Sinogram Enhancement with Generative Adversarial

Networks using Shape Priors

**Abstract:** Compensating scarce measurements by inferring them from computational models is a way to address ill-posed inverse problems. We tackle Limited Angle Tomography by completing the set of acquisitions using a generative model and prior-knowledge about the scanned object.

Using a Generative Adversarial Network as model and Computer-Assisted Design data as shape prior, we demonstrate a quantitative and qualitative advantage of our technique over other state-of-the-art methods. Inferring a substantial number of consecutive missing measurements, we offer an alternative to other image inpainting techniques that fall short of providing a satisfying answer to our research question: can X-Ray exposition be reduced by using generative models to infer lacking measurements?

**Key Words:** Generative Adversarial Networks, Image Inpainting with Edge Information, X-Ray Computed Tomography, Computer Assisted Design Data, Shape Priors

# I - Introduction

X-Ray Computed Tomography (XCT) is a versatile 3D imaging technique that allows the estimation of volumetric X-Ray attenuation profiles. It produces cross-sectional images of bodies sensitive to this radiation by sampling an object from different viewing angles to reconstruct an image from this sequence of acquisitions. Yet, X-Rays are toxic for in-vivo diagnosis and time-consuming in industrial testing. There is a trade-off between sufficient sampling for high-quality images and X-Ray intake for time and health constraints. Can computational methods exploit prior knowledge about the scanned object to compensate scarce acquisitions by inferring measurements?

We use Generative Adversarial Networks (GAN) to complete the sequence of scarce acquisitions by inferring them from Computer-Assisted Design data. When imaging a slice through an object from the viewing angle , with a detector with pixels, the measurement can be described as

with the index of pixels on the detector, the coordinates of the object and the density function of the object, that maps a spatial position to a material density. Let be the set of viewing angles at which the object is sampled. The set of measurements

is the sinogram of the image . It can be represented as an image of size . As such, missing acquisition in the sequence is represented as zero-valued pixels along one dimension of the sinogram. The problem of inferring missing acquisitions from a scarce sinogram is then similar to the one of inferring arbitrarily large regions in images based on image semantics, known as semantic inpainting.

Semantic image inpainting is a constrained image generation problem 1. Missing parts of an image are inpainted using a generative network and solving an optimisation problem. GAN 2 and their fully-convolutional version Deep-Convolutional GAN (DCGAN) 3 are adapted to this task: they were used for image 4 and sinogram 5 inpainting. The optimisation relies on finding the "closest" encoding in the latent space of the GAN distribution by minimising a penalty function that encompasses contextual and conditional information. This method suffers from several limitations:

* Walking the latent space of the distribution can only yield certain improvements in the image generated by the GAN: 6 shows that images can only be transformed to some degree (brightness, zoom, rotation). Not only is the transformation corresponding to inpainting is not defined, but it has no certainty to be achievable by "steering" the generated output, especially when guided by a generic loss function and not by a supervised walk.
* The optimisation function adds computational time and hyperparameters. In addition to training the generative model, the optimisation process is time-consuming and also requires fine-tuning of the learning rate and number of iterations.

As an alternative to this process, we use CAD data as a prior and train a Unet-GAN 7,8 to infer missing parts of the sinogram given shape information about the scanned object.

Shape information is often available in both medical and industrial imaging but is rarely used in XCT reconstruction. For instance, in medical imaging, projects such as 9 and 10 demonstrate the potential of using numerical shape models of a generic human body to minimise X-Ray exposure. An alternative is to extract prior information from earlier scans 11,12. For manufactured components, Computer Assisted Design (CAD) drawings are often available, providing strong constraints on object shape. These types of priors provide estimates of object boundary locations, though might not contain information on exact X-Ray absorption within an object, nor do they contain information about unknown defects and inclusions.

In this paper, we minimise X-Ray intake by inpainting the scarce sinogram with a GAN and a shape prior. The main advantage compared to other inpainting methods is the side-stepping of the optimisation process. Unlike the other image inpainting methods, we focus on inpainting the missing part

of the sinogram only, reducing the complexity of

the task. Our main findings are:

* Exploiting the specificity of XCT data enhances the CAD prior. Before feeding the CAD to the GAN, we rescale its values so that they match what has been observed in the scarce sinogram.
* Prior information about the shape of the object significantly improves the quality of state-of-the-art (SOTA) sinogram-enhancing techniques. We show that including the CAD prior to other methods that address a similar problem yields a significant improvement of their performance.
* Not only does our method enhance the sinogram in terms of Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM) compared to other SOTA techniques, but we also report an improvement of the reconstructed image quality.

# II - PROPOSED APPROACH

Our method is focused on inpainting a scarce sinogram using information from a shape prior, instead of an optimisation process. As such, it requires an initial training of the generative model, in our case a GAN, followed by the inpainting process. The training procedure is visually detailed in Fig. \ref{fig:training\_procedure\_explanation}.

1. **GAN Introduction**

A GAN is made of two networks that compete against each other. The generator *G* tries to generate samples that follow the same distribution as the training examples by up-sampling a noise vector drawn from a noise distribution . The discriminator *D* tries to discriminate between real training samples and those generated by G. GANs have been proposed as an alternative to avoid difficulties commonly found in deep generative models, such as explicit density estimation. GAN optimisation is done by solving the minimax problem:

where is the expectation over the training dataset. This architecture uses fully-connected units, which limit the maximum size of images the GAN can generate. To scale to larger images, the Deep Convolutional GAN (DCGAN) architecture \cite{radford2015unsupervised} was proposed.

1. **Deep Convolutional GAN and Unet - The pix2pix Architecture**

GAN are notoriously difficult to train 13 and do not handle prior-knowledge as they sample a latent distribution. To address these issues, 8 proposes the pix2pix architecture that combines the DCGAN improvements and includes shape priors by changing the generator into a U-net architecture.

The main change offered by the DCGAN is the replacement of fully-connected units by convolutional layers. Other modifications include the replacement of the maxout 14 activation by ReLu and Tanh in the generator and LeakyReLu in the discriminator, the inclusion of batch-normalisation 15 and the replacement of pooling units with learnt sampling units.

U-nets were initially designed for image segmentation. They are fully convolutional auto-encoders that have residual connections between the down-sampling and up-sampling units. They are convenient architectures that can map a multi-channel input image to a one-channel output image, or vice versa. The pix2pix architecture makes use of this design to add prior-knowledge to the GAN.

1. **Adapted Loss Function**

To encourage the generator to produce images close to their target value, in the sense of the L1-loss, an additional term is added to the training loss as prescribed in 1. We use the L1-Loss which is a good choice for image quality when Poisson noise is present, as is the case in XCT images 16 .Unlike other inpainting processes, we train the network to infer only the missing part of the image. We thus use the loss function:

where and are the sinograms with all and missing data respectively, is an image that encodes the shape prior and a weighting parameter. Given a sinogram and a shape prior, *D* determines if the sinogram was generated by the GAN or drawn from the true dataset. Given a sinogram with missing data, generates missing acquisitions that are close to the ground-truth acquisitions in terms of L1-Loss17.

1. **Architectural Details**

We follow the network's description given in 8. Let Ck be a Convolution-BatchNorm-ReLU layer with k filters. Let CDk be a Convolution-BatchNorm-Dropout-ReLU layer with a dropout rate of 50% 18 .All convolutions are 4 by 4 spatial filters applied with a stride of 2. Convolutions in the encoder down-sample by a factor of 2, whereas in the decoder they up-sample by a factor of 2. After the last layer in the decoder, a convolution is applied, followed by a Tanh function and an element-wise multiplication with a mask to infer missing acquisitions only, as explained in \ref{encoding\_the\_shape\_prior}. As an exception to the above notation, BatchNorm is not applied to the first C64 layer in the encoder. All ReLUs 19 in the encoder are leaky 20 , with slope 0.2, while ReLUs in the decoder are not leaky. This results in the following generator:

**Encoder :** C64-C128-C256-C512-C512-C512-C512-C512

**Decoder :** CD512-CD1024-CD1024-C1024-C1024-C512-C256-C128

The discriminator works on image patches of size 16x16. For the discriminator, after its last layer and after extracting the patch, a convolution is applied to map to a one-channel output, followed by a Sigmoid function. This results in the following discriminator:

**Encoder** **:** C64-C128-C256-C512

1. **The Inpainting Process**

Once the network is trained and its parameters frozen, the inpainting process is straightforward. It requires a scarce sinogram, a sufficiently sampled CAD sinogram and a mask indicating the acquisitions to infer. Given this input data, *G* infers the missing acquisitions in a one-channel image. This image is then added pixel-wise to the scarce sinogram to produce the inpainted sinogram. A visual explanation of the method is given in Fig. \ref{fig:inpainting\_method\_explanation}.

1. **Other Methods for Comparison**

An immediate solution to inferring missing acquisitions is to use linear interpolation. Using the shape prior, replacing the missing acquisitions by the ones expected by the CAD is another solution. An improvement of the acquisition prior is brought by scaling it with attenuation values matching the ones seen in the measured sinogram: this scaling is detailed in \ref{encoding\_the\_shape\_prior}. We here implement these three solutions for comparison.

The pix2pix architecture used here has never been used for XCT data but other similar methods exist. Indeed, 21 replaces missing acquisitions by linearly interpolated ones and then uses a Unet to enhance their quality. We implement this architecture without the shape prior as in the original paper and also produce a version that includes our shape prior to make it more comparable to our new approach. A GAN architecture is also used for sinogram synthesis 5 , with an optimisation procedure to perform the inpainting. This model does not use shape priors. We implemented this architecture and trained it with as much care as we trained the others but could not obtain a satisfactory result.