistic results. The outcome of these experiments is in line with the amount of training data that was used for training of the GAN networks; The input datasets provided many CT slices, fewer but still a decent number of slices containing pancreas segmentations, and far fewer slices with tumour segmentations (only present in a single dataset).

Now that realistic examples can be generated, the next steps include the manual selection of large number of realistic synthetic CT scans and pancreas/pancreatic tumour segmentations. This synthetic data will be included as an additional augmentation strategy to be used in segmentation networks that aim to segment the pancreas and pancreatic tumours from CT scans. The effect of providing varying amounts of synthetic data on the resulting output will be investigated.



### 5.3 Intracranial hemorrhage

Intracranial hemorrhage refers to any bleeding within the intracranial vault, including the brain parenchyma and surrounding meningeal spaces (Caceres & Goldstein, 2012). Acute intracranial hemorrhage (ICH) is a potentially life-threatening condition that requires fast and accurate detection because of its frequently rapid progression during the first several hours. Intracranial hemorrhage (ICH), a subtype of stroke, can be classified into five sub-types according to bleeding location: Intraventricular (IVH), Intraparenchymal (IPH), Subarachnoid (SAH), Epidural (EDH) and Subdural (SDH). The ICH that occurs within the brain tissue is called Intracerebral Hemorrhage (Figure 15). Although ICH is less frequent than ischemic stroke, it presents a higher mortality rate. The degrees of severity and interventions vary with bleeding types (Ye et al., 2019). Classification of ICH and distinguishing it from ischemic stroke is critical due to prompt appropriate treatment and mitigate neurological deficit, and mortality.

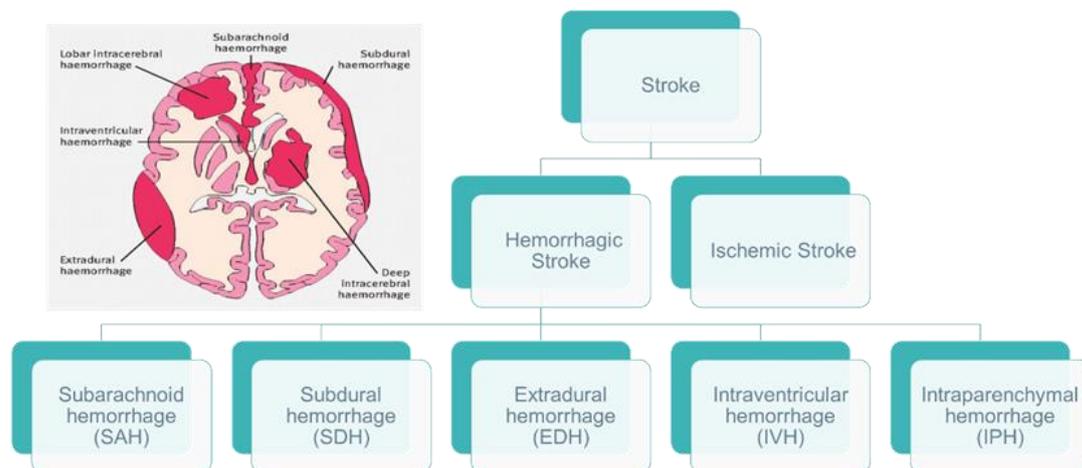


Figure 15: Sub-types of stroke and hemorrhagic stroke.

Recent advances in deep convolutional neural networks (CNNs) have shown that their potential in automating ICH detection and segmentation, and has proven that they can assist junior radiology trainees when experts are not available. Due to their capability of self-learning of nonlinear image filters and self-extraction of relevant features, CNNs have proven to be superior to methods that demand complicated engineering feature including skull stripping, image registration, and feature extraction from voxel intensity and local moment information (Muschelli et al., 2017; Ye et al., 2019).

There are a number of publicly available datasets that can be used for training for CNNs (Wang et al., 2021). This data is often released for classification challenges. Publicly available datasets containing intracranial haemorrhage scans are listed below:

- The Radiological Society of North America (RSNA) Intracranial Hemorrhage Detection Challenge dataset consists of over one million images from 25272 examinations, which provides the largest multi-institutional and multinational dataset for the study of acute ICH detection and subtype classification (Figure 16).



- The PhysioNet-ICH dataset includes a total of 75 participants (36 with ICH and 39 normal controls) with a total of 2814 slices (318 with bleeding and 2496 normal images). The number of CT scans (slices) for each subtype is 5(24) IVHs, 16(73) IPHs, 7(18) SAHs, 21(173) EDHs, and 4(56) SDHs.
- The CQ500 dataset, collected from multiple radiology centers in New Delhi, India, consists of a total of 491 CT scans. The 205 ICH scans contain all five subtypes, including 28 IVHs, 134 IPHs, 60 SAHs, 13 EDHs, and 53 SDHs.

In these datasets, however, the distribution of each ICH subtype is highly unbalanced. EDH incidence is higher among adolescents and young adults because with the increase of age, the dura mater becomes more adherent to the overlying bone. For example in the RSNA dataset, only 1.5% of all CT scans contain EDH, which is consistent with everyday clinical observations. Collecting enough data for these rare ICH subtypes are challenging and models tend to overfit if trained using limited data.

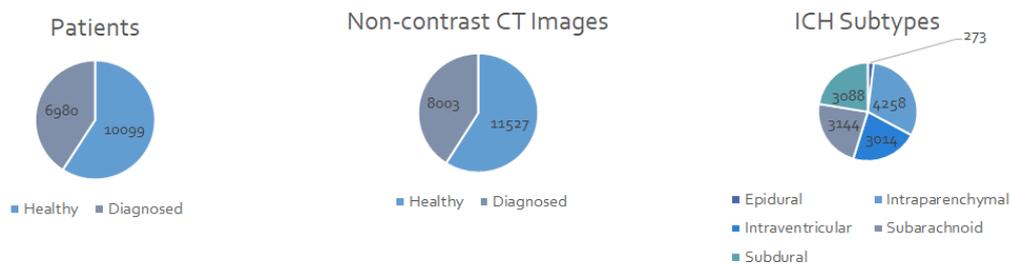


Figure 16: RSNA intracranial hemorrhage detection challenge dataset.

Standard image transformation techniques are traditional techniques to overcome unbalanced training data problem but generative methods that can extract synthetic images from entire dataset have been applied to enhance segmentation performance of CNNs. Karki, M., Cho, J. and Ko, S. (2020) examined data augmentation with synthetic images and segmentation label masks that are extracted by a lesion-conditional GAN for a haemorrhagic segmentation network and demonstrated that performance of segmentation models trained with smaller sizes of data benefited more from addition of synthetic data. Ganeshkumar M et al. (2022) proposed a novel data augmentation method to solve the problem of class imbalance in the ICH dataset using CycleGAN which improved the ICH identification performance significantly. The proposed method achieved a macro average F1-score of 0.91 and a specificity of 0.99 and a sensitivity of 0.80 in the ICH identification task and as a segmentation tool and achieved a dice score of 0.32 and a mean Intersection Over Union (IOU) of 0.22 for all the five ICH sub-types.

In an attempt to solve the issue of class imbalance present in the ICH training dataset, we used FastGAN and StyleGAN2-ADA approaches to generate synthetic epidural images. ADA is an adaptive discriminator augmentation mechanism that significantly stabilizes training in limited data regimes. 2019 RSNA Challenge and PhysioNet-ICH dataset epidural haemorrhage images were used as our model training data. The training data includes 3145 images with epidural haemorrhage at 512x512 resolution. The pre-trained Flickr-Faces-HQ model was used as starting model of transfer learning. Example results from methods are illustrated in Figure 17.

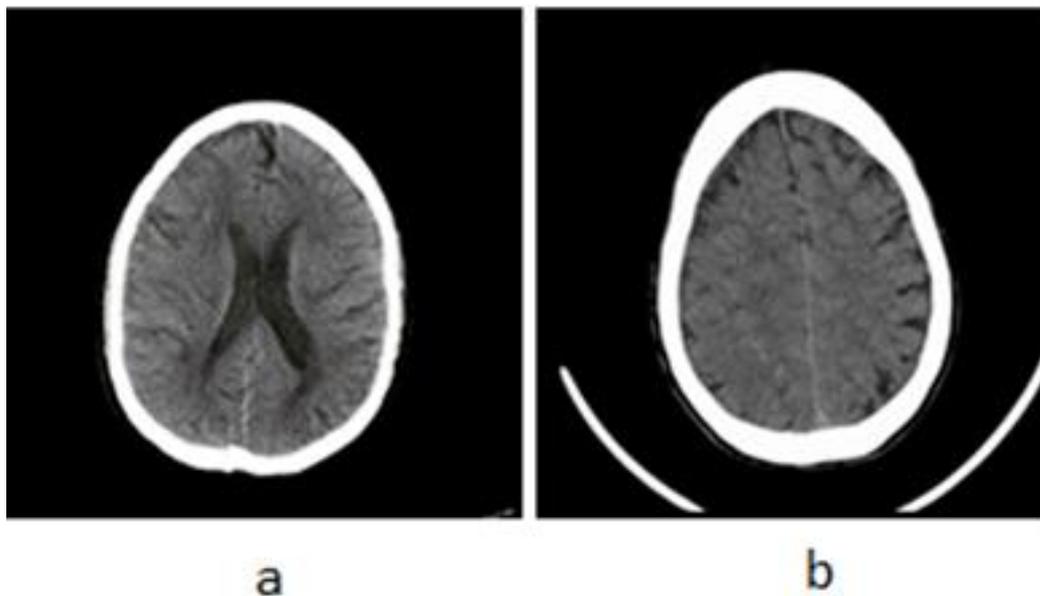


Figure 17: (a) Synthetic image generated with FastGAN (b) Synthetic image generated with StyleGan2-ADA

## 5.4 Image guided lung interventions

Several biopsy and treatment procedures in lung involve image guidance. Especially peripherally in the lung, beyond the reach of the bronchoscope's camera. There, 2D X-ray (fluoroscopy) and, for the more challenging cases, 3D X-ray imaging is used. Although X-ray enables visibility of anatomy and devices deep in the lung, it exposes the patient to radiation. Especially for 3D X-ray imaging, it is essential to find the right balance between radiation dosage to staff and patient, against the ability to confidently see devices and anatomical structures. In addition, the ability of software algorithms to detect devices and anatomy in 3D volumes with high consistency.

The process of developing and optimizing 3D acquisition protocols (dose (kV/mA), rotation time, frame rate, X-ray filters, reconstruction parameters, etc.) ideally is performed without exposing patients to unnecessary radiation. Several other options are available, include performing 3D scans of phantoms, of animal studies, and synthetic data generation.

The contribution of phantoms is limited, as they lack the subtle structures and motion which the human body has. Animal studies are cumbersome to arrange, a burden for the animals, and present unfamiliar anatomy to clinical users. Synthetic data generation from clinical scans can be a viable option, especially when studying lower dose protocols and protocols involving C-arm rotations with less accuracy (for example, when using smaller or lower cost C-arm geometries and motorization). This is because a good understanding of the X-ray physics and acquisition steps enables the downgrading or approximation of existing clinical data to match the imperfections and signal loss associated with lower dose and less accurate motion. The benefit of using a physics-based approach is that large data sets are not required, and that the image quality can be determined based on physics. The downside is a limited ability to



model higher dose scans, and the continued legal limitation of sharing the clinical data.

#### 5.4.1 Generation of Synthetic 3D CBCTs based on Clinical CBCT Data

To generate a synthetic 3D CBCT volume from a clinical CBCT dataset, we manipulate the 2D images corresponding to the 3D acquisition, and the corresponding meta-data such as the positions of the C-arm, X-ray tube and detector, and information about the amount of X-ray emitted to acquire the 2D image. These 2D images then are input to a new 3D reconstruction, often also using different reconstruction parameters, see Figure 18.

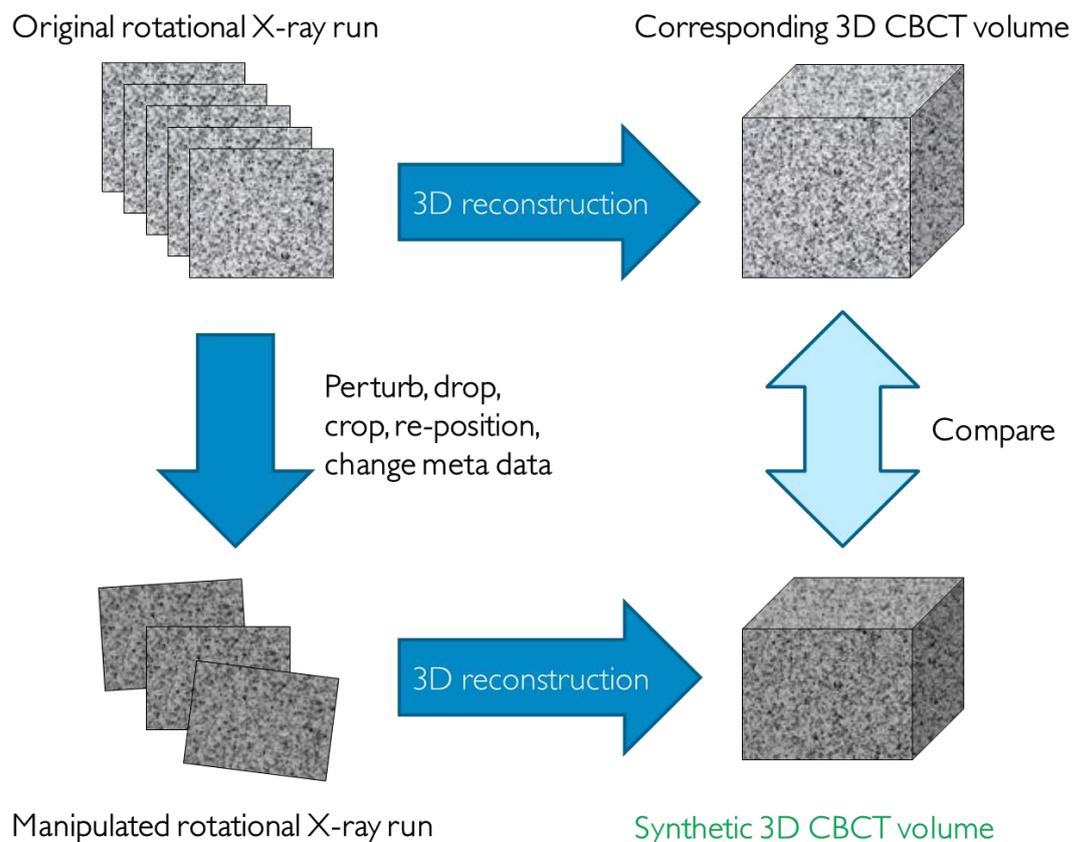
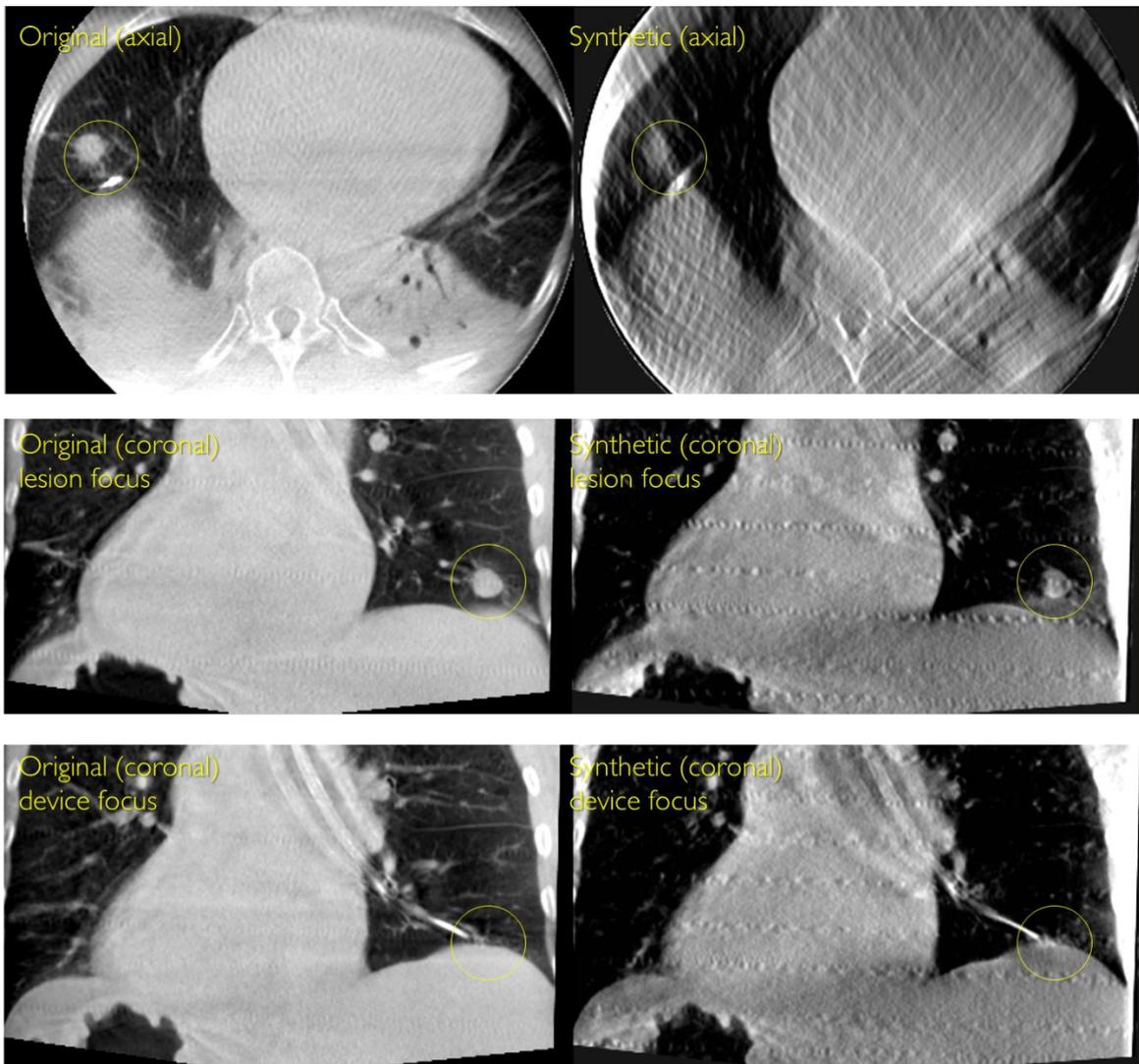


Figure 18: Steps to generate synthetic 3D CBCT volume.

For example, physics indicates that a reduction of X-ray can be emulated by adding Poisson noise to acquired images. A reduced frame rate and a smaller detector size can be emulated by respectively dropping images from the dataset, or cropping the images. Non-reproducible motion artefacts can be emulated by perturbing the X-ray source and X-ray detector positions and rotations as they are input into the reconstructor. The absence of certain meta data can emulated by running reconstructor steps with different inputs.

The synthetic volume can then be assessed (or processed) and results can be compared against the ground truth, see Figure 19 for some examples.



*Figure 19. Example of synthetic data representing how a low X-ray dose short CBCT rotation was generated in order to analyse the impact of such a 3D scan on 'device-in-lesion confirmation'. Both lesion and device remain visible and enable a device-in-lesion assessment, with the device clearly adjacent to the lesion.*

Synthetic 3D CBCT volumes are being used to evaluate and optimize alternative X-ray protocols and X-ray system geometries. Synthetic 3D volumes are also used to test the robustness of algorithms processing 3D CBCT volumes.



## 6 Quality metrics for synthetic medical images

There is currently no consensus on how to evaluate the quality and diversity of synthetic images, since synthetic images generation is a rather new research field. A straightforward approach is the training of different networks using the synthetic images, and to then test the trained networks on real images, but this can take a very long time (especially when trying many different hyper parameters).

### 6.1 Quality of synthetic images from a mathematical perspective

Evaluation metrics for images generated using GANs, or other AI models, can be categorized into quantitative and qualitative metrics. It is difficult to obtain the perfect measure, but a small comparison to find the best model can be performed using these measures. Borji (2019) mentions in his paper the desired properties which an efficient evaluation measure should have. The qualitative measures include that a generative model can generate high fidelity and diverse samples, can handle controllable sampling, has well-defined bounds, is sensitive to image distortions and transformations, must have low computational complexity, and finally must agree with human perceptual judgments. While qualitative models help to inspect and tune models they still suffer from various problems. To overcome these problems different quantitative measures can be used. In the paper by Borji (2019) a total of 24 quantitative measures are presented, which can be used to quantify the quality of the synthetic images. Some of the measures discussed are “model agnostic” where the generator is used as a black box and does not require the density estimation from the model. On the other hand, there are some measures like average log-likelihood which have a requirement of estimation of the probability distribution from the samples.

For 2D images, several quantitative measures have been introduced. Salimans et al. (2016), introduced inception score (IS) which involves using a pre-trained deep learning neural network model for image classification to classify the generated images. Further the probability of images belonging to each class is predicted and then summarized into the inception score, where it takes into account the image quality and diversity. Heusel, et al. (2017) introduced another metric called Fréchet inception distance (FID) to assess the quality of the images generated by generative model like GANs. Unlike (IS) that evaluates only the distribution of generated images, (FID) compares the distribution of generated images with the distribution of real images on which the generator was trained. (FID) is considered as a standard measuring metric for generative models.

Both (IS) and (FID) are based on using 2D CNNs pre-trained on ImageNet, therefore these metrics are not as useful for medical images or volume data, since ImageNet does not contain medical images or volumes and since there are very few pre-trained 3D CNNs. For medical images, a better approach would be to pre-train CNNs on different openly available medical datasets.



## 6.2 Quality of synthetic images for treatment planning

In radiation therapy treatment planning a dose calculation engine computes the 3D spatial dose distribution in the patient resulting from the optimized irradiation on the patient. The distribution of dose is highly dependent on the radiodensity of the irradiated tissue, quantified using a Hounsfield scale. The Hounsfield units are obtained from a linear transformation of the measured attenuation coefficients of the CT. Thus, to use synthetic image data for treatment planning, the Hounsfield units of the synthetic images must be accurate, allowing the resulting dose distribution to be calculated correctly.

If the synthetic images are generated from another image of the patient, e.g., a synthetic CT is generated from a daily CBCT or MR, the quality of the synthetic image can be evaluated by comparing the dose distribution on the synthetic CT with the dose distribution on the original CT. This caveat of this test is that it assumes the presence of an original CT, which is registered to the synthetic CT have been registered. Ideally, the CT should also have been acquired the same day to minimize anatomical differences between the images, but it should be noted that certain anatomical differences (e.g. filling of the rectum or bladder), can occur in a smaller timescale, hampering this experiment.



## 7 Legal aspects of synthetic images

The capacity of AI to create convincing and representative synthetic data is a novel phenomenon that is likely to have a large societal, legal and scientific impact. Because this type of technology was only developed in the second half of the 2010s, there are limited legal and ethical reflections with regard to AI generated synthetic data. Here we focus on three important legal and ethical aspects of the use of synthetic data in a research context.

(1) *Synthetic data as a means of anonymization.* Currently the use and re-use of personal data (such as medical images) is in Europe constrained by the requirements from data protection law, notably GDPR 2016/679. Currently, many anonymization techniques entail that the utility of the data is diminished. Creating anonymous synthetic datasets (Jordon et al. 2018, Xie et al. 2018) that have the same utility as the original personal data would have a major beneficial impact on data sharing and re-use in research. However, whether synthetic data really can be qualified as anonymous in terms of data protection is a complex question that needs closer investigation (Bellovin et al. 2019). While synthetization overall seems to be an effective method in creating anonymous data, synthetic datasets can contain accidental *Doppelgänger*s that happen to be very similar to examples from the original training set. There is currently no consensus regarding how different two medical images need to be for this purpose. From this perspective, it is safer to share synthetic images that have been compared to the training set, as opposed to sharing a trained AI model that can accidentally generate examples that are close to the original training examples. It should be mentioned that there are GANs who can be specifically trained to preserve privacy.

(2) *Ethics for data sharing.* If a dataset containing medical images is going to be shared openly, the participants normally need to provide consent prior to data collection. It is not clear if consent is required if synthetic images are shared, since the synthetic images do not belong to a specific person, even though their medical data was used to train the models that generate the synthetic images. As long as the synthetic medical images cannot be connected or traced to the participants (see previous paragraph), it can be argued that consent is not required. Larson et al. (2020) argue that clinical data should be treated as a form of public good, to be used for the benefit of future patients, and further argue that consent is not required before collected data are used for secondary purposes when obtaining such consent is prohibitively costly or burdensome (e.g. contacting 10,000 persons). How ethical review committees in different countries look at this question can only be determined by applying for ethical approval to share synthetic images. It should however be noted that ethical committees do not normally consider the legal aspects or consequences of a project.

(3) *Intellectual property (IP) and synthetic data: who has intellectual ownership over the data?* Intellectual property law builds on the idea that human authors, inventors, and database creators should be granted IP rights that will act as an economic incentive for intellectual labour and investments. A second important assumption is that creativity and inventiveness is a purely human capacity. AI-generated synthetic data have challenged this assumption and raised the question if AI could be an author in term of copyright law (Craig and Kerr, 2019) or an inventor in terms of patent law. Are synthetic medical images protected by copyright law, patent law or database rights? Or do synthetic images fall outside the scope of IP protection? Which role the



creation of synthetic data will fulfil in the economics of research is highly dependent on how such creation is qualified in terms of IP.



## 8 Conclusion

In this document we have presented state-of-the-art methods for creating synthetic images (from noise or from other images), and how this can be applied to some of our use cases. Furthermore, we have briefly mentioned how synthetic images can be seen from a legal perspective. While synthetic images are promising, there are still many aspects that require further research.



## 9 References

Amirrajab, Sina, et al. "Xcat-gan for synthesizing 3d consistent labeled cardiac MR images on anatomically variable xcat phantoms." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2020.

Bellovin, S. M., Dutta, P. K., & Reitering, N. (2019). Privacy and synthetic datasets. *Stan. Tech. L. Rev.*, 22, 1.

Borji, Ali. "Pros and cons of GAN evaluation measures." *Computer Vision and Image Understanding* 179 (2019): 41-65.

Bu, T., Yang, Z., Jiang, S., Zhang, G., Zhang, H., & Wei, L. (2021). 3D conditional generative adversarial network- based synthetic medical image augmentation for lung nodule detection. *International Journal of Imaging Systems and Technology*, 31(2), 670-681

Caceres, J.A., & Goldstein, J.N. (2012). Intracranial Hemorrhage. *Emerg Med Clin North Am.* 30(3), 771–794.

Cha, K. H., Petrick, N. A., Pezeshk, A. X., Graff, C. G., Sharma, D., Badal, A., & Sahiner, B. (2019). Evaluation of data augmentation via synthetic images for improved breast mass detection on mammograms using deep learning. *Journal of Medical Imaging*, 7(1), 012703.

Chan, Eric R., et al. "Efficient geometry-aware 3D generative adversarial networks." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Cid-Mejías, Antón, et al. "A deep learning approach using synthetic images for segmenting and estimating 3D orientation of nanoparticles in EM images." *Computer Methods and Programs in Biomedicine* 202 (2021): 105958.

Cirillo, M. D., Abramian, D., & Eklund, A. (2020). Vox2Vox: 3D-GAN for brain tumour segmentation. In *International MICCAI Brainlesion Workshop* (pp. 274-284). Springer, Cham.

Chan, Eric R., et al. "Efficient geometry-aware 3D generative adversarial networks." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

Costa, Pedro, et al. "End-to-end adversarial retinal image synthesis." *IEEE transactions on medical imaging* 37.3 (2017): 781-791.

Craig, C. J., & Kerr, I. R. (2019). The Death of the AI Author. *SSRN 3374951*

Dai, W., Doyle, J., Liang, X., Zhang, H., Dong, N., Li, Y., Xing, E.P., 2017b. Scan: structure correcting adversarial network for chest x-rays organ segmentation. arXiv:1703.08770.



Deng, J., ... Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *IEEE conference on computer vision and pattern recognition*

Dou, Qi, et al. "Pnp-adanet: Plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation." arXiv:1812.07907 (2018).

Eklund, A. (2019). Feeding the zombies: Synthesizing brain volumes using a 3D progressive growing GAN. *arXiv preprint arXiv:1912.05357*

Eilertsen, G., Tsirikoglou, A., Lundström, C., & Unger, J. (2021). Ensembles of GANs for synthetic training data generation. *arXiv preprint arXiv:2104.11797*.

Foroozandeh, M., & Eklund, A. (2020). Synthesizing brain tumour images and annotations by combining progressive growing GAN and SPADE. *arXiv:2009.05946*

Frid-Adar, Maayan, et al. "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification." *Neurocomputing* 321 (2018): 321-331.

Ganeshkumar M, Ravi, V., Sowmya V, Gopalakrishnan E. A, Soman K. P and Chakraborty, C., 2022. Identification of intracranial haemorrhage (ICH) using ResNet with data augmentation using CycleGAN and ICH segmentation using SegAN. *Multimedia Tools and Applications*

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680)

Heusel, Martin, et al. "GANs trained by a two time-scale update rule converge to a local nash equilibrium." *Advances in neural information processing systems* 30 (2017).

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in Neural Information Processing Systems* 33 (2020): 6840-6851.

Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Jordon, J., Yoon, J., & van der Schaar, M. (2018). PATE-GAN: Generating synthetic data with differential privacy guarantees. In *International Conference on Learning Representations*.

Jung, E., Luna, M., & Park, S. H. (2021). Conditional GAN with an Attention-Based Generator and a 3D Discriminator for 3D Medical Image Generation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 318-328). Springer, Cham



Kamnitsas, Konstantinos, et al. "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks." International conference on information processing in medical imaging. Springer, Cham, 2017.

Kang, S. K., An, H. J., Jin, H., Kim, J. I., Chie, E. K., Park, J. M., & Lee, J. S. (2021). Synthetic CT generation from weakly paired MR images using cycle-consistent GAN for MR-guided radiotherapy. *Biomedical Engineering Letters*, 11(3), 263-271.

Karki, M., Cho, J., & Ko, S. (2020). Lesion conditional image generation for improved segmentation of intracranial hemorrhage from CT images. arXiv preprint arXiv:2003.13868.

Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196. 2017 Oct 27.

Karras, Tero, et al. "Alias-free generative adversarial networks." Advances in Neural Information Processing Systems 34 (2021): 852-863

Kossen, Tabea, et al. "Synthesizing anonymized and labeled TOF-MRA patches for brain vessel segmentation using generative adversarial networks." *Computers in biology and medicine* 131 (2021): 104254.

Kwon, G., Han, C., & Kim, D. S. (2019). Generation of 3D brain MRI using auto-encoding generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 118-126). Springer, Cham

Larson, D. B., Magnus, D. C., Lungren, M. P., Shah, N. H., & Langlotz, C. P. (2020). Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework. *Radiology*, 192536.

Leece, R., Xu, J., Ostrom, Q. T., Chen, Y., Kruchko, C., & Barnholtz-Sloan, J. S. (2017). Global incidence of malignant brain and other central nervous system tumours by histology, 2003–2007. *Neuro-oncology*, 19(11), 1553-1564.

Ma, Li, et al. "Combining DC-GAN with ResNet for blood cell image classification." *Medical & Biological Engineering & Computing* 58.6 (2020): 1251-1264.

Muschelli, J., Sweeney, E., Ullman, N., Vespa, P., Hanley, D. and Crainiceanu, C. (2017). PltchPERFeCT: Primary Intracranial Hemorrhage Probability Estimation using Random Forests on CT. *NeuroImage: Clinical*, 14, pp.379-390.

Neff, Thomas, et al. "Generative adversarial networks to synthetically augment data for deep learning based image segmentation." Proceedings of the OAGM workshop. 2018.

Nijhawan, A., Improving brain tumour segmentation using synthetic images from GANs, master thesis, Linköping university, [urn:nbn:se:liu:diva-182048](https://nbn-resolving.org/urn:nbn:se:liu:diva-182048) , 2021



Poldrack, R. A., Barch, D. M., Mitchell, J., Wager, T., Wagner, A. D., Devlin, J. T., ... & Milham, M. (2013). Toward open sharing of task-based fMRI data: the OpenfMRI project. *Frontiers in neuroinformatics*

Salimans, Tim, et al. "Improved techniques for training GANs." *Advances in neural information processing systems* 29 (2016).

Sandfort, Veit, et al. "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks." *Scientific reports* 9.1 (2019): 1-9.

Schlegl, Thomas, et al. "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery." *International conference on information processing in medical imaging*. Springer, Cham, 2017.

Van Essen, D. C., ..., Ugurbil, K., & Wu-Minn HCP Consortium. (2013). The WU-Minn human connectome project: an overview. *Neuroimage*, 80, 62-79

Wang, X., Shen, T., Yang, S., Lan, J., Xu, Y., Wang, M., Zhang, J., & Han, X. (2021). A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans. *NeuroImage: Clinical*, 32, 102785. <https://doi.org/10.1016/j.nicl.2021.102785>

Wolterink, Jelmer M., et al. "Deep MR to CT synthesis using unpaired data." *International workshop on simulation and synthesis in medical imaging*. Springer, Cham, 2017.

Wu, J., Zhang, C., Xue, T., Freeman, B., & Tenenbaum, J. (2016). Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. *Advances in neural information processing systems*, 29

Ye, H., Gao, F., Yin, Y., Guo, D., Zhao, P., Lu, Y., Wang, X., Bai, J., Cao, K., Song, Q., Zhang, H., Chen, W., Guo, X. and Xia, J. (2019). Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *European Radiology*, 29(11), pp.6191-6201.

Yi, Xin, Ekta Walia, and Paul Babyn. "Generative adversarial network in medical imaging: A review." *Medical image analysis* 58 (2019): 101552.

Yi, Xin, and Paul Babyn. "Sharpness-aware low-dose CT denoising using conditional generative adversarial network." *Journal of digital imaging* 31.5 (2018): 655-669.

Yi, X., Walia, E., Babyn, P., 2018. Unsupervised and semi-supervised learning with categorical generative adversarial networks assisted by Wasserstein distance for dermoscopy image classification. *arXiv:1804.03700*.

Yoon, Jinsung, Lydia N. Drumright, and Mihaela Van Der Schaar. "Anonymization through data synthesis using generative adversarial networks (ads-gan)." *IEEE journal of biomedical and health informatics* 24.8 (2020): 2378-2388.



Xie, L., Lin, K., Wang, S., Wang, F., & Zhou, J. (2018). Differentially private generative adversarial network. *arXiv preprint arXiv:1802.06739*.

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232).