Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Physics

Ruben Skjelstad Dragland

Dynamic Micro-CT utilising Golden Ratio Sampling and Three-dimensional Generative Adversarial Networks

Master's thesis in Nanotechnology
Supervisor: Dag Werner Breiby
Co-supervisor: Anders Kristoffersen
June 2023

NTNU
Norwegian University of
Science and Technology

equinor

Ruben Skjelstad Dragland

Dynamic Micro-CT utilising Golden Ratio Sampling and Three-dimensional Generative Adversarial Networks

Master's thesis in Nanotechnology
Supervisor: Dag Werner Breiby
Co-supervisor: Anders Kristoffersen
June 2023

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Physics

NTNU
Norwegian University of
Science and Technology

ABSTRACT

Temporal computed tomography, dynamical computed tomography, or just 4D-CT, is an X-ray imaging technique where dynamical processes can be resolved in three dimensions over time. However, CT is a technique that requires many sampled projections in order to produce reconstructions of high quality. Consequently, the achievable temporal resolution is low. Therefore, in order to extract the maximal amount of valuable insights from a temporal CT scan, it is imperative that one has the ability to prioritise between temporal and spatial resolution in the post-processing of the data. Then, it would be possible to reconstruct with high spatial resolution when the dynamical process allows for it, and sacrifice this resolution when spatial undersampling is required to resolve the dynamical process. In the latter case, there is a great demand for innovative methods for minimising the mentioned sacrifice. A technique that enables both the mentioned properties would be vital in multiphase flow studies in porous media, where insights could increase the oil recovery factor from reservoirs, and accelerate the transition towards carbon-neurality through carbon capture and storage (CCS) technologies.

In this master's thesis, a temporal CT technique based on golden ratio sampling and undersampled reconstruction enhancement was developed. The motivation behind this technique was, as mentioned, to enable laboratory 4D-imaging with resolution flexibility *a posteriori*. The spatial resolution could be given lower priority due to the enhancing and denoising abilities of a diversely trained three-dimensional convolutional generative adversarial network (GAN), which aimed at removing undersampling artefacts from CT reconstructions. With the derived technique, a dynamical CT scan of an ordinary hourglass was performed, as well as a static CT scan of a cracked cylindrical sandstone sample. The optimal performance of the trained GAN was found to be 52 projections for the hourglass sample, resulting in a mean structural similarity index (MSSIM) of approximately 0.80. In terms of Fourier shell correlation (FSC), the performance was as good as for twice the number of projections. From a human perception point of view, this number of projections resulted in nearly no undersampling artefacts after GAN-enhancement. The maximal temporal resolution tested was 26 s, which captured the sand-air interface with minor dynamical artefacts. The GAN struggled with distinguishing between undersampling artefacts and the actual features of the sandstone sample. However, improved results were achieved by employing the SIRT algorithm, training using the true sandstone probability distribution, and by training with a batch size of 4. In these cases, the resulting MSSIM was approximately 0.63 for 85 projections with the FDK algorithm. The same result was achieved when using 150 iterations and 34 projections of the SIRT algorithm in addition to a GAN trained with the true sandstone probability distribution. The GAN was observed to remove high-frequent noise and perform contrast enhancement when trained on FDK reconstructions, while it recovered contrast from blurring when trained on SIRT reconstructions. Proposed development of the GAN included training at elevated undersampling artefacts, adjustments to the generality of the dataset, hyperparameter tuning, and adaptation of the architecture to include sinogram data.

SAMMENDRAG

Temporal datatomografi, dynamisk datatomografi, eller bare 4D-CT, er en røntgenavbildningsteknikk hvor tredimensjonale dynamiske prosesser kan avbildes. Ulempen med CT er de mange projeksjonene som kreves for å rekonstruere tredimensjonale modeller av høy kvalitet. Dermed blir den oppnåelige tidsoppløsningen lav ved høy romlig oppløsning. Derfor er det avgjørende å ha en teknikk som gir fleksibilitet i valget mellom tids- og romlig oppløsning etter datatomografiopptak. Dette ville muliggjøre rekonstruksjoner med høy romlig oppløsning når den dynamiske prosessen tillater det, og rekonstruksjoner med høy tidsoppløsning fremfor romlig oppløsning når den dynamiske prosessen krever det. Etterspørselen er stor etter teknikker som minimerer konsekvensene av underoppløsning i datatomografi. En teknikk som forbedret dynamisk datatomografi i disse henseende ville være særdeles viktig i studier av strømning i porøse medier. Innsikt fra slike studier kan øke oljeutvinningseffektiviteten fra reservoarer og akselerere utviklingen av karbonfangst og -lagring.

I denne masteroppgaven ble det utviklet en teknikk for dynamisk datatomografi. Motivasjonen bak teknikken var å muliggjøre 4D billedtagning med fleksibel tids- og romlig oppløsning. Teknikken baserte seg på de spesielle egenskapene til det gylne snitt, samt maskinlærte rekonstruksjonsforbedringsfilter for underoppløste volum. Tidsoppløsning kunne dermed prioriteres over romlig oppløsning i datatomografiopptak på grunn av nettopp dette maskinlærte, tredimensjonale, generative motsetningsnettverket (GAN), som hadde som mål å fjerne artefakter fra underoppløste rekonstruksjoner. Den utviklede teknikken ble testet på den dynamiske prosessen i et timeglass, samt på en sprukket sandsteinprøve uten dynamikk. Det generative motsetningsnettverket fungerte best på rekonstruksjoner bestående av 52 projeksjoner for timeglasset. Da ble den midlere strukturelle likheten (MSSIM) lik 0.8, og den fouriertransformerte korrelasjonen (FSC) var like god som for rekonstruksjoner bestående av dobbelt så mange projeksjoner. Fra et menneskelig perspektiv var det nesten ingen underoppløsningsartefakter i filtrerte rekonstruksjoner med dette antallet projeksjoner eller flere. Den maksimale tidsoppløsningen som ble testet var på 26 s, der grensesnittet mellom sand og luft ble fanget med mindre dynamiske artefakter. Rekonstruksjonsforbedringsfilteret fungerte ikke like godt på sandsteinprøven. I dette tilfellet var det vanskelig å skille mellom underoppløsningsartefakter og signal. Bedre resultater ble imidlertid oppnådd ved å trene basert på SIRT-data, trene med sandsteinens sanne sannsynlighetsfordeling og ved å beregne gradientene ut fra en partistørrelse på fire. I disse tilfellene ble en MSSIM på 0.63 oppnådd. Dette gjaldt både ved 85 projeksjoner med FDK-algoritmen og ved 34 projeksjoner, 150 iterasjoner med SIRT-GAN. I tilfeller der treningsdatasettet besto av FDK-rekonstruksjoner, fjernet det generative nettverket høyfrekvent støy, og forbedret kontrasten, mens utsmørte kontraster ble gjenopprettet etter at nettverket hadde blitt trent på SIRT-rekonstruksjoner. Foreslått videre utvikling av modellen gikk ut på å trene nettverket med kraftigere underoppløsningsartefakter, samt å justere generaliteten i treningsdatasettet. I tillegg var hyperparameter-justering og arkitekturtilpasninger mulige løsninger, blant annet ved å inkludere sinogrammet som inndata til nettverket.

PREFACE

This master's thesis completes the five-year journey of the *Integrated Master's Degree in Nanotechnology* study programme at the Norwegian University of Science and Technology (NTNU). I have had the pleasure of being a part of a newly initiated collaboration between Equinor ASA's CT laboratory at Rotvoll and the X-ray research group at NTNU, lead by PhD Anders Kristoffersen and Professor Dag W. Breiby, respectively.

The topics of this master's thesis are temporal CT using golden ratio sampling and reconstruction enhancement using deep three-dimensional convolutional neural networks. Kim Robert Tekseth carried out the technical implementation of the golden ratio sampling procedure. My focus was instead to describe the procedure, optimising the sampling parameters, and carrying out initial tests. The adaptation, implementation, optimisation, dataset generation, and examination of the three-dimensional convolutional generative adversarial network were the other performed tasks during this project.

This work is the result of many contributions from helpful people. Firstly, I would like to thank my supervisors, Professor Dag W. Breiby and PhD Anders Kristoffersen, for their guidance and support throughout the project. A thank you to Kim Robert Tekseth for his technical implementation of the golden ratio sampling procedure, for the help with the TIGRE reconstruction toolbox, and for his general support. The rest of the team at the CT laboratory at Rotvoll also deserve a big thank you for teaching me their CT instrument, providing me with a nice training dataset, and for the overall good company these past months. I would also like to thank Associate Professor Basab Chattopadhyay, as it initially was his idea to use generative adversarial networks for reconstruction enhancement.

Once again, I would like to thank Dr. Magnus Moreau for the time working together for CrayoNano AS. Our simulation project prepared me well for this year's challenges.

Finally, I have to thank my friends and family. I would like to thank my class as well as my *München gang* for being a distraction from the thesis work when possible. Conversely, I would like to thank my family for returning my focus to the thesis when necessary.

Abschließend möchte ich der Technische Universität München *Servus!* und *Prost!* sagen. Viele von den Themen, die in dieser Arbeit relevant waren, habe ich während meiner Zeit an der TUM gelernt. Obwohl ich nur ein Jahr an der TUM studiert habe, war sie eine sehr wichtige Zeit.

Ruben Skjelstad Dragland
June 15, 2023

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AD Automatic Differentiation

ADAM Adaptive Moment Estimation

AI Artificial Intelligence

ANN Artificial Neural Network

AoR Axis of Rotation

CCS Carbon Capture and Storage

cGAN Conditional Generative Adversarial Network

CGD Conjugate Gradient Descent

CNN Convolutional Neural Network

CPU Central Processing Unit

CT Computed Tomography

DL Deep Learning

DNN Deep Neural Network

favg frame average

FBP Filtered Back Projection

FDK Feldkamp-Davis-Kress

FC Fully Connected

FFT Fast Fourier Transform

fps frames per second

FSC Fourier Shell Correlation

GAN Generative Adversarial Network

GD Gradient Descent

GPU Graphics Processing Unit

i.i.d Independent Indentically Distributed

ML Machine Learning

MLE Maximum-Likelihood Estimation

MSE Mean Squared Error

MSSIM Mean Structural Similarity Index Measure

NTNU Norwegian University of Science and Technology

PReLU Parametric Rectified Linear Unit

PSNR Peak Signal-to-Noise Ratio

RoI Region of Interest

SAXSTT Small Angle X-ray Scattering Tensor Tomography

SGD Stochastic Gradient Descent

SIRT Simultaneous Iterative Reconstruction Technique

SSIM Structural Similarity Index Measure

0.1 INTRODUCTION

The launch of ChatGPT [39] marked the entry of advanced language models into everyday life. Overnight, interactions with advanced artificial intelligence (AI) became a common way of obtaining information, just like searching the web. Businesses were forced to assess the impact of this launch, and figure out how to best enhance their own performance utilising this new technology. The drift towards artificial intelligence is also apparent within academia, mainly through the use of machine learning and deep neural networks in research [63]. The field of physics poses countless optimisation problems where machine learning can be the key to unlocking new insights through enabling innovative techniques. Computed tomography (CT) is one field within physics where the progress made in computer vision is easily applicable, and can be inherited and adapted effortlessly [63].

CT is a widely known 3D imaging technique that exploits the penetrating power of X-rays to sample 2D projections of an object from many azimuthal angles [22]. From the sampled projections, a 3D reconstruction of the object can be computed. This technique is highly applicable for nondestructive characterisation of material's internal structures, such as microporous structures imaged using micro-CT [1]. Furthermore, several CT reconstructions sampled in series can capture internal dynamics within an object, and is referred to as temporal CT, dynamical CT, or just 4D-CT because of the added time dimension. With this technique, experiments involving advanced dynamical processes can be performed, where the development of fluid phases can be tracked through time, and thereby provide insights to multiphase flow in porous media [1]. This is a field of research of particular interest to both the industry and academia. Equinor ASA, previously the Norwegian state oil company Statoil, has a keen interest in temporal CT, as insights from dynamical CT experiments can improve the oil recovery factor in their reservoirs. Moreover, just like computer vision has been directly applicable to CT, experiences from temporal CT techniques and dynamical experiments can be directly transferred to the low-carbon transition in an effort to accelerate the development of carbon capture and storage (CCS) technologies. CCS is a key component for achieving a carbon-neural industry, and studies of CO_2 encapsulated in porous media under high pressure are of particular importance for the technology [28]. However, the temporal CT technique is not without its challenges. In order to produce high-quality reconstructions, the number of projections must be high. Consequently, the scan time is long, and important dynamical processes are not captured due to a low temporal resolution.

In this master's thesis, a temporal CT technique based on golden ratio sampling and undersampled reconstruction enhancement is developed. The motivation for this work is to enable laboratory 4D-imaging with flexibility in the choice between temporal and spatial resolution, and to push the theoretical limit of undersampling. To elaborate, with the proposed scanning technique, one is able to find a suitable compromise between temporal and spatial resolution in the post-processing of the data, *a posteriori*. Then, the spatial resolution can be enhanced by applying a conditional generative adversarial network (cGAN) [36], which through supervised adversarial training [20] can account for streaking artefacts and noise associated with undersampled CT reconstructions. With the derived technique, a simple dynamical experiment of an hourglass is performed as an initial demonstration of application. For this sample, the achievable temporal resolution for the sample is determined, and the effect of

the GAN-enhancement is evaluated. Further assessment of the GAN is performed by evaluating its performance on a more complex sample, namely a cracked cylindrical sandstone. As this is the very beginning of the collaboration between Equinoir ASA and NTNU, several other tests are performed to increase the initial understanding of the derived technique. Experimentation with transfer learning, loss functions, different datasets, and different network architectures are performed for the purpose of optimisation.

Part I

REVIEW OF THE LITERATURE

1 COMPUTED TOMOGRAPHY

Computed Tomography (CT) is an X-ray imaging technique were a 3D model is reconstructed from a series of 2D projections. Small Angle X-ray Scattering Tensor Tomography (SAXSTT) is one type of CT, which was the topic of the project thesis preceding this master's thesis. Therefore, much of the literature review in this chapter is directly retrieved from the research conducted in the project thesis. Notices of this will also be made in sections where this is the case.

1.1 X-RAY RADIATION

X-ray radiation was thoroughly covered in the mentioned project thesis on SAXSTT, and is therefore included as is in this section. For a quantum mechanical introduction to the topic of X-ray absorption, see Appendix A. In the project thesis, X-rays were defined to be electromagnetic waves with energy in the orders of keV. Planck's Equation,

$$E = \hbar\omega = 2\pi\hbar\frac{c}{\lambda}, \tag{1.1}$$

relates the energy of a photon E to the angular frequency ω or wavelength λ of the corresponding electromagnetic wave. $c = 2.997\,92 \times 10^8$ m/s is the speed of light. The other constant is the reduced Planck's constant $\hbar = 1.0543 \times 10^{-34}$ Js [5]. Whence, X-rays typically have wavelengths in the sub-nanometer range.

Excitation and acceleration are the most common phenomena where X-ray radiation occurs. X-ray radiation from excitation, called *Characteristic X-ray radiation*, occurs when a highly energetic electron collides into a target atom. The accelerated electron transfers enough energy to eject an inner-shell electron with energy E_i, and an outer shell electron with energy E_f lowers its energy-state by filling the vacancy,

$$E_{photon} = -\Delta E = -(E_f - E_i). \tag{1.2}$$

Due to conservation of energy, this process causes emission of a photon, as described in Equation (1.2). As the atomic energy levels are discrete, this process gives rise to a spectrum of discrete X-ray emission lines.

Additionally, scattering events occur when electrons pass through an anode material. These events accelerate the electrons in new directions, and X-rays known as *Bremsstrahluhng* are emitted.

1.2 MATHEMATICAL DESCRIPTION

In a continuation of the previous section, the mathematical description of CT was also investigated in the project thesis, and is included in this section. The quantum mechanical description of X-ray absorption described in Appendix A was simplified

by the attenuation coefficient $\mu(\nu)$. The X-ray intensity I as a function of penetration depth is then given from Beer-Lambert's law,

$$I(s) = I(0)\exp(-\int_0^s \mu(\nu)d\nu), \tag{1.3}$$

where s is the thickness of the sample, and $I(0)$ is the initial intensity. The spectral dependence $\mu(\nu)$ is here neglected, as it often is, as the beam is assumed to be almost fully monochromatic [7]. This assumption results in beam hardening, which is an artefact that will be explained in below in Section 1.4. A simple manipulation of the expression gives the projection line integral

$$p(s) = -\ln(\frac{I(s)}{I(0)}) = \int_0^s \mu(\nu)d\nu. \tag{1.4}$$

The projection line integral in Equation (1.4) is closely related to the Radon transform of an object function $f(x, y)$ for a single orientation θ [62]. Confidence in this statement may be achieved by comparing Equation (1.4) with a single-angle Radon transform,

$$p_\theta(r) = \int_{-\infty}^{\infty} f(r, \nu)d\nu. \tag{1.5}$$

Collected Radon transform data is called a sinogram, and is used to reconstruct the original object function $f(x, y)$ [18].

An example of a sinogram, calculated for a cross section image of a walnut, is shown in Figure 1.1. The image was retrieved by performing CT on a walnut at the Technical University in Munich.

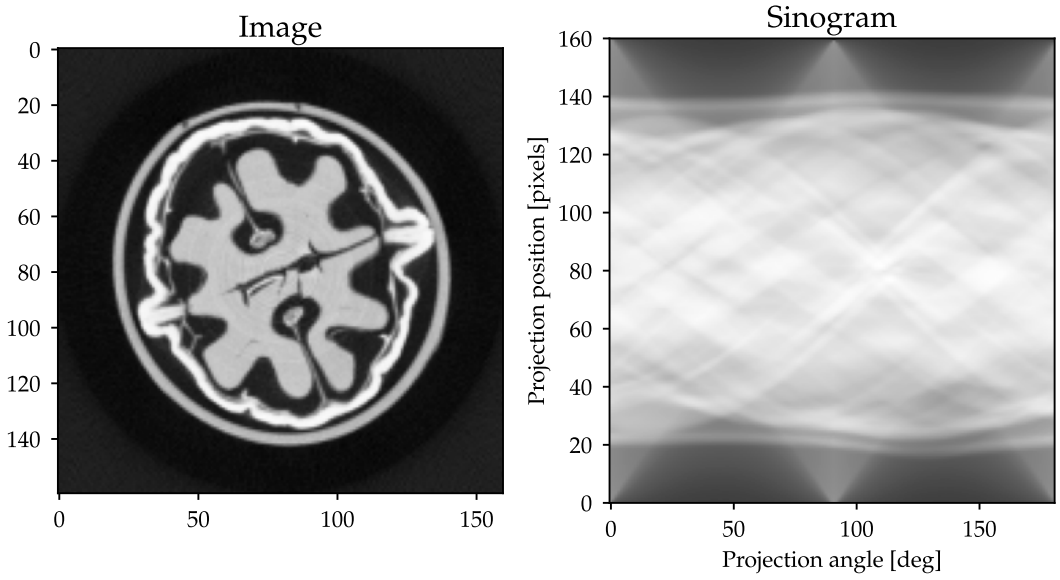


Figure 1.1: The cross section of a walnut, and the corresponding sinogram. The sinogram was calculated by performing the Radon transform for 360 different angles using the skimage-package in Python.

The key in computed tomography is to determine the spatial dependency of the attenuation coefficient $\mu(\mathbf{r})$. By sampling many projections, meaning line integrals

from different orientations, data necessary to reconstruct a three-dimensional image is collected. For a given cross section of the sample $f(x, y)$, the detected intensity is plotted as a function of projection number and detector number, thereby creating a sinogram. By utilising this sinogram and the Fourier slice theorem, the object $f(x, y)$ may be determined by other means than computing the full inverse Radon transform.

The Fourier slice theorem states that the Fourier transform of a projection is a slice of the 2D Fourier Transform of the region from which the projection was obtained [18]. Consequently, the full 2D Fourier transform $F(\omega_x, \omega_y)$ of an object $f(x, y)$ can be constructed from a series of 1D Fourier transforms $P(\omega)$ of projections $p(s)$ with different orientations [62]. In other words, the Fourier transform of a single projection makes up one angle of the full 2D Fourier transform,

$$P(\theta, \omega) = \mathcal{F}(\omega \cos\theta, -\omega \sin\theta). \tag{1.6}$$

With the fully 2D Fourier transform sampled, the filtered back projection (FBP) algorithm reconstructs the object by forward and inverse Fourier transforms. Firstly, a sinogram of projections is mapped to frequency space in polar coordinates by subsequent 1D Fourier transforms, mathematically described as:

$$P(\theta, \omega) = \int_{-\infty}^{\infty} p(\theta, r) e^{-2\pi i \omega r} dr. \tag{1.7}$$

With this the 2D Fourier transform $F(u, v)$ of the object $f(x, y)$ is found. The final step is an inverse 2D Fourier transform with, for instance, a *ramp*-filter of $|\omega|$ to account for the radial distribution of points in polar coordinates. This filter is also the Jacobian of the area integration element in the polar Fourier space. Consequently, the object function can be expressed as

$$f(x, y) = \int_0^{\pi} \int_{-\infty}^{\infty} |\omega| P(\theta, \omega) e^{-2\pi i \omega (x \cos\theta - y \sin\theta)} d\omega d\theta. \tag{1.8}$$

The Feldkamp-Davis-Kress (FDK) algorithm is a popular implementation of the FBP algorithm [15] that approximates the three-dimensional divergence of the X-ray source by a cone, as opposed to a fan, which only describes divergence in two dimensions.

1.3 SIMULTANEOUS ITERATIVE RECONSTRUCTION TECHNIQUE

The Simultaneous Iterative Reconstruction Technique (SIRT) is an iterative algorithm in the family of algebraic reconstruction techniques (ART) [29]. The algorithm is based on the idea of calculating the difference between measured and calculated projections, where an initial guess of the object function is used to calculate the projections. After having calculated the difference for all pixels in all projections, each cell is updated by the average calculated projection difference for the cell. With this, one iteration is completed, and the process is then repeated until convergence [29]

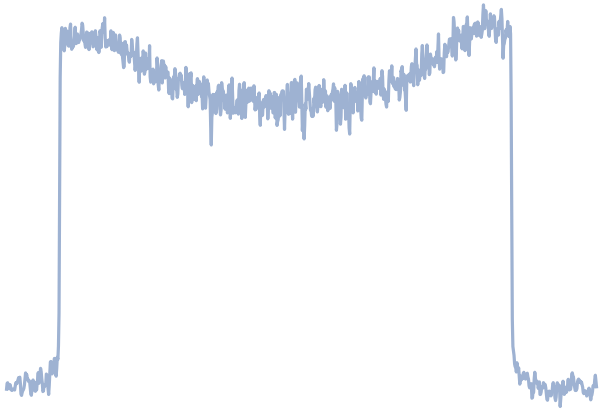


Figure 1.2: A one-dimensional cross section of a CT scan reveals cupping artefacts due to beam hardening. Because soft X-rays are attenuated faster than more energetic X-rays, absorption is lower in the centre of the sample, resulting in a cupping in reconstructed intensity.

1.4 CT ARTEFACTS

Artefacts are deviations from the theoretically expected behaviour of a CT system for projections or reconstructed images. One essential task in the field of CT is to develop methods to reduce the effects of artefacts. One of the most common artefacts is beam hardening, which is caused by an increase in mean photon energy as the soft X-rays are attenuated more than the more energetic radiation. Consequently, Beer-Lambert's law in Equation (1.3) is in fact a simplification of the real characteristics of the X-rays [8]. In addition to attentuation due to penetration depth, the intensity attenuation is also dependent on the energy of the X-rays, as derived in Equation (A.10) for the absorption cross section. Hence, there is a non-linearity between the actual X-ray attenuation and the measured intensity in the projections. The consequences of beam hardening are shown in Figure 1.2. The first consequence is a cupping artefact in the centre of the reconstruction, as the mean photon-energy has increased with the travelled path length, resulting in less attenuation in the centre [8]. Secondly, streak artefacts will occur between high-attenuating regions for the same reason as the cupping artefact.

Alisasing, or undersampling, artefacts refer to CT scanning that does not comply with the Nyquist-Shannon theorem [8]. The Nyquist-Shannon theorem states that the sampling frequency must be at least twice the highest frequency in the signal. Therefore, if not enough projections are taken, the Fourier slice theorem cannot be fulfilled satisfactory. This will result in stripes from the edges of high-attenuating regions in the reconstructed image. Additionally, there is also a practical understanding of the undersampling issue. If the desired 3D model has a cylindrical symmetry and a radius of $\frac{D}{2}$ voxels, then the circumference of the cylinder is πD voxels. With too sparse sampling, some voxels on the circumference will therefore not be sampled at all, resulting in a loss of information. A special case of undersampling artefacts is *missing wedge artefacts*, which occurs if only a limited angle interval is scanned. Such an artefact is shown in Figure 1.3, where the reconstruction is based upon only 37.5 % of the total revolution.

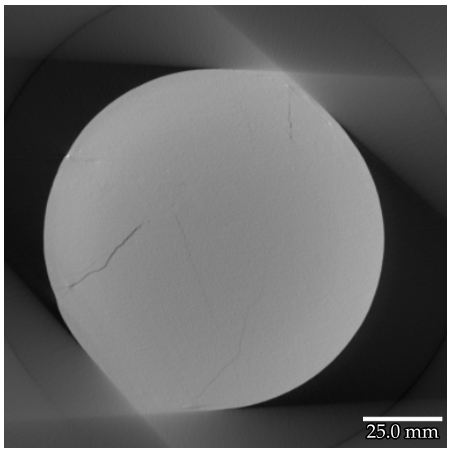


Figure 1.3: Missing wedge artefacts in a reconstruction because only 37.5 % of the total revolution was sampled. The full 2D Fourier transform is therefore not sampled, and the Fourier slice theorem (1.6) is not satisfied. In this example, the missing wedge results in two corners along the edge of the cross section due to the lacking information.

Blurring in the reconstruction, as seen in Figure 1.4, is oftentimes caused by having assumed the wrong axis of rotation (AoR). Even few pixels off the true AoR can cause severe blurring in the reconstructed image [61]. Therefore, a major part in the post-processing is to iteratively find the correct AoR by evaluating the sharpness of centre slice reconstructions for different AoRs. Note, however, that this is just one instance of a procedure to determine the correct AoR.

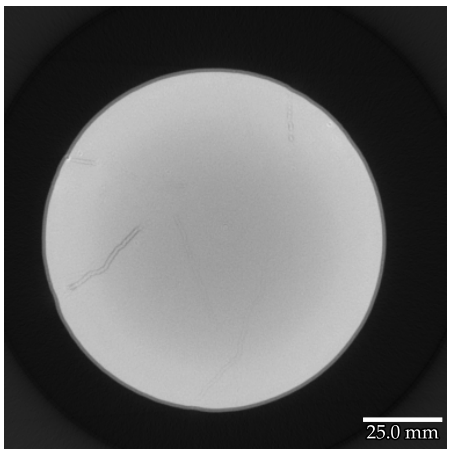


Figure 1.4: Misalignment of the axis of rotation causes blurring in the reconstructed image. As seen in the misaligned reconstruction of the sandstone, numerous double edges are present around details and at the bounds of the sample.

Finally, ring artefacts are highly recognisable defects caused by defect detector pixels. They can be either hot, meaning always on, or cold, meaning always off. Luckily, these artefacts are easy to remove if the defect pixels are known. In this case, applying a median filter to the projections at the defect pixel locations should be sufficient to remove the artefacts.

2 MACHINE LEARNING

2.1 INTRODUCTION

Artificial intelligence (AI) is the broadest term for replicating human intelligence in machines. The application of AI has exploded in recent years, and is now present in everyday life as well as in many scientific disciplines. Within AI, machine learning (ML) is another broad term for machines that recognise patterns in provided data. The data can typically be used for learning or optimisation, and it can come as raw or structured data. Moreover, the data can be labelled or unlabelled, meaning that the data is either already classified or not. Artificial neural networks (ANN) are most commonly instances of supervised machine learning, where labelled data is used to enable the network to perform a certain task. Training of networks with many layers is often referred to as deep learning (DL). Deep neural networks (DNN) are due to the complexity of the model able to recognise patterns in fully unstructured data without human intervention or preparation [14].

One of the most prominent fields of machine learning, and especially deep learning, is computer vision. Computer vision is the field of AI that enables machines to perform image classification, enhancement, segmentation etc.

The topics of gradient descent and automatic differentiation are two essential parts of machine learning that were the core of the research conducted in the preceding project thesis. Therefore, much of the theory regarding these subjects have been retrieved from the project thesis.

2.2 GRADIENT DESCENT

Gradient descent is an optimisation algorithm that updates the model's parameters based on the gradient of the cost function and the step size α [45]. The cost function $J(\theta)$ is the error of the model across the entire dataset, where the error function of a single observation is denoted the loss function $l_i(\theta)$. It is important to realise that even though the dataset is a sufficient representation of the true unknown distribution P, it is nevertheless only a subset of this distribution [6]. Ultimately, the calculated gradient $\nabla_\theta J(\theta)$ has a certain variance compared to $\nabla_\theta P(\theta)$. It is therefore said that machine learning algorithms minimise the *empirical risk* in an effort to minimise the generalisation error, or *risk* [19]. Nevertheless, a step of gradient descent consists of a forward pass where the empirical risk is calculated, followed by a backward pass where the gradient of the model is calculated. After each backward pass, the parameters of the model θ are updated according to the following equation:

$$\theta = \theta - \alpha \nabla_\theta J(\theta). \tag{2.1}$$

The idea of this algorithm is that by following the gradient of the cost function, the estimation will converge to the correct solution of the problem, which naturally

results in the minimal error. Gradient descent in parameter space is visualised in Figure 2.1.

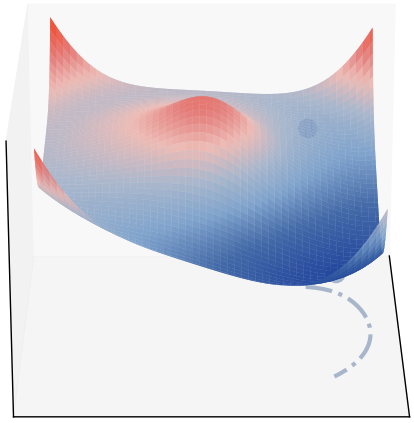


Figure 2.1: A surface plot of the error in parameter space for a model. The trajectory is drawn to show the path of parameter optimisation based on gradient descent. Notice that the smooth curve is an indication of many steps with a small step size. An animation of the process is available at https://github.com/RubenDragland/XRD_CT/tree/main/Plotting/animations. Both the animation and the plot were retrieved from the mentioned project thesis.

Oftentimes the optimisation task involves a massive model with millions of parameters trained on a huge dataset. In this case, calculating the gradient to the cost function across the entire dataset is a too computationally expensive task. Therefore, stochastic gradient descent (SGD) is usually employed instead. By calculating the gradient and performing an update of the model's parameters based on a single instance of data, often denoted online learning, the model usually learns much faster. This comes at the cost of noisy updates due to the high variance of the calculated gradients. However, the high variance can be a powerful tool in tackling local minima in parameter space, which improves the chances of convergence. Therefore, SGD can result in better solutions even with noisy updates [32]. With that being said, full batch learning is still employed when the conditions of convergence are required to be understood, and if it is beneficial to employ accelerated gradient descent techniques, such as conjugate gradient descent (CGD) or Quasi-Newton methods [47].

A compromise where the variance of the updates are tweaked by including several, but not all, instances of data per update is called minibatch learning. Consequently, the minibatch size is introduced as a hyperparameter. A hyperparameter is shortly explained the human-controlled parameters, or penalty factors, occurring in the model and during training [41]. Determining the optimal hyperparameters will enhance the performance of the model, and speed up the training.

Several other hyperparameters are introduced when determining the measures of loss and optimisers. There are numerous examples of loss measures. The mean

squared error (MSE) averages the Euclidean distance squared between the prediction $\widehat{Y}_i$ and the target Y_i as

$$\text{MSE} = \frac{1}{n}\sum_{i}^{n}\left(Y_i - \widehat{Y}_i\right)^2. \tag{2.2}$$

This loss measure is typically used in encoder decoder networks, image enhancement algorithms, segmentation algorithms, and generally if the target variable is numeric. The model targets are not necessarily numeric, but can be actual labels. In the case of two distinct categories, the optimisation task is a binary classifier [26], and the binary cross entropy loss is typically employed:

$$\text{BCE} = \frac{1}{n}\sum_{i}^{n}\left[Y_i \ln(\widehat{Y}_i) + (1 - Y_i)\ln(1 - \widehat{Y}_i)\right]. \tag{2.3}$$

In such a classifier, the target value Y_i is either 0 or 1, and the output of the model is necessarily a probability. To ensure the latter, the Sigmoid activation function,

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \tag{2.4}$$

can be exploited so that $\widehat{Y}_i' = \sigma(\widehat{Y}_i) \in (0, 1]$, and the resulting updated expression for the binary cross entropy loss becomes

$$\text{BCE}_{\sigma(x)} = \frac{1}{n}\sum_{i}^{n}\left[Y_i \ln(\sigma(\widehat{Y}_i)) + (1 - Y_i)\ln(1 - \sigma(\widehat{Y}_i))\right]. \tag{2.5}$$

Furthermore, during training, the parameter update does not need to be based entirely on the gradient of the cost function. By updating the parameters using *Momentum* [19], an exponentially decaying moving average of past gradients are used in addition to the current gradient to ensure accelerated learning. The technique is especially useful in the cases of noisy gradients. As mentioned, this is the case in online learning, as the gradient is calculated based on a single instance of data, and not an average of the entire dataset. To ensure a more consistent approach towards convergence, momentum is used to enhance the direction shared by several consecutive gradients. Correspondingly, momentum will exploit high curvature in parameter space and series of small but consistent gradients [19]. When introducing momentum to the optimisation step, the updated gradient descent rule becomes

$$\begin{aligned} \mathbf{v} &= \beta\mathbf{v} - \alpha\nabla_\theta J(\theta) \\ \theta &= \theta + \mathbf{v}, \end{aligned} \tag{2.6}$$

where β is the momentum hyperparameter, and α is the learning rate.

In addition to tweaking the direction of the step, the size of the step can be adjusted. Here, *RMSProp* [19] is a well-known technique for adjusting the learning rate by an exponentially decaying average of the squared gradients. Formally, RMSProp is implemented as

$$\begin{aligned} \mathbf{g} &= \nabla_\theta J(\theta) \\ \mathbf{r} &= \rho \mathbf{r} + (1-\rho)\mathbf{g} \odot \mathbf{g} \\ \theta &= \theta - \alpha \frac{\mathbf{g}}{\sqrt{\mathbf{r}+\delta}}, \end{aligned} \tag{2.7}$$

where ρ is the decay rate of the moving average, δ prevents zero-division, and $\odot$ is the element-wise multiplication operator. Note that element-wise division is also performed in the last line.

Finally, *ADAM*, short for adaptive moments, is an example of an optimiser that employs more than one of the above techniques. Specifically, ADAM is a combination of RMSProp and momentum. Consequently, it is one of the most robust and popular optimisers [19]. From the gradient calculation, momentum is calculated in the first moment of the gradients, and RMSProp is calculated in the second moment of the gradients. In contrast to the original momentum and RMSProp algorithms, ADAM also applies a correction bias to the first and second moments. The ADAM algorithm can be summarised as

$$\begin{aligned} \mathbf{g} &= \nabla_\theta J(\theta) \\ \mathbf{v} &= \beta \mathbf{v} - (1-\beta)\mathbf{g} \\ \mathbf{r} &= \rho \mathbf{r} + (1-\rho)\mathbf{g} \odot \mathbf{g} \\ \mathbf{v}' &= \frac{\mathbf{v}}{1-\beta^t} \\ \mathbf{r}' &= \frac{\mathbf{r}}{1-\rho^t} \\ \theta &= \theta - \alpha \frac{\mathbf{v}'}{\sqrt{\mathbf{r}'}+\delta}, \end{aligned} \tag{2.8}$$

where t is the time step, and β and ρ are the decay rates of the moving averages.

2.3 AUTOMATIC DIFFERENTIATION

As mentioned, automatic differentiation (AD) was the main focus of the project thesis, and a detailed review of the technique from the project thesis is therefore provided in this section. It is an algorithmic technique for computing the analytical gradients of a function using computational graphs and the chain rule. It is important to contrast AD from numerical differentiation using finite differences, which comes with substantial numerical error. Moreover, AD should not be confused with symbolic differentiation, which is a method for calculating the full symbolic expression for the gradient, like one would do by hand [2]. Most often, AD is implemented using an object-oriented operator overloading approach. In other words, the AD object inserts the operators from the cost function to a computational graph. Moreover, the rules of differentiation for the operators are pre-implemented. Therefore, it is required that the evaluated function only consists of operators that are supported by the AD object. With the function evaluation completed, the backward pass is initiated. Let $w(y(x))$ be a functional depending on the function $y(x)$, which is again a function of the input variable x. Applying the chain rule,

$$\frac{\partial w}{\partial x} = \frac{\partial w}{\partial y}\frac{\partial y}{\partial x}, \tag{2.9}$$

the gradients with respect to the input variables are calculated given the output value. Equation (2.9) shows how the gradient of any complex computational graph can be written as a product series of basic gradient expressions. The backpropagation algorithm recursively applies the chain rule on the computational graph, eventually ending up with the gradients of the input variables [2]. Figure 2.2 is such an instance of a computational graph, with the chain rule applied, to the function $w(x, y) = \exp(xy)$.

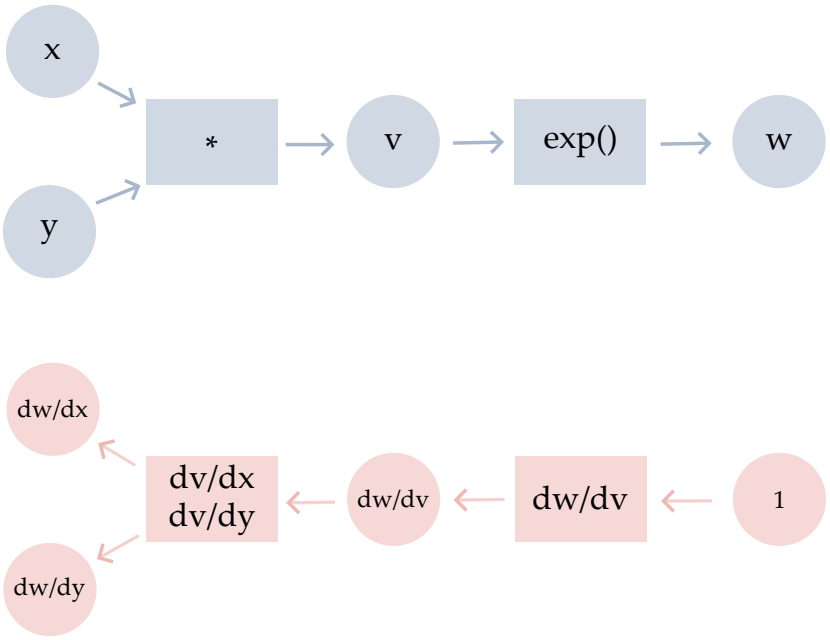


Figure 2.2: A computational graph for the function $w(x, y) = \exp(xy)$ on the top. Boxes represent operators, and the circles represent the input or output variables. The chain rule is applied to the computational graph in the bottom figure. Here, circles are the currently calculated gradient, and the boxes are the basic gradient expressions which are multiplied to the current gradient. The same figure was used in the mentioned project thesis preceding this master's thesis.

To exemplify the application of AD in the training of neural networks in a very simplified manner using Figure 2.2, one could consider w to be the loss of the neural network. Analogously, x and y would be the network weights. During the forward pass, the network evaluates the input data, from which the loss is calculated. Simultaneously, the mentioned computational graph, which tracks all operations performed on the input data, is constructed. In order to optimise the network, the weights are updated according to the rules of gradient descent (2.1). For this purpose, backpropagation is applied to the computational graph, which results in the gradient of each network weight being derived from the chain rule (2.9).

2.4 PREPARATION OF DATA

Generalisation, meaning the performance on unobserved data, is a central problem within the field of machine learning. When minimising the empirical risk based on

training data, so-called overfitting will eventually occur. Overfitting is a phenomenon where the training error decreases without the validation error, which is the predicted error on unseen data during training. The counterpart to overfitting is underfitting, which is a model's inability to learn the underlying pattern in the training data altogether. In addition to the validation error, there exists a test error, also denoted *generalisation error*, which is the error retrieved from testing the model on unseen data after having finished hyperparameter-tuning and training [19].

Even though the term is called generalisation, the model cannot be expected to perform well on all possible data. The most essential constraint for generalisation is that all data must be independent and identically distributed (i.i.d.) [38]. In other words, the data must be independently drawn from the same probability distribution. This is because the model is trained to learn this underlying distribution.

There are numerous tricks to improve the generalisation of the model, called regularisation techniques. Shortly summarised, they include, but are not limited to, dropout, early stopping, weight decay, and L2 regularisation [37]. Dropout is a technique where a random subset of weights is ignored ignored during training. Early stopping is to abort the training process when the validation error starts to increase, which is a sign of overfitting. The aim of weight decay is to penalise large weights, which is a common feature of an overfitting model. One instance of weight decay is L2 regularisation, which is to add a constant times the sum of the squared weights to the loss function.

However, the most common and overall efficient technique for generalisation regardless of model architecture is data augmentation. Data augmentation is to artificially alter the training data by applying random transformations. The most common transformations are rotations, translations, inversion, gaussian noise, and scaling. Consequently, the effective size of the training data is increased, which in turn improves the generalisation of the model, because the model has not the ability to memorise the now greatly expanded training dataset [19]. At the same time, a data augmented dataset is more robust to noise and outliers, which is a common feature of real-world data. In this way, augmented datasets could be considered more representative of real-world data.

2.5 ARTIFICIAL NEURAL NETWORKS

An artificial neural network (ANN) is built up of many artificial neurons, just like the biological neural network in the human brain is built up of neurons. In this way, the essential building block of any ANN is the artificial neuron, a mathematical equation that, for a given number of inputs, applies its *weights* to produce an activation, which in turn is sent through a non-linear activation function to produce an output [4]. Matrices and vector notation may be applied to provide the most concise mathematical description of the feed forward process of an artificial layer consisting of N input nodes and M output nodes,

$$\begin{aligned} \mathbf{a} &= W\mathbf{x} \\ \mathbf{z} &= f(\mathbf{a}). \end{aligned} \tag{2.10}$$

Equation (2.10) describes first the calculation of the activation of the layer through a matrix-vector product, a linear transformation, and then the application of a non-linear activation function, f. Further illustration is provided in Figure 2.3. Due to the

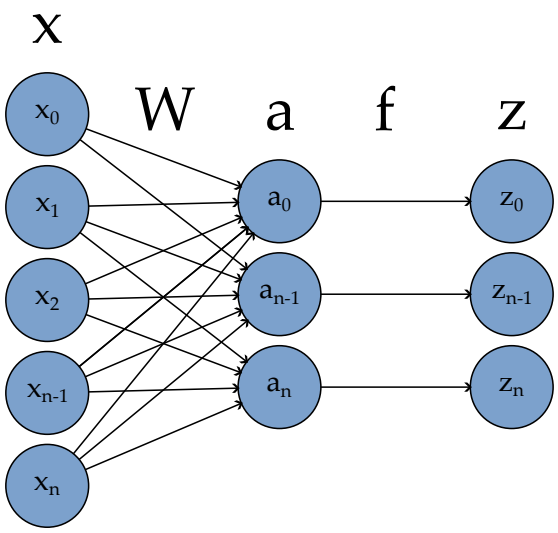


Figure 2.3: The composition of a linear layer. x is the input vector, which by matrix-vector multiplication with the weight matrix W results in the activation vector a. By applying the non-linear activation function f, the output vector z is produced.

non-linearity of the activation function, addition of a series of so-called *hidden layers* between the input and output layers increases the complexity of the model, which is the key to the abilities of deep neural network, which are ANNs with many hidden layers.

The layer described in Equation (2.10) is a linear layer, also called a fully connected (FC) layer. One issue with linear layers is the high number of trainable parameters, which require both a large amount of memory and computational power to train. Moreover, the risks of overfitting increase when a neural net has overparametrised layers, meaning that it has the capacity to memorise the training data. Therefore, convolutional layers revolutionised the world of machine learning, especially in the field of computer vision. A convolutional layer applies the same *convolutional kernal* to the input data, and explots the fact that nearby pixels are strongly correlated [4]. Consequently, the number of parameters is greatly reduced and is constant regardless of input size, and the model generalises more efficiently. Moreover, the model automatically becomes translation invariant by applying convolutional layers instead of linear layers.

As a fundament, the continuous convolution operation is defined as

$$f * g = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau, \tag{2.11}$$

where f and g are two functions, and $*$ is the convolution operator. In the world of computer vision, two adjustments are usually assumed for convenience when describing convolutional kernels applied to input data.

Firstly, the discrete convolution operation is used instead of the continuous convolution operation, since the input data is a discrete number of features. Therefore, Equation (2.11) is adjusted to consist of a discrete sum instead of an integral:

$$f * g = \sum_{\tau=-\infty}^{\infty} f(\tau)g(t-\tau). \tag{2.12}$$

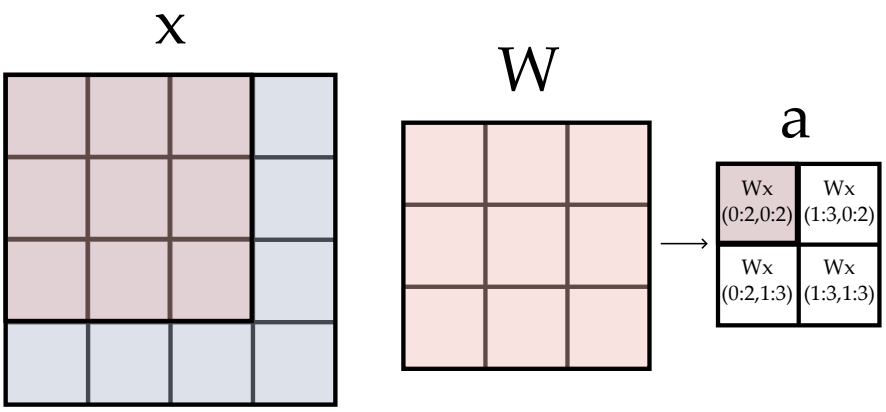


Figure 2.4: Convolution of a 2D image with a 2D kernel. The topleft pixel of the 2×2 output is effectively the dot product of the 3×3 kernel and the 3×3 topleft corner of the input image. Correspondingly, the bottomright pixel of the output is the dot product of the kernel and the bottomright corner of the input image. Here, a stride of 1 is assumed.

Secondly, the cross-correlation operation is used instead of the convolution operation, which is defined as

$$f \star g = \sum_{\tau=-\infty}^{\infty} \overline{f(\tau)} g(t+\tau), \tag{2.13}$$

where $\star$ is the cross-correlation operator. Specifically, cross-correlation in three dimensions becomes

$$f \star g = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} \sum_{\tau_3=-\infty}^{\infty} \overline{f(\tau_1, \tau_2, \tau_3)} g(t_1+\tau_1, t_2+\tau_2, t_3+\tau_3), \tag{2.14}$$

Practical illustrations of Equation (2.12), (2.13), and generally convolutional layers are provided in Figure 2.4. For each output pixel, a dot product is performed between the kernel and a corresponding region of the input image.

As indicated in Equation (2.14), dimensionality is a central feature of convolutional layers. The input data has its spatial dimensions in addition to a number of channels, exemplified by the RGB colour channels, for instance. Consequently, the convolutional kernel has the same number of channels as the input data, the same number of spatial dimensions, and a kernel width of 3, 5 or 7 in each spatial dimension. However, this would result in a single output channel, which is not always desirable. Instead, the entire kernel of a convolutional layer consists of a series of mentioned convolutional kernels, each responsible for a single output channel. To summarise, the input data has typical shape $(N, C_{in}, \mathbf{S})$, where N is the number of samples in the batch, C_{in} is the number of input channels, and $\mathbf{S}$ represents the spatial dimensions. A kernel of shape $(C_{out}, C_{in}, \mathbf{F})$ is applied to the input data, resulting in an output of shape $(N, C_{out}, \mathbf{S})$. Figure 2.5 illustrates a convolutional layer in the context of a neural network.

It is worth realising that the many channels produced from many convolutional layers in series are the key to the success for convolutional neural networks (CNNs) [21]. Each convolutional kernel works as a filter, and performs a specific feature extraction, for instance edge detection. This is crucial for many appplications in computer vision.

However, an increasing number of channels also increases the required computational power and memory. Fortunately, the solution is simply to reduce the spatial

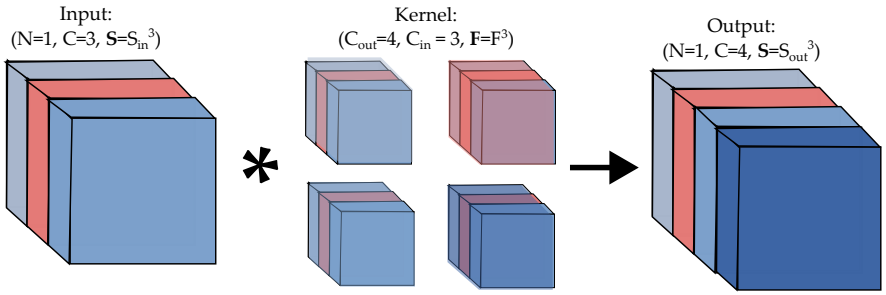


Figure 2.5: A convolutional layer with input consisting of a single sample, batch size $N = 1$, three channels C, and three spatial dimensions. The layer itself consists of four convolutional kernels, each responsible for a single output channel. Each kernel has also three spatial dimensions F, and the same number of channels as the input data $C_{in} = C$. Because there are four kernels, the output has four channels $C_{out} = 4$.

dimensions of the data as it passes through the network, while the number of feature channels increases. There are generally two ways to do this: *pooling* and *strided convolutions*. Pooling is a non-linear operation that simply reduces the data of a given region to a single value, for instance the maximum value or the average value. With the 2×2 *Max Pooling* operation, the most dominant value is selected in a region where every spatial dimension has size 2, and because the convolutional stride is also 2, the size of each spatial axis is reduced by a factor of 2. Similar to pooling, a reduction in the spatial dimensions can also be achieved if the convolutional layers are not padded and the convolutional stride is greater than 1. This phenomenon is summarised in the following expression:

$$S_{out} = \frac{S_{in} - F + 2P}{S} + 1, \tag{2.15}$$

where S_{out} and S_{in} are the spatial dimensions of the output and input data, respectively, F is the spatial size of the convolutional kernel, P is the padding, and S is the convolutional stride.

As a result, a common network architecture for CNNs is the *U-Net* [44], which in 2015 was a new convolutional neural network architecture proposed to perform biomedical image segmentation. A U-net consists of a feature extraction, where the spatial dimensions are reduced while the number of channels is increased, and a fusion, where upsampling is applied at the same time as channels are merged [44]. The crucial component of the U-net architecture is the skip connection. During upsampling, stored data from the feature extraction is concatenated with the output data from the upsampling layer, and the result is passed on to the next convolutional layer. Consequently, the convolutional layers are not required to store every feature, but may instead focus on the features of interest that minimise the loss function.

2.6 GENERATIVE ADVERSARIAL NETWORKS

A generative adversarial network (GAN) is a network architecture consisting of two opposing neural networks, a generator and a discriminator. It was first proposed by Ian Goodfellow in 2014 [20] with the goal of creating a zero-sum game between the generator and the discriminator, where the generator attempts to generate realistic samples, and the discriminator attempts to distinguish between real and fake sam-

ples. This idea is theoretically based upon the Nash equilibrium in game theory [55]. The generator, which is the model of interest for inference, is trained to learn the probility distribution of the training data, because the discriminator will not be able to distinguish between real and fake samples if the probability distribution of the generator output is equal to the probability distribution of the training data. This procedure may be described by a min-max game with value function $V(D, G)$ [36],

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))], \tag{2.16}$$

where D is the discriminator, G is the generator, $p_{data}(x)$ is the probability distribution of the training data, and $p_z(z)$ is the probability distribution of the latent space. It is important to notice that the aim of the generator is to generate data from noise.

However, a GAN may also be conditional, where the generator is trained to generate data conditioned on a given input, called a conditional GAN (cGAN) [36]. An example of a cGAN may be a generator that transforms a low-resolution input image to a high-resolution output image. If generated sucessfully, the discriminator would not be able to distinguish between the generated and real high-resolution images.

Even though GANs have been successful in generating realistic images, they are also known to be difficult to train [20]. Since the GAN consists of two neural networks that are trained simultaneously, the strength relation between the networks is crucial. The discriminator has a tendency to quickly learn the key differences between the real and generated samples, and will therefore be increasingly certain in its predictions. This certainty is a cause for vanishing gradients in the adversarial loss function, which in turn causes the generator to be unable to learn. The root cause for this tendency is related to the fact that the discriminator task, binary classification, is substantially easier than the generator task, which is generation. Therefore, the generator networks needs to include more layers and more parameters than the discriminator network. At the same time, the more advanced the discriminator network is, the more essential features of the real samples it will learn, which in turn would enhance the generator training. In instances with a too weak discriminator, it would simply guess randomly between real and fake samples, providing no useful information to the generator. There exist numerous techniques to stabilise GAN training. Through trial and error, a combination of, for instance, adjusting the generator and discriminator learning rates, the number of discriminator updates per generator update, the generator loss function, and the minibatch size, may prevail.

2.7 TRANSFER LEARNING

A final important concept in machine learning is transfer learning. This technique allows for exploiting the knowledge of a pre-trained model for a different task [58]. One common example is to freeze the feature extraction weights of a deep convolutional net, and then only train the last layers of the network so that it can be used for your designated task. Not only does this allow for a faster training time, it is also probable that better results are achieved with pre-trained weights, because the pre-training has been performed on an immense scale, with a gigantic dataset on several powerful machines. Moreover, the pre-training and information sharing is also more sustainable in terms of energy consumption. One example of a pre-trained model is *ResNet* [25]. This net also revolutionised the field of deep neural networks because of its use of

residual connections, meaning to add the input of a convolutional layer to the output of the same convolutional layer, thereby limiting the vanishing gradient problem. Other instances of transfer learning include the so-called perceptual loss term [33]. Instead of depending entirely on a pixel-wise loss term between target data and output data, it is possible to first apply a pre-trained feature extractor, for instance ResNet [25], to both the target and output data. Then, the loss term is based on the MSE loss between the feature maps of the target and output data. In this way, one can force the model to learn what is perceived as important features, rather than only evaluating the independent pixel-wise loss. One disadvantage with perceptual loss, however, is the increased computational cost, since these pre-trained deep neural networks consist of many layers with many parameters. The resulting computational graph when utilising perceptual loss would therefore be significantly extended, increasing both computational time and memory consumption.

3 TEMPORAL CT

3.1 INTRODUCTION TO TEMPORAL CT

Temporal CT is a technique where a dynamic process is studied using a CT scan. It combines the advantages of CT, specifically the ability to investigate entire 3D volumes of a bulk object, with the ability to study the evolution of a process in time. One example of the application of temporal CT is time resolved internal processes in the human body, such as the heart [54]. Another example is to study the flow of different fluids in porous media [1], which is the designated application for the work performed in this master's thesis.

Even though the abilities of temporal CT are advantagous, the execution of the technique is easier said than done. The main issue is sufficiently efficient data acquisition to obtain satisfactory spatial and temporal resolution. A fully sampled CT scan will typically consist of 1000 projections for a volume of $512\times512\times512$ voxels. The scan time for this number of projections is typically 30 minutes. Any dynamic process with a typical timescale of less than the scan time will provide dynamic sampling artefacts. Alternatively, one could undersample each CT scan by reconstructing based on, for instance, the first 300 projections. With this approach, the temporal resolution is improved, but the spatial resolution is reduced due to undersampling artefacts and missing wedge artefacts, as referred to in Section 1.4. To solve the issue of missing wedge artefacts, one could perform several full revolutions of the sample, with a smaller number of projections per revolution. However, also this approach comes with disadvantages. Firstly, one is forced to decide on the number of projections per revolution, which ultimately is compromising between spatial and temporal resolution before even starting the dynamical experiment. Secondly, one will necessarily sample the same projections for each revolution, and thus only provide new information in the time domain, not the spatial domain. Thirdly, the issue of undersampling has with this solution not been adressed.

The following sections will therefore discuss a more efficient and innovative approach to temporal CT, which is based on golden ration sampling and machine learning.

3.2 UNDERSAMPLED CT RECONSTRUCTIONS

An important aspect of temporal CT is the exploitation of undersampled CT reconstructions, or sparse-view reconstructions, to improve temporal resolution. From this aspect, two terms are necessary to define: the necessary number of projections for full sampling, and the undersampling factor. Full sampling is defined as the number of projections P necessary to achieve a sampling frequency of twice the maximal signal frequency [13]. The circumference of an assumed cylindrical sample is given by πD, where D is the diameter of the scanned object with number of pixels as unit. The

number of pixels is given by the field of view divided by the pixel size. Anyway, the number of projections necessary for full sampling is thus defined as

$$P = \frac{\pi D}{2}, \tag{3.1}$$

with the factor $\frac{1}{2}$ to exploit the fact that each projection intersects the circumference twice.

From Equation (3.1), one can define the undersampling factor as the ratio between the sampled number of projections and the number of projections for full sampling P. One alternative to this definition, which will be denoted the *relative undersampling factor*, is the ratio between the number of projections in the undersampled reconstruction and the number of projections used in the designated ground truth. Hence, this ratio can be either larger or smaller than P, depending on the number of projections used in the reference reconstruction. Both definitions are useful in different contexts, and will be referred to as undersampling factor and relative undersampling factor, respectively.

3.3 GOLDEN RATIO SAMPLING

The golden ratio is one of the most mysterious and interesting ratios in mathematics. It is related to the Fibonacci numbers together with all objects with five-fold symmetry, and it can be shown that the golden ratio is the most irrational among irrational numbers [12]. More importantly, it has been shown that the ratio is central for optimisation within the field of biology, computer science, and physics. The most prominent example is the distribution of leaves in a sunflower [12].

Even though the golden ratio seems mysterious, it appears from a simple geometric construction [12]. Suppose that a line is divided into two parts as visualised in Figure 3.1.

Figure 3.1: A line segment divided by the golden ratio. The total line length divided by the long part is equal to the long part divided by the short part. Both ratios are equal to the golden ratio ϕ.

The long part has length a, and the short part has length b. The respective lengths of the lines are such that

$$\phi = \frac{a+b}{a} = \frac{a}{b}, \tag{3.2}$$

from which the golden ratio ϕ can be defined. Assume further that a has unit length, while the entire line has length x. The updated equation for the golden ratio becomes

$$\phi = x = \frac{1}{x-1}. \tag{3.3}$$

Written as a quadratic equation, the golden ratio can be derived by solving

$$x^2 - x - 1 = 0, \tag{3.4}$$

for which

$$\begin{aligned} x_1 &= \frac{1+\sqrt{5}}{2} = 1.618... \\ x_2 &= \frac{1-\sqrt{5}}{2} = -0.618... \end{aligned} \tag{3.5}$$

are the solutions.

The next clue in understanding the properties of the golden ratio lies in investigating the Fibonacci numbers, which was performed in [12]. The Fibonacci numbers are an example of an additive sequence of integers, where each number is the sum of the two preceding numbers. This sequence can be expressed as

$$\begin{aligned} F_0 &= 0 \\ F_1 &= 1 \\ F_n &= F_{n-1} + F_{n-2}. \end{aligned} \tag{3.6}$$

In other words, the recursion relation of the Fibonacci numbers is given by

$$A_{n+2} = A_{n+1} + A_n, \tag{3.7}$$

where A_n is the nth. Fibonacci number. Suppose now that for a sufficiently high n, the Fibonacci numbers are also a geometric sequence, meaning that the recursion relation

$$A_{n+1} = \alpha A_n, \tag{3.8}$$

also holds. These two constraints provide the Fibonacci quadratic equation:

$$\alpha^2 A_n = \alpha A_n + A_n \implies \alpha^2 - \alpha - 1 = 0, \tag{3.9}$$

which is equivalent to the first derivation of the golden ratio. As a matter of fact, if one divides increasingly large Fibonacci numbers by their preceding Fibonacci numbers, the ratio converges to the golden ratio [34], for instance $987/610 = 1.618033$.

Moreover, some manipulation of the quadratic equation may yield the golden ratio as continued root operations:

$$\begin{aligned} \phi &= \sqrt{1+\phi} \\ \phi &= \sqrt{1+\sqrt{1+\phi}} \\ \phi &= \sqrt{1+\sqrt{1+\sqrt{1+...}}} \end{aligned} \tag{3.10}$$

Additionally, by dividing by α in Equation (3.9), it is possible to write the golden ratio as the simplest continued fraction in existence [34]:

$$\begin{aligned} \phi &= 1 + \frac{1}{\phi} \\ \phi &= \frac{1}{1+\frac{1}{1+\frac{1}{1+\frac{1}{1+...}}}}. \end{aligned} \tag{3.11}$$

The fact that this continued fraction is only composed of ones ensures that it converges very slowly. According to [34], the golden ratio can therefore be considered the most irrational among irrational numbers.

With sufficient investigation of the properties associated with the golden ratio, it is time to investigate the unique opportunities of golden ratio sampling. Briefly explained, golden ratio sampling is to sample points along a line by splitting the longest segment by the golden ratio. Mathematically, this is expressed as

$$x_n = \mod\left[\frac{l}{\phi}n, l\right], \tag{3.12}$$

where x_n is the nth sample point, l is the length of the line, and ϕ is the golden ratio.

Golden angle sampling may be considered a special case of golden ratio sampling, where points along the circumference of a circle is sampled. Firstly, the golden angle is defined as,

$$\varphi = \min\left[\frac{360}{\phi}, 360 - \frac{360}{\phi}\right] = 137.5^\circ, \tag{3.13}$$

thereby being the acute angle when dividing a circle into two parts with the golden ratio. By sampling point n given by

$$\theta_n = \mod\left[\frac{360}{\phi}n, 360\right], \tag{3.14}$$

the sampled point on the circle will always split the longest segment by the golden ratio [16]. Moreover, the distribution of points along the circumference is roughly uniform from the beginning, and becomes increasingly uniform as the sample size increases, not to mention the fact that each point is uniquely sampled only once, as the golden ratio is irrational [16]. Interestingly, the uniformity peaks every time the number of points sampled is equal to a Fibonacci number.

3.4 MACHINE LEARNING APPLIED IN COMPUTED TOMOGRAPHY

CT is one of the most promising fields in physics where machine learning can be applied, due to its close resemblance to the highly developed field of computer vision [63]. Reconstruction of 3D-models, segmentation, classification, and enhancement are all examples of tasks that have been perfected by the development of computer vision, and which are very relevant for CT. This section will shortly summarise examples of machine learning applied in CT by referring to the literature on the topic.

Firstly, the reconstruction algorithm itself can be replicated using machine learning. This was proven in the article "*Tomographic reconstruction with a generative adversial network*" [60], where a generative adversarial network (GAN) called GANrec was self-trained to perform the inverse radon transform directly. Analytical reconstruction algorithms cannot reconstruct satisfactory 3D-models in cases with missing wedge artefacts, because entire sections of the object have not been sampled. Subsequently, entire sections of the object in Fourier space would be missing [60], and the condition for FBP, The Fourier slice theorem, would not be satisfied. Instead, GANrec could compensate for missing information because it was an optimisation problem that assumed the existence of a global minimum. This is an instance of a so-called End-to-End approach with some degree of physics-based prior knowledge, because

it compared the measured and projected sinograms to assess the error of the reconstruction [22]. It is important to realise that the GANrec was self-trained, and did not require training data. At the same time, it required a large amount of computational power for each reconstruction, unlike the typical applications of supervised GANs.

Alternatively, one can split the approach into two steps. The first step is a physics-assisted reconstruction algorigthm, while the second step is an image enhancement. This was performed in "*Physics-assisted generative adversarial network for X-ray tomography*" [22]. It employed an advanced physics-assisted maximum likelihood estimation (MLE) to reconstruct the 3D-model, and further improved the reconstruction using a conditional generative adversarial network (cGAN) called PGAN. The main purpose of the PGAN was to remove undersampling artefacts and other forms of noise, and supervised learning using simulated data was applied to train the network, which will be further explained in Section 3.5.

Finally, supervised learning can also be applied to tasks such as segmentation and super-resolution. The latter is exemplified in the article "*CT-image of rock samples super resolution using 3D convolutional neural network*" [56], where a 3D-convolutional neural network (CNN) was trained to increase the resolution of CT-images of rock samples. High quality CT reconstructions were downsampled and interpolated to simulate low-resolution images, for which the CNN was trained to undo the transformations. Numerous other examples of machine learning applied in CT could be listed, but the most relevant to the topic of this thesis are the aforementioned examples.

One instance that goes beyong the scope of convolutional neural networks is the article *Dual-domain sparse-view CT reconstruction with Transformers* [48]. Their goal was to find an alternative to the artefacts caused by filtered back projection and the time consumption caused by iterative reconstruction algorithms in sparse-view CT. The proposed solution was called CTTR, Computed Tomography Transformer, which combines convolutional neural networks with the transformer architecture, and exploits information provided from both undersampled filtered backprojection and the sinogram itself. To elaborate, the reconstructed slice was processed by convolutional layers before being flattened and sent to the transformer encoder. The task of the encoder was to learn what parts of its input are important. Meanwhile, the decoder received the *encodings*, and utilised them to reconstruct the slice using the sinogram as input. This mechanism, where each input to the decoder is weighed based on the input from the encoder, is called *attention* [52]. The mechanism of attention addresses the one major disadvantage of convolutional neural networks, which have no way of weighting the importance of different parts of the input.

3.5 RECONSTRUCTION ENHANCEMENT

From the mentioned applications of machine learning in CT, supervised learning for reconstruction enhancement is the most relevant to the topic of this thesis. Therefore, it will be the focus of this section, while using the article "*TomoGAN: Low-Dose Synchtrotron X-ray Tomography with Generative Adversarial Networks*" [33] as a reference. This same reference was used to form the basic architecture of the GAN used in this master's thesis, and the distinctions between the models will be specified and explained in the following sections. Nevertheless, the basic idea of the reconstruction enhancement is the same in both cases. A two-step approach with separate recon-

struction and enhancement is used. Consequently, the algorithm is more flexible than the GANrec, because it is possible to improve either the reconstruction or the enhancement separately. Secondly, the reconstruction enhancement is performed by a conditional GAN, which is as mentioned trained using supervised learning. In this way, the reconstruction enhancement can be performed on any CT-image, as long as the training data is representative of the data to be enhanced. The typical artefacts that are removed by the reconstruction enhancement are undersampling artefacts, high-frequent noise, and streaking artefacts. Transfer learning was also employed, where a deep classification network was used to extract features from the CT-images in order to assess whether all features had been maintained after enhancement [25].

3.6 QUANTIFYING RECONSTRUCTION ENHANCEMENT

One common means of quantifying the quality of a reconstruction is the peak signal to roise ratio (PSNR). It is defined as

$$\mathrm{PSNR}(I, J) = 10 \log_{10} \left(\frac{\max(I)^2}{\mathrm{MSE}(I, J)} \right), \tag{3.15}$$

where $\max(I)$ is the maximum possible pixel value, typically 255 or 1, and $\mathrm{MSE}(I, J)$ is the mean squared error between the two images I and J.

Additionally, structural similarity index measure (SSIM) is common metric to quantify the similarity between two images [49]. For two corresponding images I and J, the global SSIM is defined as

$$\mathrm{SSIM}(I, J) = \frac{(2\mu_I\mu_J + c_1)(2\sigma_{IJ} + c_2)}{(\mu_I^2 + \mu_J^2 + c_1)(\sigma_I^2 + \sigma_J^2 + c_2)}, \tag{3.16}$$

where μ_I and μ_J are the mean values of I and J, respectively, while σ_I and σ_J are the standard deviations, and σ_{IJ} is the covariance between I and J. The constants c_1 and c_2 are chosen to stabilise division, and are related to the dynamic range of the images. Therefore, the images are usually rescaled to the range $[0, 1]$ before calculating the SSIM. With this normalisation, 0.01^2 and 0.03^2 for c_1 and c_2, respectively, was used in [49].

The expression in (3.16) is derived from weighting the image qualities of luminance, contrast, and structure [27]. These qualities are defined as

$$\begin{aligned} l(I, J) &= \frac{2\mu_I\mu_J + c_1}{\mu_I^2 + \mu_J^2 + c_1'}, \\ c(I, J) &= \frac{2\sigma_I\sigma J + c_2}{\sigma_I^2 + \sigma_J^2 + c_2'}, \\ s(I, J) &= \frac{\sigma_{IJ} + c_3}{\sigma_I\sigma_J + c_3'}. \end{aligned} \tag{3.17}$$

For simplicity, the third constant c_3 is set to $c_3 = \frac{c_2}{2}$ [57]. As shown, for luminance one compares the mean of the images, while the standard deviation is compared to evaluate the contrast. Finally, the covariance of the image is related to the structural similarity.

However, the global SSIM score can in some cases be misleading, for instance if most of the background is black, while fine details in regions of interest are distorted. In such cases, the global SSIM score will be high, while the image quality is perceived by humans to be poor. Therefore, the mean structural similarity index measure (MSSIM) is often used instead. It defines a gaussian kernel, often 11×11 pixels, which is convolved with the images. At each step of the convolution, the SSIM score is calculated, and the mean of these scores gives the MSSIM score,

$$\mathrm{MSSIM}(I, J) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{SSIM}(I_i, J_i). \tag{3.18}$$

This MSSIM score is supposedly more representative of the perceived image quality [57].

It is also possible to compare 2D images or 3D volumes in the frequency domain, by calculating the Fourier shell correlation (FSC) [23]. FSC is defined as

$$\mathrm{FSC}(\mathbf{R}) = \frac{\sum_{r_i \in \mathbf{R}} \mathcal{F}_I(r_i) \cdot \mathcal{F}_J(r_i)^*}{\sqrt[2]{\sum_{r_i \in \mathbf{R}} |\mathcal{F}_I(r_i)|^2 \cdot |\mathcal{F}_J(r_i)|^2}}, \tag{3.19}$$

and provides information about the correlation at different spatial frequencies between two images or volumes, I and J. Here, $\mathcal{F}_I(r_i)$ is the Fourier transform of I at the spatial frequency r_i, and $\mathbf{R}$ is the set of spatial frequencies to be considered.

A range of different filters can also be applied to images or volumes to enhance and compare certain properties. Two examples of such filters are the Sobel filter and the Laplacian filter. The Sobel filter is an approximation of the gradient of the image, where the S_x and S_y 3×3 kernels detect the horizontal and vertical edges, respectively [17]. They are defined as

$$S_x = \begin{pmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{pmatrix} \qquad S_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{pmatrix}. \tag{3.20}$$

Similarly, the Laplacian filter L is an approximation of the second derivative of the image. The signs of the elements in the matrix may vary, but SciPy [53] defined the 3×3 kernel as

$$L = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{3.21}$$

Part II

PROJECT WORK

4 EXPERIMENTAL

4.1 EQUIPMENT

North Star Imaging Inc. is the provider of the CT scanner used in this thesis. North Star Imaging model X5000, from now on denoted *Industrial CT*, is equipped with a Gulmay MP1 450 kV X-ray source. However, for the experiments in this thesis, a voltage of 250 kV, a current of 2700 µA, and a focal spot size of 400 µm were used. As for the detector, it was a quadratic, (2048×2048), Perkin Elmer [XRD 1620/1621 AM/AN] with pixel pitch of 200 µm and 10 fps frame rate. For the cone beam geometry calibration, a source to detector distance of 1350 mm, and a source to sample distance of 930 mm, were used. A filter of copper and thin was used in front of the source.

The *micro-CT* is the North Star Imaging model X3000. Its X-ray source is a XRayWorX 225 kV operated at 124 kV with 100 µA and a focal spot size of 12.4 µm. The detector is a rectangular, (1536×1920), VarianLo7 [C:10S1-403] with pixel pitch of 127 µm and 5 fps frame rate. Here, the source to detector distance is 502 mm, and the source to sample distance is 47 mm.

4.2 SAMPLES

The sample used for temporal CT is an hour glass, and will for simplicity be referred to as the *hour glass* sample. The hour glass was blown by hand with transparent colourless borosilicate glass with thickness between 1.55 mm and 1.83 mm, and dimensions of 135 mm × 65 mm × 65 mm. Its content is coloured sea sand, SiO_2.

The *sandstone* sample is a reservoir rock from the *Snøhvit* reservoir in the North Sea. Its dimensions are approximately 240 mm in height and 140 mm in diameter. The most distinct features of the sample are a series of wide cracks running inside the sample. Its main purpose was to assess the spatial resolution obtained with enhanced temporal CT, and compare PSNR and SSIM obtained with different procedures.

4.3 DERIVED GOLDEN RATIO SAMPLING PROCEDURE

An efficient scanning routine for temporal CT using industrial CT scanners and micro-CT was derived through trial and error. The resulting procedure was a compromise between flexibility in the post-processing and the collected information per time during scanning. Moreover, the procedure was constrained by versatility of the hardware, and limited by the experimental set-up.

Originally, the procedure was supposed to perform an exposure every 137.5°, meaning every golden angle. This would maximise the amount of information collected per exposure, and provide full flexibility in the post-processing. To elaborate, each sampled projection would provide unique information, and any number of projections could be included for a given reconstruction, representing a time frame. This would enhance the spatial resolution of this reconstruction at the cost of temporal resolution.

However, due time consumed by rotating the sample at a balanced velocity of 21°/s, or 12°/s for the micro-CT, more frequent exposures were necessary to increase the exposure-rotation ratio, and thus increase the amount of information collected per time.

Metallic means, more acute irrational angles, were tested in order to increase the information per time [51]. A higher metallic mean would result in a smaller, but still irrational, angle, and thus more frequent unique sampling. However, using a metallic mean, less new information was collected per exposure, and the number of projections required for even sampling increased quadratically. Not only did this limit the flexibility in the post-processing, this procedure was ultimately discarded due to hardware limitations. The golden angle sampling, and the developed metallic mean sampling, were implemented by manipulating the scanner to perform numerous CT scans consisting of only a single projection. However, this resulted in much wasted time due to software initiation and completion.

Therefore, the procedure was modified to perform more projections within each software-defined CT scan, and golden ratio sampling was utilised to find from which angle to start. Obviously, the angle increment was derived by distributing the number of projections evenly over the revolution. Next, the starting angle of the next revolution was found through golden ratio sampling on the interval of the first angle increment. However, due to tubes and other hardware, the sample stage would alternate between rotating clockwise and counterclockwise. Therefore, the derived sampled angle alternated between being the starting angle and the final angle of the full revolution. From Equation (3.14), one possible way to calculate the golden ratio sampled angle for revolution n was

$$\theta_n = \mod\left[\frac{n360}{\phi p}, \frac{360}{p}\right]. \tag{4.1}$$

A visualisation of the sampling procedure for the first four revolutions of 13 projections each is included in Figure 4.1. The starting point of each revolution is marked with a black diamond, and alternates between belonging to the first and last increment of the revolution, due to the alternation between clockwise and counterclockwise rotation.

4.4 VERIFICATION OF PROCEDURE

The validity of the final golden ratio sampling procedure was assessed by performing a scan of a static sandstone sample with the industrial CT scanner. The resulting projections and reconstructions were compared to the same sample scanned with a standard procedure. Moreover, it was verified that no limitations of the hardware distorted the golden ratio reconstructions. A total of 1440 projections was performed using standard CT, while 1000 projections were collected using the derived golden ratio sampling procedure. In both cases, the number of projections should be sufficient to provide information to each voxel of the reconstructed volumes, respectively. Additionally, a total of 12 frames was used to average each exposure, resulting in an effective exposure time of 1.2 s for each projection.

The obtained projections were processed using flat-field and dark-field corrections, together with a 3×3 median filter around the listed defective pixels. Briefly explained, the projections are calibrated by subtracting the obtained signal when all detector

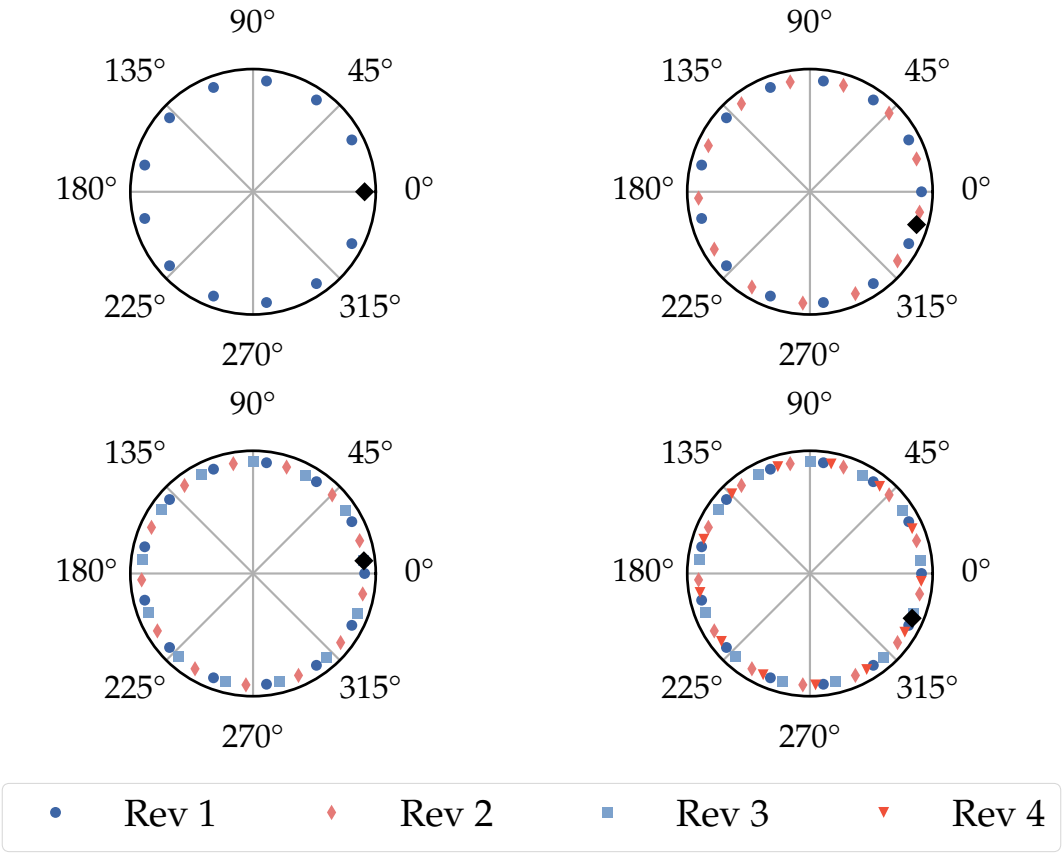


Figure 4.1: Each subplot illustrates the cumulative sampling when utilising the derived sampling procedure. One completed revolution is shown in the top left subplot. Two completed revolutions are shown in the top right subplot. Additionally, the respective contributions to the sampling from each revolution is illustrated by the different markers. For each revolution, the longest unsampled interval is divided into two shorter intervals by the golden ratio. The starting point of each revolution is marked with a black diamond, and alternates between belonging to the first and last increment of the revolution, due to the alternation between clockwise and counterclockwise rotation.

pixels are *off*. This is the dark-field correction. Calibration is completed by dividing the obtained signal to the maximal detector range, which is the difference between the flat-field and dark-field image. Then, the measured intensities were expressed as a function of absorption, utilising Equation (1.4). The centre of rotation was determined by sharpness analysis. To elaborate, the raw projections were shifted by a varying number of pixels, then used for a reconstruction, for whom the sharpness of the reconstructed volume was assessed by applying a Sobel filter (3.20). The centre of rotation was determined by the pixel offset that resulted in the best sharpness score. Alternatively, one could use the offset that resulted in the highest variance of the reconstructed volume [61]. Golden ratio search was used to find the centre of rotation with a precision of 0.1 pixel as long as the sharpness was approximately concave in the vicinity of the centre of rotation. Otherwise, a linear search was applied, with the sharpness values being plotted and manually inspected to find the centre of rotation.

4.5 FURTHER OPTIMISATION

An effort was made to further optimise the golden ratio sampling procedure by adjusting the number of frames averaged per exposure (favg). To evaluate the effect of adjusting the exposure dose, the sandstone was scanned thrice, with 6, 12, and 24 frames averaged per exposure, respectively. Whence, the resulting exposure times were 0.6 s, 1.2 s, and 2.4 s, respectively. The idea behind reducing the exposure time was to shorten the scan time, and thus increase the temporal resolution, at a negligible cost of spatial image quality. At the same time, the idea behind increasing the exposure time was to enhance the quality of each projection, and thus possibly enabling a lower number of projections in a given reconstruction, which would also increase the achievable temporal resolution.

Another post-processing effort to optimise the achievable temporal resolution, as well as the exposure ratio, was to perform missing wedge reconstruction with 180° coverage. Thereby, the rotation time could be halved for each reconstruction. The idea was that the quality gain in terms of temporal resolution and exposure ratio would outweigh the quality loss due to cone-beam artefacts in the missing wedge reconstruction.

5 DATA AQUISITION

5.1 SIMULATED DATA

All eleven two-dimensional simulated datasets from *TomoBank* [9] were downloaded. These phantoms had dimensions of 256×256 pixels, and had originally been generated using the *TomoPhantom* package [30]. Each pixel had in most cases either the value 0 or 1, and in some instances some gray values. With the TomoPhantom package, it could in theory be possible to generate more simulated datasets efficiently. However, time was not prioritised for such a procedure. Instead, the simulated datasets were given a third dimension. To elaborate, each simulated dataset was copied 32 times in the third dimension. Following this, 32 256×256 slices of zeros were added to the existing data. Another volume consisting of 64 slices produced like described above was rotated slightly and added to existing simulated volume. This procedure was repeated until the width of the third dimension was 512 pixels. Hence, eleven 2D datasets were used to create eleven high resolution 3D datasets with dimensions of 256×256×512 pixels.

From this ground truth dataset, the TIGRE toolbox [3] was used to generate projections of each sample. Subsequently, the projections were reconstructed using the FDK algorithm. Given the width of 256 pixels, a fully sampled reconstruction would require approximately 402 projections, as defined in Equation (3.1). The number of projections for the undersampling ranged between 201 and 6, corresponding to undersampling factors of 2 to 64, respectively.

Slices in the XY-plane of the target simulated data can be studied in Figure 5.1, and are denoted with the prefix *SIM*, while the corresponding 8 times undersampled reconstructions are shown in Figure 5.2, with the MSSIM value relative to the ground truth listed.

5.2 EQUINOR ASA DATASET

As for experimental data, high quality reconstructions of different porous rocks were provided by Equinor ASA. The data had been reconstructed in the period from 2016 to 2018 using the FDK algorithm and the NSI software provided by the producer listed in Section 4.1. The projection data was acquired using the micro-CT mentioned in Section 4.1. Like the simulated dataset, undersampled reconstructions were generated using the TIGRE toolbox by a forward projection operation followed by backprojection. Because the dimensions of the different reconstructions differed, each dataset was created with a common undersampling factor. Training datasets with undersampling factor 8, 12, and 16 were generated, but an undersampling factor of 8 was chosen for the final training. The experimental dataset target XY-slices are denoted *EQNR* in Figure 5.1, as they were provided by Equinor ASA, with the corresponding undersampled reconstructions shown in Figure 5.2. Shown in Figures 5.3 and 5.4 are the middlemost XZ-slices of the same dataset.

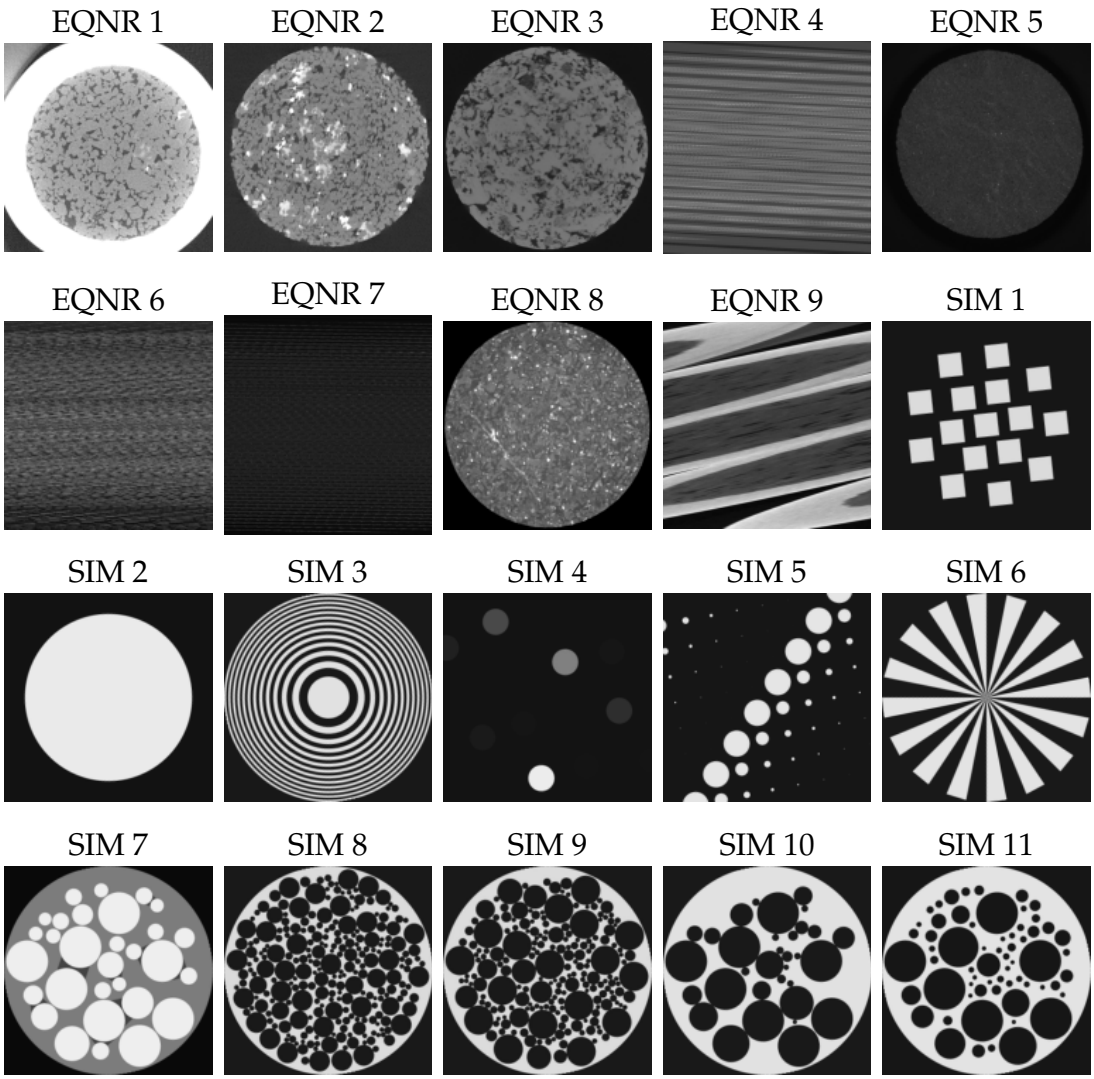


Figure 5.1: The target reconstructions of the training dataset. The middlemost cross section in the XY-plane is shown. *EQNR* indicates that the data was provided by Equinor ASA, while *SIM* indicates that the data was retrieved from *TomoBank* [9].

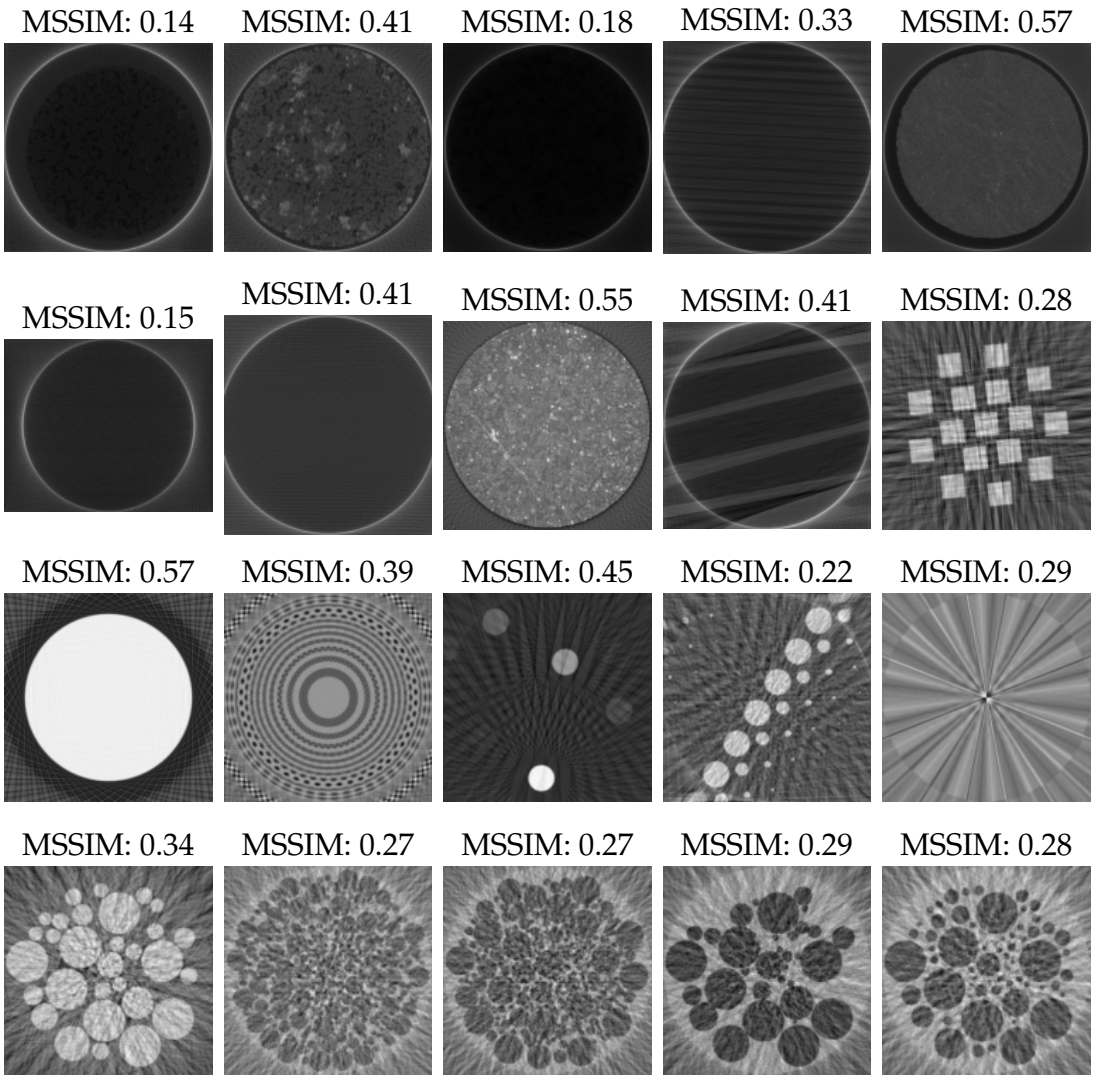


Figure 5.2: The undersampled FDK reconstructions of the training dataset with an undersampling factor of 8. The middlemost cross section in the XY-plane is shown. The MSSIM value with respect to the ground truth is listed.

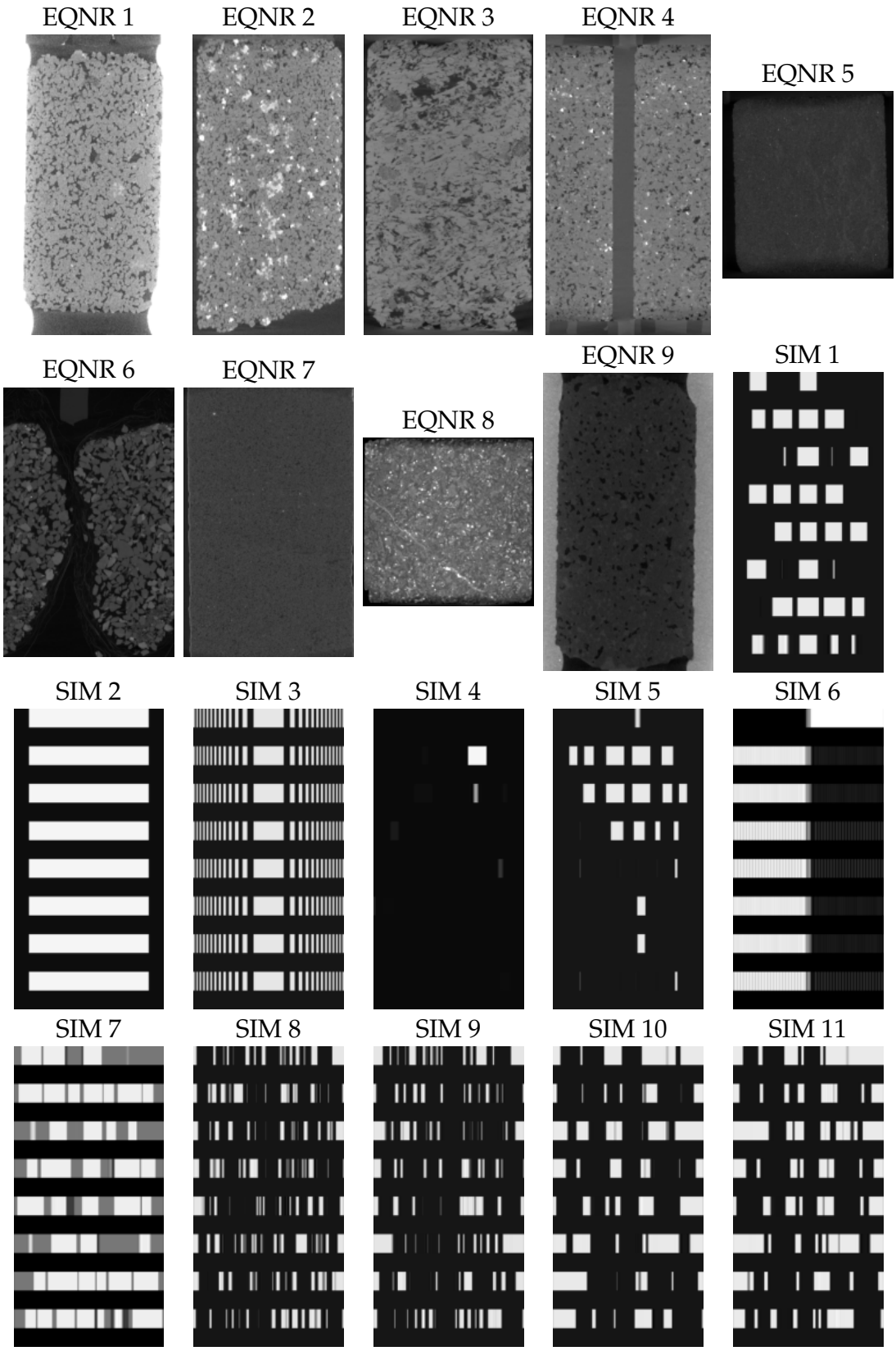


Figure 5.3: The target reconstructions of the training dataset. The middlemost vertical slice in the XZ-plane is shown. *EQNR* indicates that the data was provided by Equinor ASA, while *SIM* indicates that the data was retrieved from *TomoBank*, and is simulated [9].

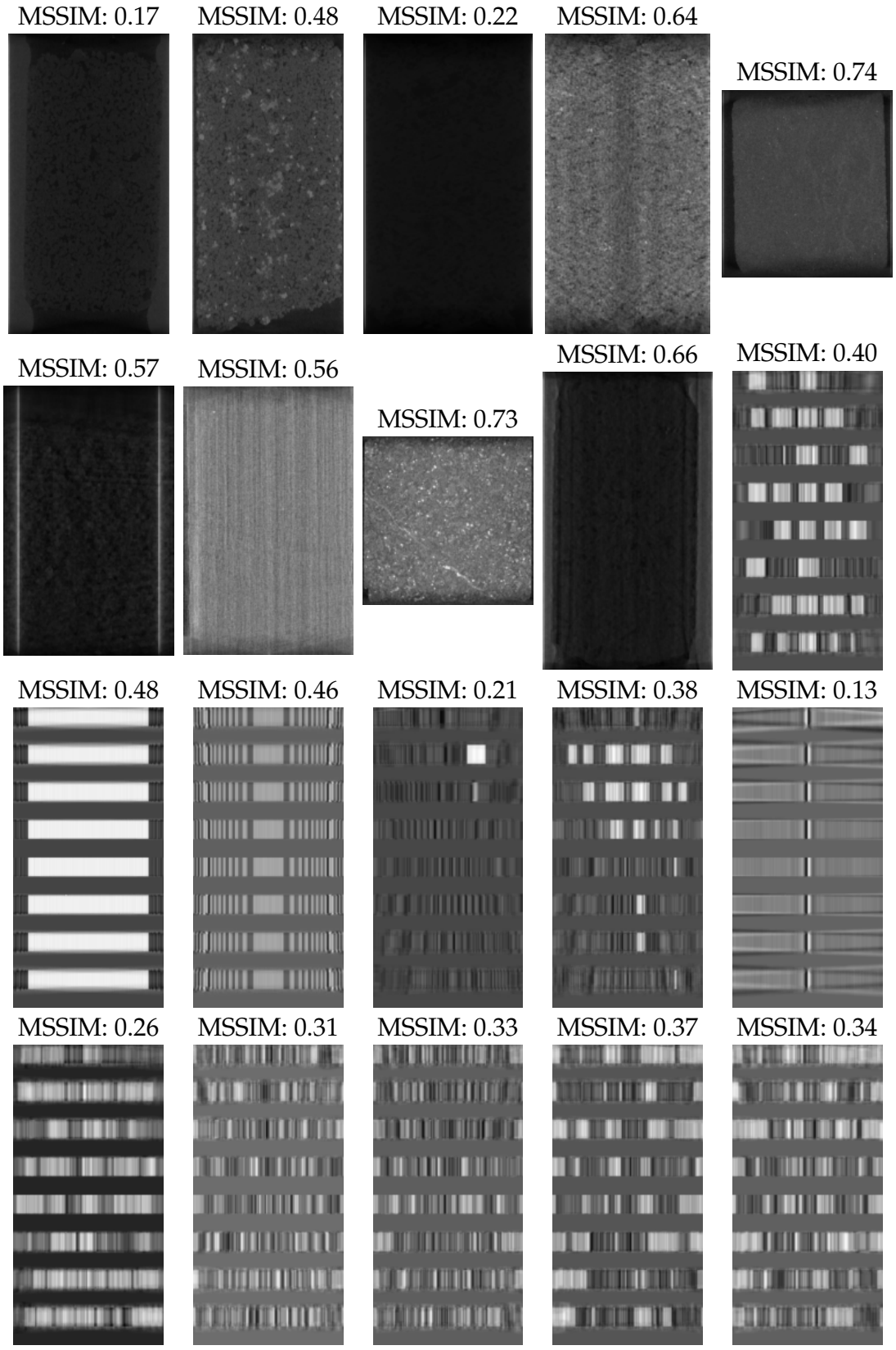


Figure 5.4: The undersampled FDK reconstructions of the training dataset with an undersampling factor of 8, illustrated by the middlemost vertical slice in the XZ-plane. The MSSIM value with respect to the ground truth is listed.

5.3 SIRT DATASET

Training data using 50 SIRT iterations and an undersampling factor of 8 were also collected. Here, the reconstruction denoted *EQNR* 3 was excluded from the training dataset, as its dimensionality caused it to be too computationally expensive to reconstruct. It should be noted that the height of the reconstruction could have been cropped, but this was not prioritised. The middlemost cross sections in the XY-plane of the SIRT reconstructions are included in Figure 5.5, while Figure 5.6 illustrates the corresponding middlemost XZ-slices.

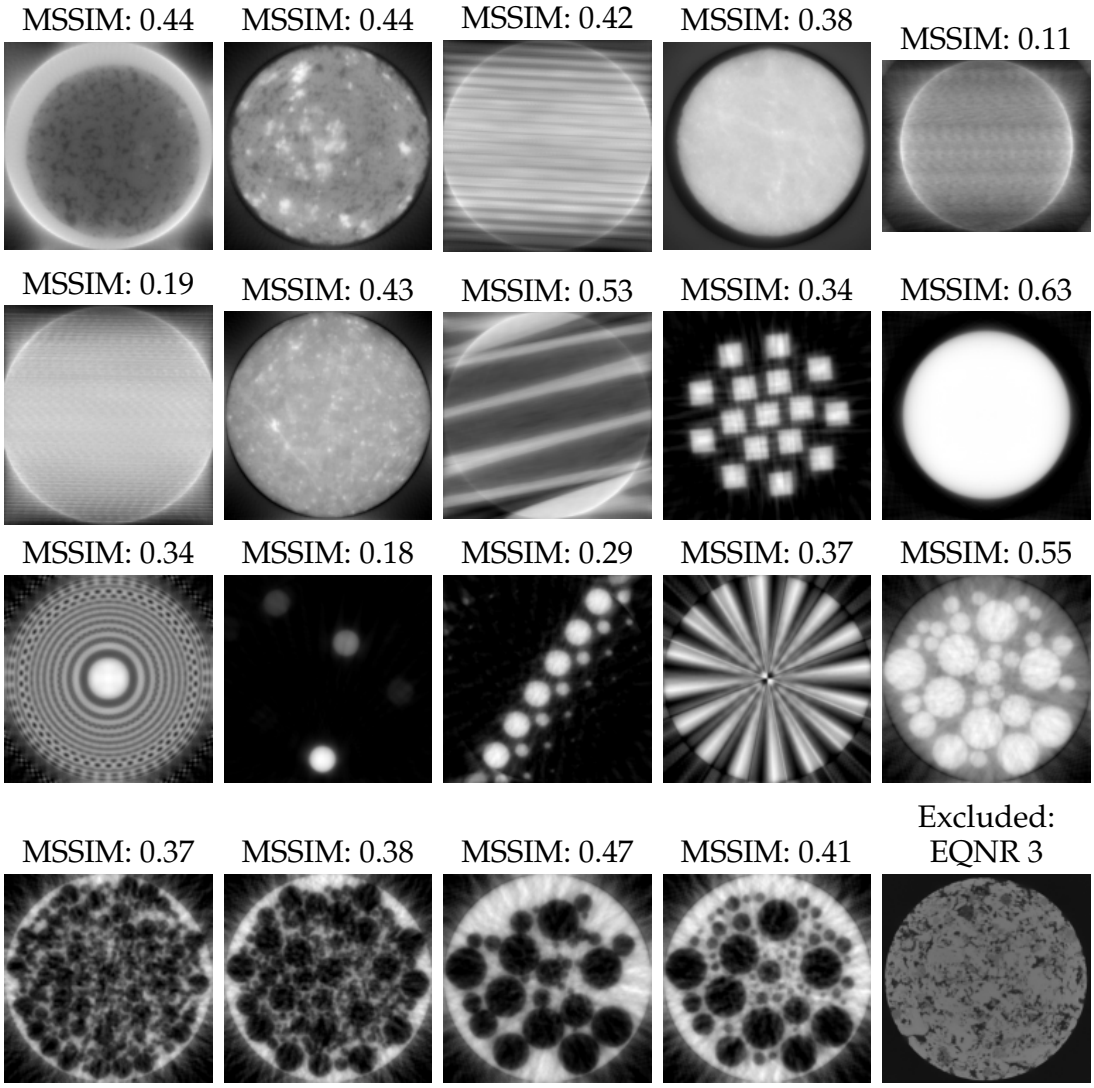


Figure 5.5: The SIRT reconstructions of the training dataset with 50 iterations and an undersampling factor of 8. The middlemost horizontal cross section is shown. The sample *EQNR* 3 was excluded from the training dataset.

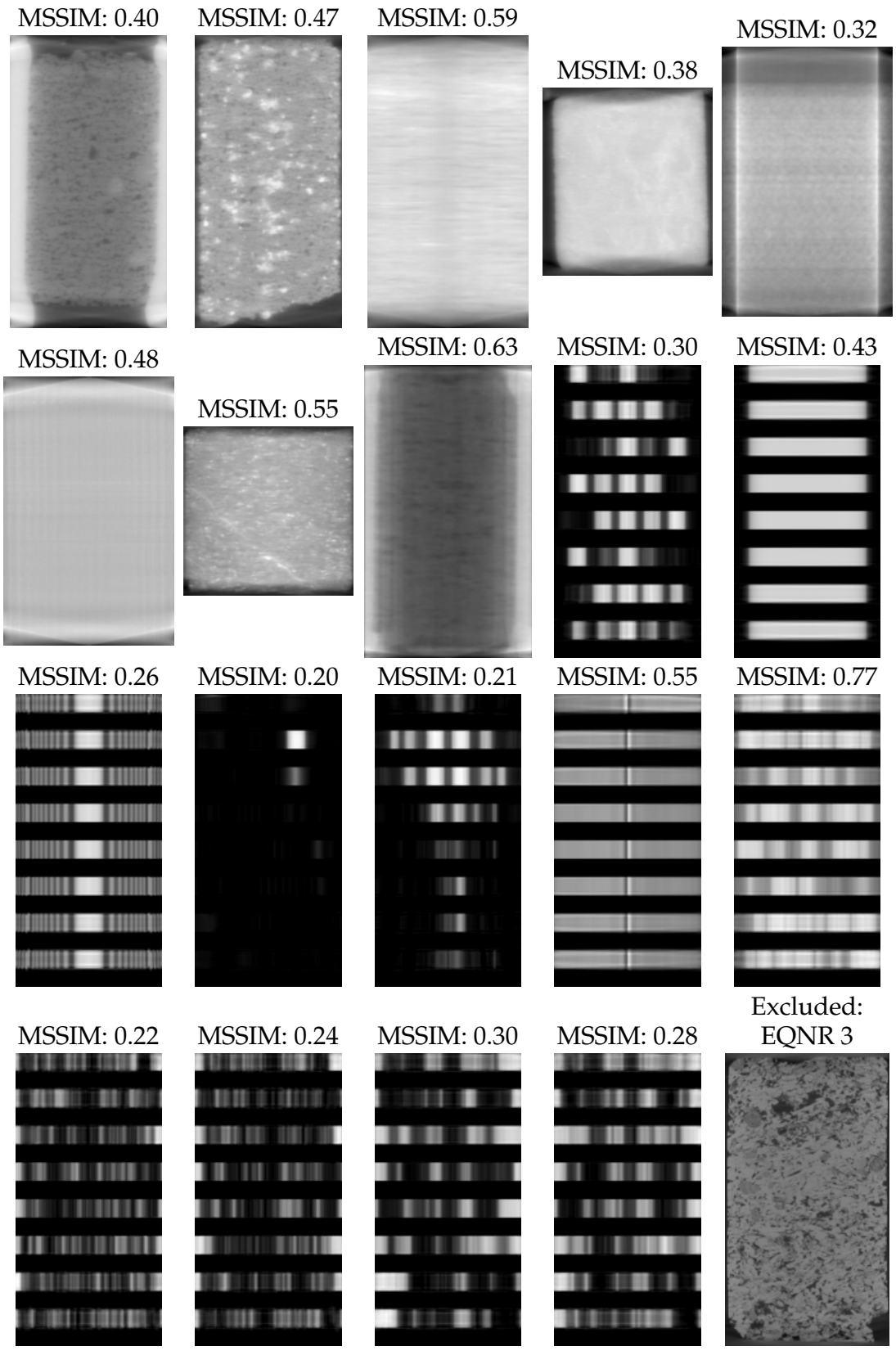


Figure 5.6: The SIRT reconstructions of the training dataset with 50 iterations and an undersampling factor of 8. The middlemost vertical slice in the XZ-plane is shown. The sample *EQNR 3* was excluded from the training dataset.

6

DEVELOPMENT OF RECONSTRUCTION ENHANCEMENT MODEL

6.1 DATA PREPARATION AND AUGMENTATION

As mentioned in Sections 5.1 and 5.2, undersampled data were generated using the TIGRE toolbox [3]. Both the FDK algorithm and the iterative SIRT algorithm were used.

Before entering the reconstruction model, both the target ground truth and the generator input were normalised to values between 0 to 1. Moreover, the reconstructions were cropped to 128^3, and further augmented by flipping, randomly inverting, and rotating the reconstructions. In some cases, rotation was not performed as due to interpolation artefacts. Nevertheless, the effective size of the datasets increased dramatically due to the high number of possible permutations. The three-dimensional transforms were implemented using the *TorchIO* package [42], which specialises in a standard set of transforms for medical CT imaging that closely resemble the available Pytorch 2D transforms [40]. The data preparation and augmentation processes in the cases of training and application are illustrated in Figure 6.1.

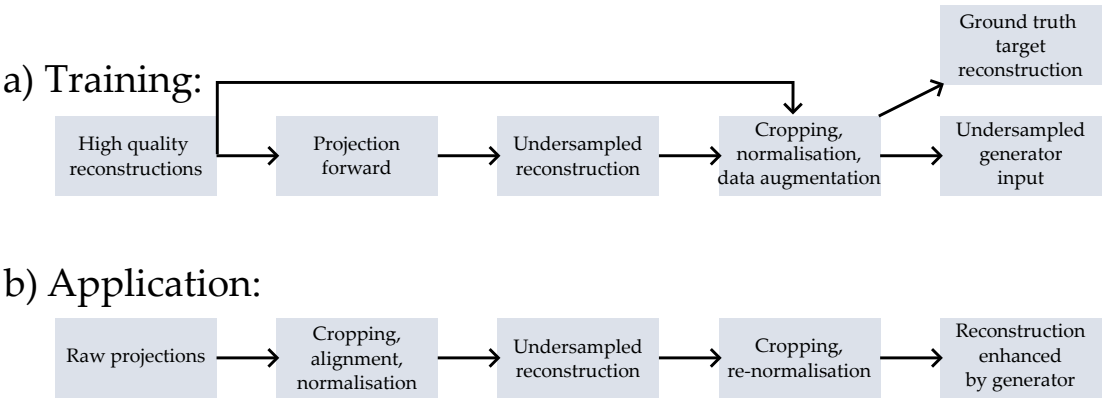


Figure 6.1: Illustration of the data preparation and augmentation processes for training and application. In a), undersampled reconstructions are generated using forward projection and subsequent backprojection. Followings this is the data augmentation process. The resulting data are used for the GAN training process. In b), a CT scan with few projections is reconstructed, and only cropping and normalisation is performed before the reconstruction enhancement model is applied. The main differences are the use of data augmentation during training, and that application of the trained model was used on new CT scans sampled using the derived golden ratio sampling procedure from Section 4.3.

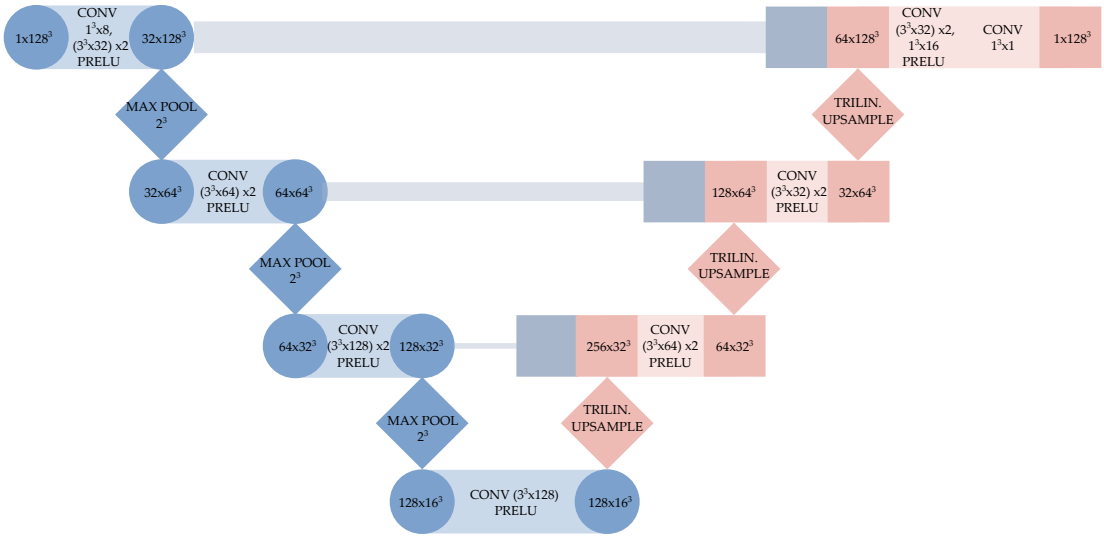


Figure 6.2: Illustration of the 3D fully convolutional max-pooling U-net generator architecture. Input and output shapes are given by the circles and squares. Convolutional block information is provided between the nodes. Pooling and upsampling are symbolised by diamonds.

6.2 ARCHITECTURE

The developed model used for reconstruction enhancement was a cGAN consisting of a 3D fully convolutional U-net generator and a 3D convolutional discriminator with two linear layers for binary classification. The respective architectures are illustrated in Figures 6.2 and 6.3. Input and outputs in the diagrams are illustrated by circles, while the connections between the circles provide information about the convolutional kernels in the given convolutional block. The kernel spatial size is given together with the number of output channels. The kernel input channel size is neglected, as this information is provided by the input node marker. Additionally, the convolutional stride is given only if it was not equal to 1, and parametric ReLU (PReLU) was predominantly the activation function of choice. Note, however, that the final layer of the generator was left without a non-linear activation function. The generator output was rescaled to values between 0 and 1 using min-max normalisation. The discriminator had instead a sigmoid activation function to ensure meaningful results from the binary cross-entropy loss function, as seen in Equation (2.5). Input and output features of the linear layers are illustrated by the nodes, while the weights, and thereby the matrix vector multiplication, are illustrated by the connectors. Special operations, such as max pooling, trilinear upsampling, and adaptive pooling are symbolised by diamond nodes. Finally, the skip connections with concatenation of channels are illustrated by the long horizontal arrows with squared ends. Shortly explained, max pooling reduces the spatial size of the input, upsampling increases the spatial size, adaptive pooling ensures a constant spatial output size, and skip connections concatenate the channels of one convolutional block with the input channels of another convolutional block.

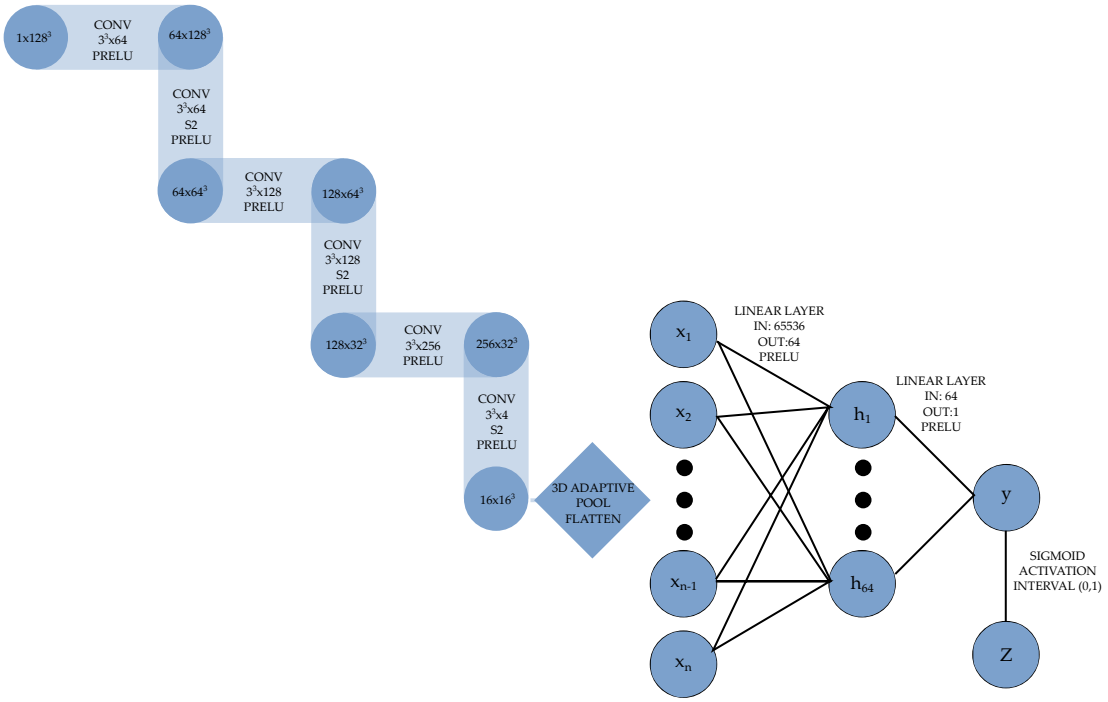


Figure 6.3: Illustration of the 3D strided convolutional binary classifier discriminator architecture. Input and output shapes are given by the circles. Strided convolutional kernels are vertically aligned. 3D adaptive pooling connects the convolutional layers to the linear layers, whose matrix-vector multiplication is symbolised by nodes and connectors.

6.3 TRAINING

One epoch in the training process consisted of generator steps, discriminator steps, and a validation. The latter was used to monitor the generalisation error of the model to prevent overfitting, and consisted therefore of only a forward pass of the validation dataset, with no updating of weights. The actual training process consisted of a forward pass followed by backpropagation of the generator and discriminator, respectively. A diagram illustrating these separate processes is provided in Figure 6.4. During the generator step, the output was compared to the target ground truth by plain mean squared error (MSE) loss (2.2), as well as MSE loss of the first parts of the ResNet convolutional feature extractor output. This loss term did, however, only compare 2D slices in one plane, as it was severely memory intensive to repeat this procedure along each axis. The 3D generator output and the 3D high quality reconstruction had to be transformed to 224×224 with three channels, passed through the first bottleneck of the freezed ResNet feature extractor, from which the outputs were compared by MSE loss. The generator output was also fed into the discriminator, which provided an adversarial binary cross entropy loss (2.5). If the generator managed to fool the discriminator, the adversarial loss would be close to 0. A weighted average of the different loss terms was used to calculate the total generator loss. Backpropagation was then applied to update the generator weights based on the ADAM optimiser (2.8).

As for the discriminator step, the discriminator performed two classifications. First the generator output was evaluated, then the high quality reconstruction. The discriminator output of the high quality reconstruction was compared to the target label 1 by binary cross entropy loss (2.5), and the discriminator output of the generated reconstruction was compared to the target label 0 by the same loss function. Update of the weights was also in this case performed by the ADAM optimiser on the sum of the two loss terms.

The equipment available for training was either a single NVIDIA GeForce RTX 3090 with 10 496 cores and 24 GB of memory, or three distributed NVIDIA QUADRO P6000 with 3840 cores and 24 GB of memory each. Even though the three QUADRO P6000 outperformed the single RTX 3090 in terms of number of cores and memory capacity, distribution across several GPUs involves an additional overhead due to inter-GPU communication.

6.4 MODEL AND TRAINING OPTIMISATION

Compared to the original TomoGAN and the first developed model and training algorithm in this project, several modifications were implemented. Firstly, the capacity of the discriminator was increased by adjusting the number of output channels in the last convolutional layers from 4 to 16. Secondly, the output activation function of the generator was changed from linear activation to a normalisation layer, so that the output was constrained to the range 0 to 1. Thirdly, the VGGNET feature extractor, that was used in the TomoGAN paper [33], was replaced by the first layers of the ResNet feature extractor [25]. As mentioned, PReLU was the activation function of choice for both the generator and the discriminator, while the original TomoGAN lacked this additional learnable parameter inbetween layers.

In terms of training optimisation, the training step relation between the generator and the discriminator was removed. Instead, a while-loop was implemented, where the generator was trained until the discriminator was fooled, and then the discriminator was trained until the generator was fooled. The criteria for the generator and discriminator to be fooled were defined as a certain adversarial loss or discriminator loss, respectively. For instance, a natural criterion was to stop generator training when the adversarial loss was below 0.69, because a BCE loss of 0.69 indicates random guessing (2.5). Likewise, a natural discriminator criteria would be 1.38. The possibility of an endless while-loop was accounted for by introducing a maximum number of 10 successive generator and discriminator steps, respectively. To fine tune the training, a cosine-shaped learning rate scheduler was implemented, that reduced the learning rate over the course of the maximum number of iterations. Finally, the generator weights were frequently saved during training, and the best performing models were manually selected based on the validation loss, relative improvement in validation MSSIM, and stability. The optimal generator loss function weights were found through trial and error, meaning adjusting the importance of the MSE, adversarial, and perceptual loss, respectively.

Furthermore, successfully trained models were saved and used as initialisation for other training sessions. This transfer learning was performed to reduce training time, increase training stability, improve the generalisation error, and adapt the GAN to new datasets. For instance, a model trained on simulated data was used as initialisation for training on experimental data. The same model was also used as initialisation for training on undersampled SIRT reconstructions.

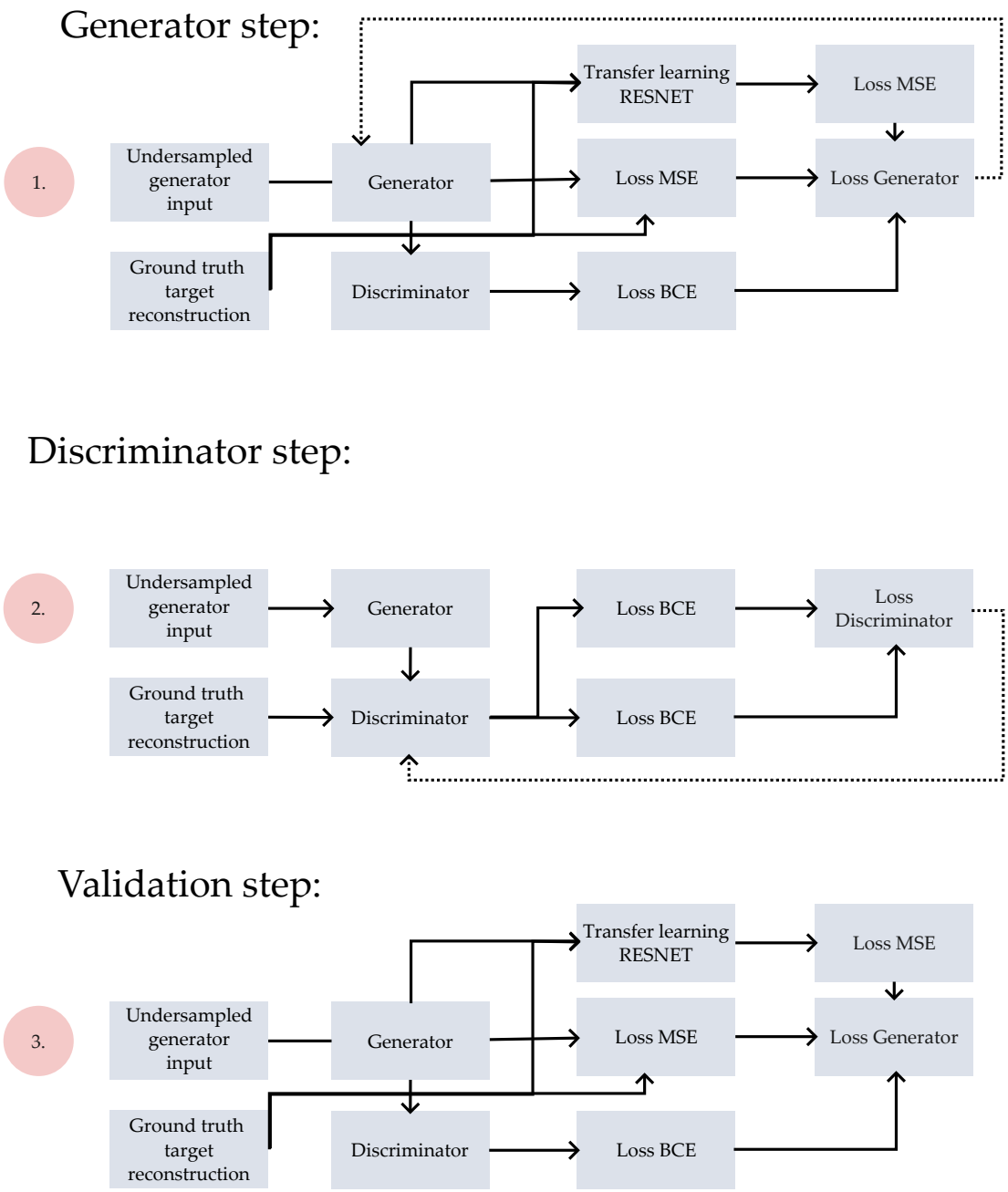


Figure 6.4: The GAN training process consisted of generator steps, discriminator steps, and validation. For the generator, MSE loss was computed directly and after ResNet feature extraction, while the adversarial loss was computed using BCE. In contrast, the discriminator loss was the sum of two BCE losses: one for the classification of generator output, and one for the classification of high quality reconstruction. No updating of weights was performed during validation.

7 PERFORMED EXPERIMENTS

7.1 4D-CT WITH ENHANCED TIME RESOLUTION

Because the aim of this thesis was to enhance 4D-CT, a simple, yet full worthy, dynamic CT experiment was performed. The hour glass sample, the industrial CT scanner, and the derived golden ratio sampling procedure, described in Section 4.2, 4.1, and 4.3, respectively, were used for the experiment. The experiment was performed twice, once with 25 projections per revolution, and once with 13. The initial post-processing of the measured projections included median-filtering of defect pixels, flat-field and dark-field corrections, rotation to upright position, determining AoR, and cropping (1536,1024) as the region of interest (RoI). The latter experiment, which consisted of 13 projections per revolution with a total of 715 projections, was used in the final analysis, because the sample was in this case tilted to reduce the symmetry, making it a more challenging reconstruction. The processed projections were binned using a 2×2 average pooling convolutional operator with stride 2, which halved the dimensions along each axis in terms of pixel count. The resulting projections were reconstructed using the FDK algorithm. One reconstruction utilised all 715 projections, while another only utilised the very first full revolution. Following reconstruction of the volume, the fully trained generator was used to enhance the undersampled reconstruction. The generator was the result of transfer learning from initial training on simulated data, and fine-tuning on the complete data set shown, for instance, in Figure 5.2 Focus was directed towards the sand-air interface, which was the most dynamic part of the sample. To further investigate this interface, a Sobel filter (3.20) and a Laplacian filter (3.21) were applied, and the sand-air interface was cropped for a more thorough investigation. In one last experiment to assess the maximal temporal resolution, the first revolution was divided into two parts of 7 projections with approximately 180° coverage each, and reconstructed using the FDK algorithm. These resulting missing wedge reconstructions were enhanced by the generator, and compared before and after enhancement. Finally, the movement of the sand-air interface was visualised by comparison of the two reconstructed time stamps.

7.2 SPATIAL COMPARISON OF RECONSTRUCTIONS

The spatial reconstruction enhancement of the hour glass sample was also assessed. For this analysis, no binning was performed to preserve the original number of pixels. The hourglass was reconstructed and enhanced using between 1 and 9 full revolutions, corresponding to 13 and 117 projections, respectively. The spatial quality metrics used to assess the enhancement were global SSIM and PSNR of volumes, as well as MSSIM and PSNR in cropped cross sections. Additionally, line profiles were visually investigated. Iterative reconstruction using the SIRT algorithm was also added to the comparison. A total of 100 iterations were performed using both 26 and 52 projections, with a computational time of approximately 25 min per reconstruction.

The same analysis was repeated for a static CT scan of the sandstone sample. Here, the number of projection per revolution was 17, and the total number of projections was 935. The ground truth was a fully sampled FDK reconstruction utilising all 935 projections. The undersampled FDK reconstructions ranged between 1 and 14 full revolutions, corresponding to 17 and 238 projections, respectively. Because the entire volume was static, the volume was cropped and visualised along all axes, with a focus on the major cracks of the sample. Also in this case, iterative reconstruction using the SIRT algorithm was included in the comparison. The settings of the iterative reconstruction were 150 iterations and 34 projections.

7.3 ENHANCEMENT OF ITERATIVE RECONSTRUCTIONS

In an experiment to test the versatility of the developed GAN, the SIRT dataset shown in Figures 5.5 and 5.6 was used for another round of training, with the generator weights from the initial training on simulated data as initialisation. The goal of the experiment was to assess whether the end result of reconstruction followed by enhancement could be better by using a more robust reconstruction algorithm. Moreover, the experiment was aimed at shortening the required computational time for SIRT. Note that the fully sampled FDK reconstruction was maintained as the target in this dataset. In this way, typical iterative reconstruction blurring would be present in the training data, but not the target data.

7.4 ADDITIONAL TESTS

A set of additional tests were performed to better understand the dynamics of the developed GAN. In order to minimise confusion, the results of these tests are presented in Appendix C.

Firstly, the original aim of this project was to develop a general enhancer that did not require training on a specific sample. Nevertheless, the developed GAN was trained on a reconstruction of the sandstone sample where the ground truth consisted of 1000 projections, and the undersampled reconstruction consisted of 50 projections. No data augmentation except for random cropping, normalisation, and random inversion were applied. Training was performed using both FDK and SIRT, using 150 iterations in the latter case. As usual, for the initialisation of the generator, the obtained weights from the initial training on simulated data were used. The motivation for this experiment was to investigate if a more accurate enhancement could be obtained by providing the true probability distribution. By evaluating the results of the experiment, it would be possible to asses whether the training dataset was sufficiently representative of the true probability distribution of the sandstone.

Finally, the suspected weakness of training the GAN using online learning was investigated. For this purpose, the spatial input dimensions were reduced to 64^3 so that a batch size of 4 could be used. These were the only two hyperparameter changes compared to the training resulting in the generator used in Section 7.1. The training was performed using the same dataset as in the main experiments, namely Sections 7.1 and 7.2. With the increased batch size was motivated by a desire to achieve a more stable convergence towards a set of weights that would satisfactorily enhance all samples in the dataset.

Part III

PRESENTATION OF RESULTS AND DATA ANALYSIS

8 GOLDEN RATIO SAMPLING

8.1 VERIFICATION OF GOLDEN RATIO SAMPLING PROCEDURE

Figure 8.1 illustrates the similarity between cross sections retrieved from standard CT sampling and the derived golden ratio procedure. The corresponding intensity distributions in Figure 8.2 revealed no distinct difference in terms of visual perception or actual pixel values. It is worth mentioning that the most probable reason for the small distinctions in intensity distributions was due to a different number of total projections in the CT scans.

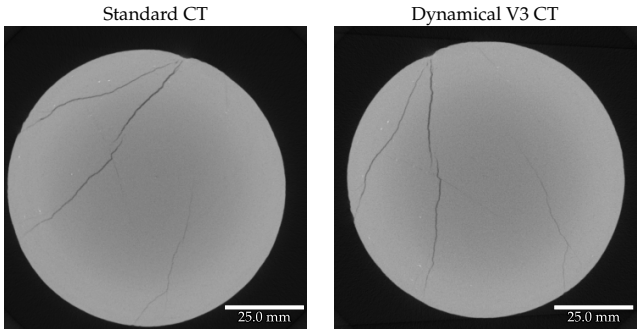


Figure 8.1: The same horizontal cross section of the sandstone sample is presented for two different CT scanning techniques. On the left is the result of a standard CT scan, where the 1440 projections were evenly distributed over the full rotation of 360°. The result of utilising the derived golden ratio procedure, explained in Section 4.3, is illustrated on the right. The scan consisted of 55 revolution, where the number of projections in each revolution was 17.

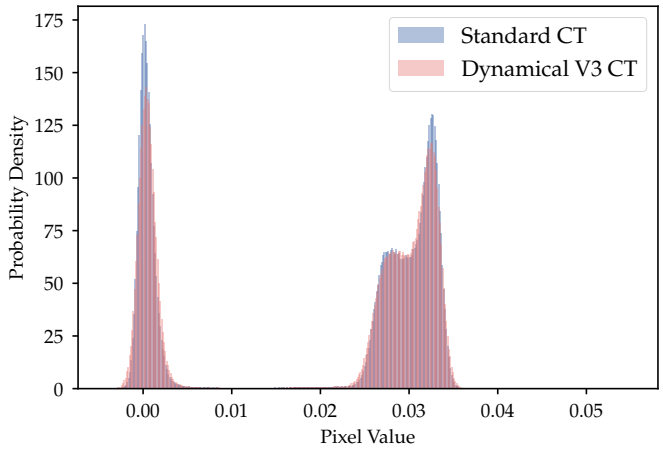


Figure 8.2: The pixel intensity distributions of the cross sections shown in Figure 8.1. Before plotting, the cross sections were cropped to minimise the influence of the background. The intensity distributions were very similar, as expected.

8.2 UNDERSAMPLED RECONSTRUCTION QUALITY

As explained in Section 4.5, an experiment was conducted to determine the optimal exposure time when performing temporal CT scans. The exposure time was tuned by adjusting the number of frames averaged (favg) in each projection. One hypothesis was that by reducing the exposure time, one would achieve a higher temporal resolution at a negligible loss in image quality. The opposing hypothesis supposed that enhanced exposure time would improve the image quality substantially, which could ultimately allow for fewer projections in each reconstruction. Firstly, a cross section of the sandstone containing several details in the form of cracks was chosen as the object of interest. Next, it was verified that the SSIMs between the fully sampled reconstructed cross sections were indeed close to 1 when comparing different exposure times. Specifically, the mentioned SSIMs were 0.9988, 0.9993, and 0.9985 for 6 favg compared to 24 favg, 12 favg compared to 24 favg, and 12 favg compared to 6 favg, respectively. Correspondingly, the PSNRs were 32.3, 35.0, and 31.9.

The reconstruction with the longest exposure time was chosen as the reference for the following analysis, because it, as expected, contained the least amount of noise, as shown the intensity distribution included in Figure 8.3. The intensity distribution shown is from the region of interest marked with a red rectangle in the subplot on the left. This region will be the basis for the following analysis. The adjusted SSIMs and PSNRs for the ground truth comparison in this region were as follows: 0.9974, 0.9984, and 0.9969, and 30.5, 33.0, and 30.0, in the same order as mentioned above. However, the MSSIM values for this RoI with a local 11×11 gaussian kernel were 0.505, 0.63, and 0.476, respectively. The fully sampled cross section of interest captured with an exposure time of 2.4 s, corresponding to 24 favg, is shown on the left.

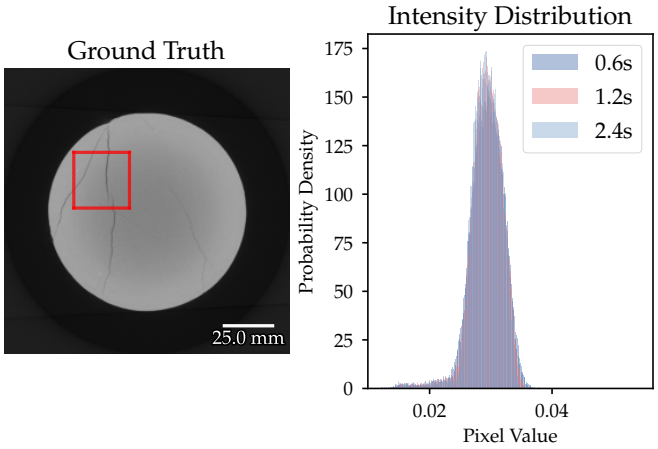


Figure 8.3: The Intensity distribution of the region of interest marked by the red box on the left was included for different exposure times in the histogram on the right. The highest peak was obtained with the highest exposure time, as expected.

As for the undersampled reconstructions, one, two, and three full revolutions of 17 projections were used in the respective reconstructions. For all undersampled reconstructions, the SSIM and PSNR were calculated with respect to the reference reconstruction. The results are included in Table 8.1. The respective projections correspond to an undersampling factor of 55, 27.5, and 18.3 with respect to the 935 projections in the reference reconstruction. From the table, one should note that the increases in

SSIM and PSNR were greatest when going from heavily undersampled and heavily underexposed. However, at 17 projections, the increase was greater by doubling the number of projections than by doubling the exposure time, both in terms of structural similarity and pixel-wise loss. Expectedly, the highest global SSIM and PSNR was obtained by combining the highest exposure time and the highest number of projections. With 51 projections and 2.4 s exposure, a global SSIM in the mentioned RoI above 0.5 was achieved.

Table 8.1: An overview of the global SSIM and PSNR values for the undersampled reconstructions in the region of interest marked by a red rectangle in Figure 8.3. The values in the respective matrices are indicated by the table headers.

SSIM				**PSNR**			
Projections	**Exposures**			**Projections**	**Exposures**		
	0.6**s**	1.2**s**	2.4**s**		0.6**s**	1.2**s**	2.4**s**
17	0.204	0.238	0.329	**17**	13.92	16.22	18.46
34	0.260	0.326	0.437	**34**	16.82	19.07	21.16
51	0.303	0.387	0.523	**51**	18.47	20.77	22.89

Furthermore, a visual inspection of the undersampled reconstructions was performed based on Figure 8.4. From the figure, it was evident that all reconstructions were heavily undersampled. The contrast from the cracks were in the most heavily undersampled and underexposed cases indistinguishable with surrounding streaking artefacts. Moreover, the reconstructions in the left column, with the lowest exposure time, had considerable random pixel-noise. The magnitude of the noise and artefacts were exemplified in Figure 8.5 by retrieving a line profile across two cracks in the sandstone. Both cracks were resolved in all reconstructions, but the surrounding noise and artefacts were of the same order of magnitude as the feature itself.

In terms of temporal resolution, the total acquisition times for the different scan settings are included in Table 8.2. However, if the number of projections listed were the number of projections per revolution, the updated scan times and exposure ratios would be those listed in Table 8.3. The latter is simply defined as the ratio between the time spent stationary counting photons and the total acquisition time. Additionally, the MSSIM scores are included in Table 8.2 for the different combinations of exposure times and undersampling factors. Generally, the MSSIM score is lower than the global SSIM score. Moreover, unexpected results, such as the fact that the MSSIM score for 17 projections and 1.2 s was the lowest observed, were present. Except for this outlier, the MSSIM score generally increased with increasing exposure time. The scan times will be further presented in the following section, but it was observed that increasing the exposure time provided a larger increase in exposure ratio than increasing the number of projections. With more projections, the effective rotation rate was reduced, which results in increased rotational time in addition to the prolonged total exposure.

8.3 TIME CONSUMPTION WITH GOLDEN RATIO SAMPLING

As mentioned in Section 4.3, the optimal sampling procedure for temporal CT experiments on the mentioned laboratory CT scanners was found to be a compromise

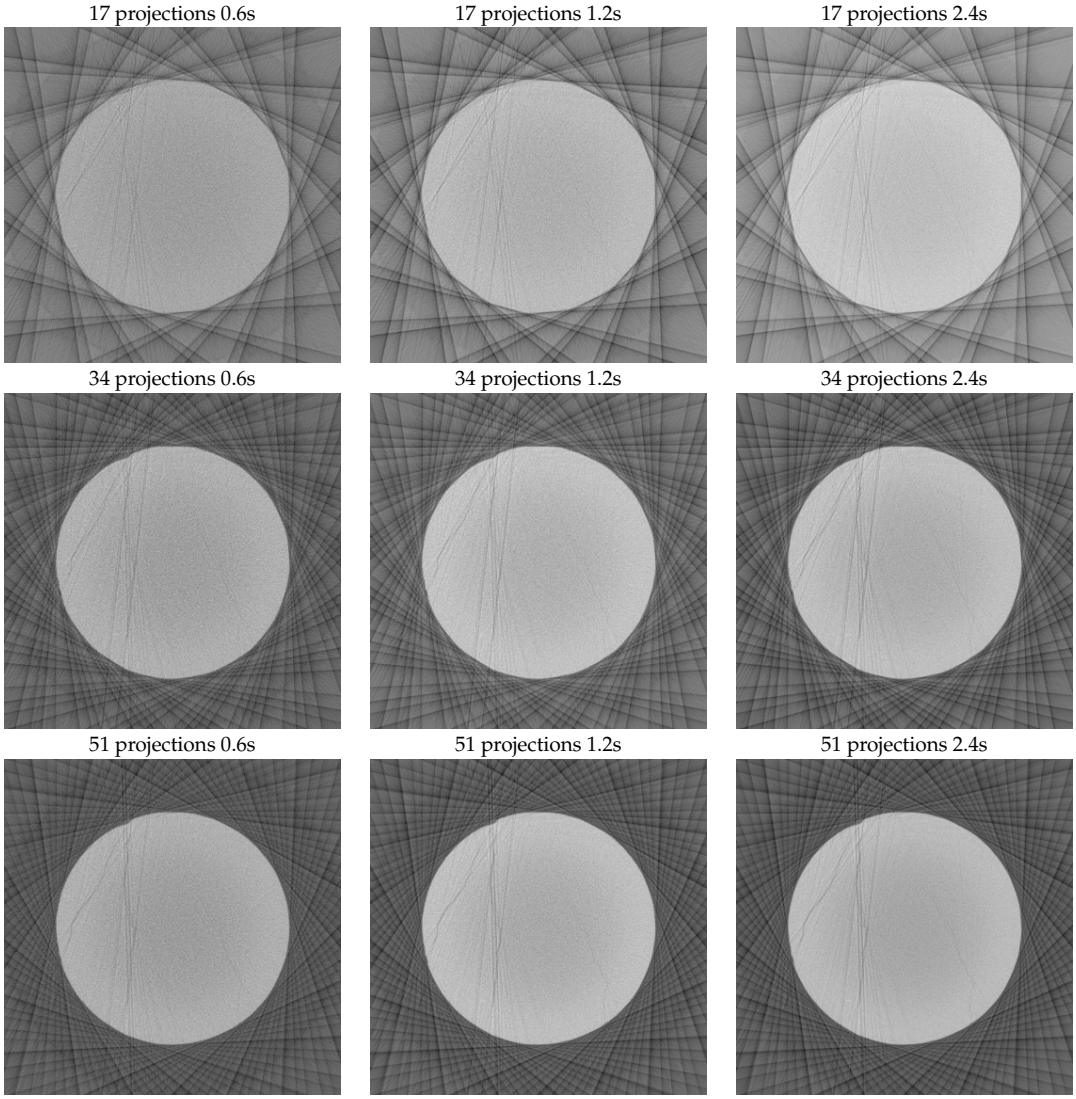


Figure 8.4: The grid sorts the reconstructions for the different combinations of exposure times and undersampling factors. Each column has the same exposure time, and each row has the same number of projections. Notice that the undersampling artefacts were the most prominent in the top row, while the random pixel noise was the most prominent in the left column. Undersampling resulted in streaking artefacts instead of resolving the cracks in the sandstone.

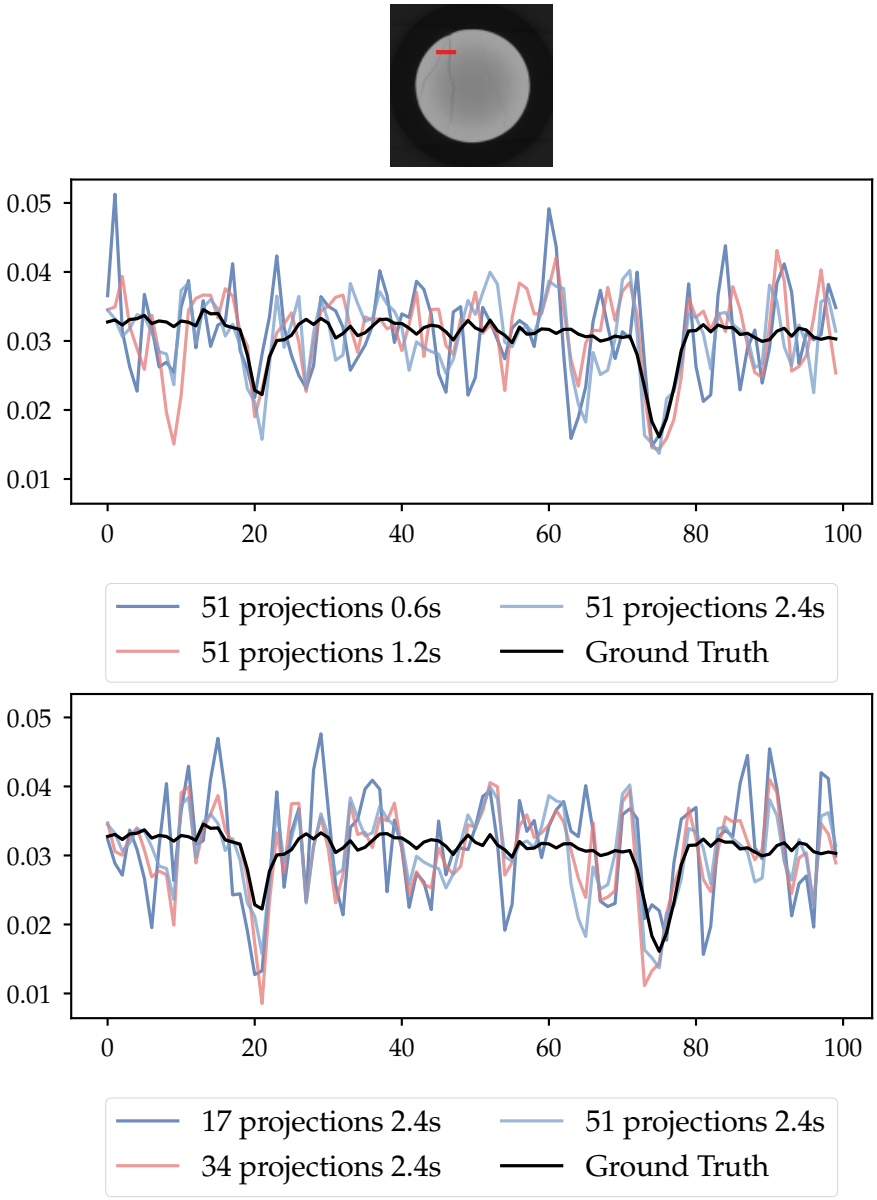


Figure 8.5: Profiles for undersampled reconstruction cross sections of two cracks in the sandstone. The red line in the top image indicates the location of the line profile. The black line indicates the ground truth line profile, while the coloured lines show the line profiles for the different reconstructions. Here, the max number of projections was combined with varying exposure times, and the max exposure time was combined with varying numbers of projections. Noise was generally reduced by increasing the variable parameter in the respective subfigures. The most accurate reconstruction of the crack was obtained with 51 projections.

Table 8.2: An overview of the experimental scan times and MSSIM values for the undersampled reconstructions in the region of interest marked by a red rectangle in Figure 8.3. The values in the respective matrices are indicated by the table headers. *Experimental scan time* is defined to be the measurement time given the settings of the performed CT scan.

Experimental Scan Time [s]				**MSSIM**			
Projections	**Exposures**			**Projections**	**Exposures**		
	0.6**s**	1.2**s**	2.4**s**		0.6**s**	1.2**s**	2.4**s**
17	46	56	76	**17**	0.141	0.137	0.208
34	91	112	153	**34**	0.143	0.168	0.255
51	137	168	229	**51**	0.154	0.191	0.316

Table 8.3: An overview of the achievable scan times and corresponding exposure ratios for different CT scan settings. The values in the respective matrices are indicated by the table headers. *Achievable scan time* is defined to be the scan time necessary to collect the given number of projections if only one revolution were performed.

Achievable Scan Time [s]				**Exposure Ratio**			
Projections	**Exposures**			**Projections**	**Exposures**		
	0.6**s**	1.2**s**	2.4**s**		0.6**s**	1.2**s**	2.4**s**
17	46	56	76	**17**	0.223	0.365	0.535
34	65	85	126	**34**	0.314	0.477	0.646
51	87	117	179	**51**	0.353	0.521	0.685

between reconstruction flexibility and information per time. This method will be referred to simply as *V3* or *golden ratio V3* in the following sections. Nevertheless, one crucial hyperparameter, in addition to the frame averaging presented in Section 8.2, was the number of projections per revolution p, as this contributed heavily to the minimal temporal resolution of the 4D-CT scan. Figures 8.6 and 8.7 include the temporal resolutions and the exposure ratio, meaning the ratio of total exposure time relative to the total scanning time, as a function of p, for the industrial and micro CT scanners, respectively. Figure 8.6 reveals that the industrial scanner has a temporal resolution below one minute for $p = 13$ at the cost of 31 % exposure time. However, the gain in exposure ratio is marginal for $p > 50$, as it is 53 % for $p = 75$. Likewise, Figure 8.7 reveals that the micro CT scanner has a temporal resolution approximately one and a half minute for $p = 13$ at the cost of 34 % exposure time. The difference in results between the two scanners is due to the typical scanning parameters mentioned in Section 4.1 and 4.3, as the micro CT scanner applies a lower rotation rate and a lower frame per second (fps) than its industrial counterpart.

The aspect of temporal resolution from performing several revolutions comes at the cost of fewer projections per time. The total scanning time is therefore increased compared to a standard CT scan, but not as much as what would be the case with golden angle sampling. Figure 8.8 compares the total scanning time for golden ratio V3, standard CT, and golden angle sampling. Here, 25 projections per revolution were used for golden ratio V3. Moreover, for golden angle sampling, both the theoret-

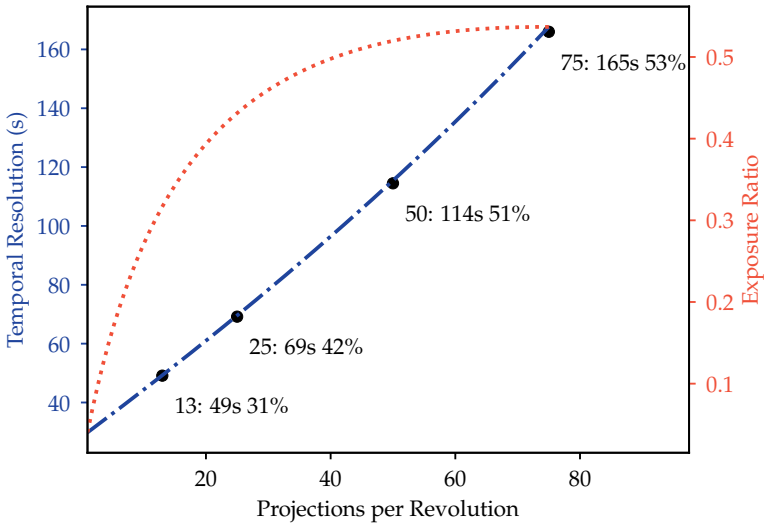


Figure 8.6: The scan settings of the *Industrial* CT scanner are assumed. Displayed are the temporal resolution and exposure ratio as a function of the number of projections per revolution p when applying the developed golden ratio scanning method. The black dots indicate the temporal resolution and exposure ratio at 13, 25, 50, and 75 projections per revolution.

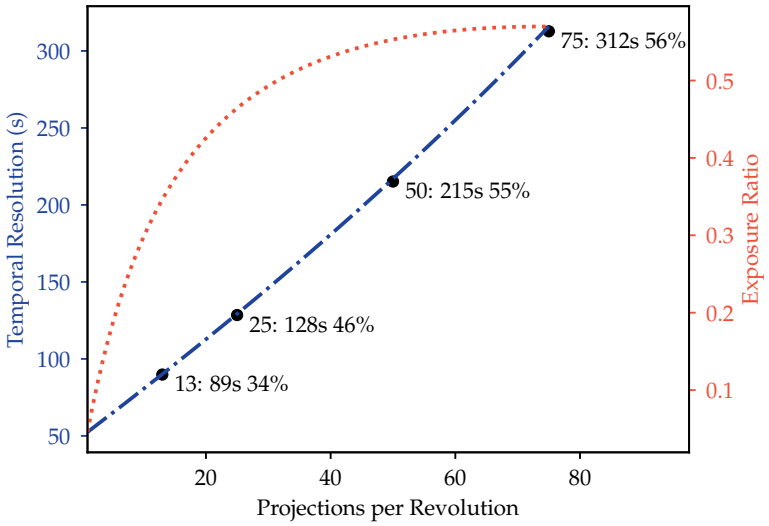


Figure 8.7: The scan settings of the *Micro-CT* scanner are assumed. Displayed are the temporal resolution and exposure ratio as a function of the number of projections per revolution p when applying the developed golden ratio scanning method. The black dots indicate the temporal resolution and exposure ratio at 13, 25, 50, and 75 projections per revolution.

ical achievable and the experimentally hardware-limited scanning times are included. Hardware limitations were rotating both clockwise and anti-clockwise to prevent intertwining of cables, as well as a software initiation and finalisation time of 5 s per projection. The case of the industrial CT scanner is included on the left, and the micro-CT scanner on the right.

An overview of the total scanning time for golden ratio V3 with varying number of revolutions as well as varying number of projections per revolution is shown in

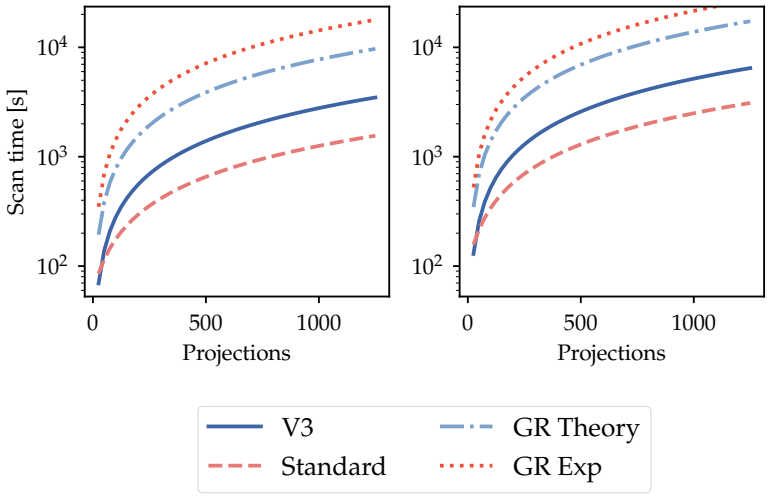


Figure 8.8: The golden ratio V3 procedure compared to standard sampling and golden angle sampling. The Industrial CT scanner settings are illustrated on the left, while the Micro-CT scanner settings are illustrated on the right. 25 projections per revolution were used for golden ratio V3. *GR Exp* accounts for hardware limitations that contribute to additional scanning time, while *GR Theory* is what in theory is achievable with golden angle sampling.

Figure 8.9. Likewise, the corresponding overview for the micro CT scanner is shown in Figure 8.10. The subplots on the right in both figures include the scanning times for different settings corresponding to a total of 1000 projections, and are essentially the diagonal in the respective subplots on the left. It is evident that many projections per revolution are advantageous for the total scanning time.

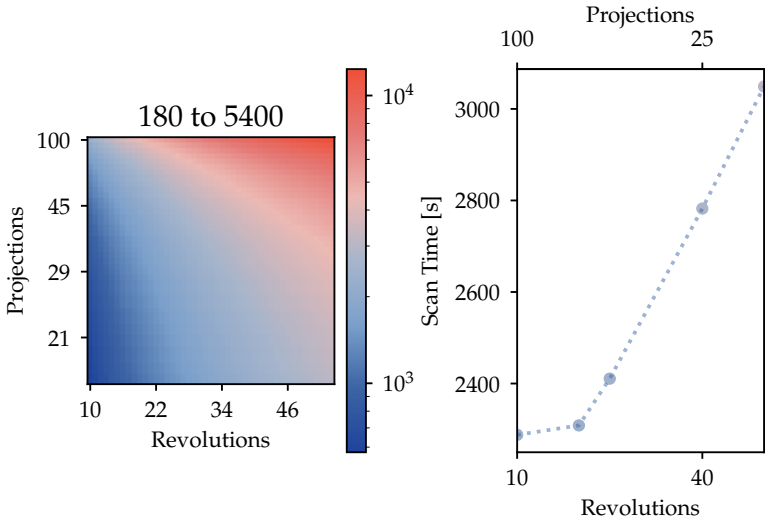


Figure 8.9: Industrial CT V3 Sampling Scan Times. On the left is a map of the typical scanning time for the golden ratio V3 procedure with varying number of revolutions and varying number of projections per revolution. From the map, the diagonal was retrieved and is shown on the right, where each points shows a scanning time for collecting 1000 projections.

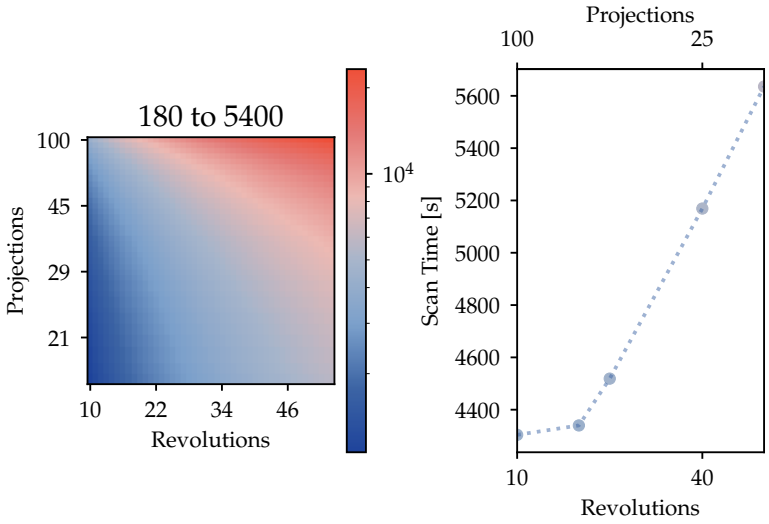


Figure 8.10: Micro-CT V3 Sampling Scan Times. On the left is a map of the typical scanning time for the golden ratio V3 procedure with varying number of revolutions and varying number of projections per revolution. From the map, the diagonal was retrieved and is shown on the right, where each point shown represents 1000 projections.

9 EVALUATION OF GOLDEN RATIO V3 SAMPLING

9.1 THE EFFECT OF EXPOSURE TIME FOR DYNAMICAL CT

Based on the visual inspection of Figures 8.4 and 8.5 together with the data in Tables 8.1 and 8.2, it is fair to state that the most significant factor for image quality is the number of unique projections, at least with this amount of undersampling. This result is expected if one considers the Fourier slice theorem (1.6), as many projections are required in order to fully sample the full two-dimensional Fourier transform of an object. This theory is experimentally shown when comparing the SSIM, PSNR, and the visual perception of the reconstructions; namely 34 projections of 0.6 s compared with 17 projections of 1.2 s, and 34 projections of 1.2 s compared with 17 projections of 2.4 s. Note that the number of projections p and the exposure time e will in some instances be listed in parenthesis to remind of the applied scan settings. In the first case, the global SSIM values increased from 0.204 ($p = 17, e = 0.6$ s) to 0.260 and 0.238 by doubling the projections and exposure time, respectively. On the other hand, the enhancement from doubling the exposure time slightly outperformed the enhancement from doubling the projections in the second case ($p = 17, e = 1.2$ s). In terms of MSSIM, the increase was in the first case insignificant in terms of projections, and actually negative, but insignificant, in terms of exposure time. The reason for the latter descrepency is unknown, as the level of noise is expected to improve with longer exposure time. However, only a single outlier is necessary to shift the intensity values in the renormalisation of the images before the MSSIM calculations. Additionally, the exposure comparison was only conducted for a single slice, and not a larger volume. Averaging over volumes could have cancelled out random variations. The second instance showed the opposite of the first instance, namely that the doubling of exposure time actually outperformed the doubling of projections in terms of MSSIM. Anyway, by doubling the projection number and exposure time, respectively, the PSNR values increased from 13.92 to 16.82 and 16.22 in the first case, and from 16.22 to 19.07 and 18.46 in the second case. This was an indication that given the current undersampling, the pixel-wise quality improvement from reducing the streaking artefacts outperformed the quality improvement from reducing the experimental poisson and gaussian random noise, as these were significant differences. It is also interesting how the PSNR values disagreed with the MSSIM values. In [31], some underlying tendencies for the different quality metrics were presented. Typically, PSNR is considered a poor metric for image quality, as it is heavily depended on changes in brightness and contrast, features which are not essential for human visual perception. Conversely, blurring is a feature that significantly affects human visual perception, but this is not well captured by PSNR [31]. All metrics are especially vulnerable to Gaussian noise. From considering these facts, there is reason to believe that the descrepency is due to the added aspect of structural similarity through the covariance term in MSSIM. However, the choice of reference must also be considered. One common ground truth was chosen as reference in all comparisons, namely the ground truth obtained with an exposure time of 2.4 s. From an additional calculations, where the metrics in Ta-

bles 8.1 and 8.2 were re-calculated with respect to the ground truth obtained within the same exposure time, the tendency changed. In this case, all metrics agreed with the fact that increasing the number of projections was the most significant factor for image quality. Therefore, there is a possibility that the MSSIM with 2.4 s exposure time is slightly over-estimated, because of a lack of independence to the reference reconstruction. Nevertheless, this reference is the closest reconstruction to the ground truth, and was therefore considered the best reference. Ideally, another scan with even longer exposure time should have been performed.

However, prioritising projections rather than exposure time comes at the cost of loss of temporal resolution, because increasing the number of projections affects the temporal resolution more than increasing the exposure time. In fact, the increases in scan time are 90 % and 45 % higher when doubling the number of projections than when doubling the exposure time, respectively. From this, some conclusive remarks can be made. For instance, one should not use a too short exposure time of 0.6 s when the number of projections is 51 or above. Instead, within the same achievable scan time, Table 8.3, one could perform the scan with 1.2 s exposure time and 34 projections. This alternative would provide a significant increase in exposure ratio, which resulted in higher SSIM, PSNR, and MSSIM. Furthermore, performing the scan with 2.4 s exposure time and 34 projections can be performed nearly as fast as 1.2 s exposure time and 51 projections, only 9 s slower per revolution. For this loss in temporal resolution, an exposure ratio of 0.646 compared to 0.521 may be achieved. This exposure ratio ensured the second highest quality metrics of the investigated cases, only surpassed by the setting with max exposure time and max projections. Essentially, based upon this analysis, one should not consider to reduce the exposure time before scanning with a severe undersampling factor. Stopping for more projections is associated with an overhead of stopping and starting the sample rotation, which affects the temporal resolution more than increased exposure time. Likewise, when performing a scan with very few projections, one should consider to increase the exposure time, as this would drastically increase the exposure ratio, and thus the image quality, with only a minor increase in scan time.

Moreover, when comparing Tables 8.1, 8.2, and 8.3, one can generally expect the best image quality from the settings with the most time invested in the scan. With except for the outliers of most projections combined with minimal exposure time, and vice versa, more time invested in the scan implies better image quality. Therefore, in order to obtain the best temporal and spatial resolution during a 4D-CT scan, one should confirm that the scan settings are a balanced in terms of exposure time and number of projections.

As a final note, it is important to realise that all reconstructions in Figure 8.4 were severely undersampled, as noisy outliers were the same order of magnitude as the important features one would like to resolve, as seen in Figure 8.5. This fact poses a challenge for the upcoming enhancement models.

9.2 UNRESOLVED CHALLENGES

As mentioned in Section 9.1, the image quality generally increases with the time invested in the scan. Therefore, the optimal sampling angle is the shortest angle that allows for an acceptable temporal resolution. The desired flexibility in the post-

processing of the measurement is another factor that has to be considered. Together with the optimal exposure time, this angle makes up one of two essential hyperparameters for the 4D-CT scan. The unresolved challenges of the derived golden ratio V3 procedure are that the number of projections per revolution and the exposure time have to be chosen *a priori*. This choice is not trivial, as it depends on numerous factors; the most important being the sample itself. The wider the sample is, the more projections are required to provide sufficient information at the edge. Additionally, a higher-quality scan is required to resolve finer details in the sample. The next essential factor is the dynamic process investigated. If the reconstructed time step is too large, dynamical artefacts will be introduced in the reconstructions. Fortunately, the golden ratio V3 procedure allows for including an arbitrary number of revolutions into each reconstructed time step. However, due to the quantisation into discrete revolutions, the number of projections per revolution has to be chosen carefully, as already mentioned. In this way, the golden ratio V3 procedure is a trade-off between golden angle sampling, which provides full flexibility, and standard CT, which provides no real flexibility except the worst-case scenario of having to perform missing wedge reconstructions 1.3.

9.3 PROPOSED IMPROVEMENTS TO THE PROCEDURE

An innovative proposition to the golden ratio V3 procedure is to perform two reconstructions per revolution. This requires no additional adjustments to the procedure; only the post-processing would need modifications. Even though the beam geometry of a laboratory CT source is a diverging cone beam, and not a parallel beam like a synchrotron source, it may still be worth performing two missing wedge reconstructions with 180° coverage per revolution. Therefore, one can effectively double the number of projections per revolution relative to the original golden ratio V3 procedure. As known, the number of projections per time increases with the number of projections per revolution, and many projections per revolution is beneficial as long as the temporal resolution requirements of the scan are met. Consequently, both the temporal resolution and the exposure ratio of the scan are increased with this modification. The temporal resolution is increased, because the amount of rotation per reconstruction is reduced, and the exposure ratio is increased, because the number of projections collected per reconstruction remains the same. This gain comes, however, at the cost of possibly introducing missing wedge artefacts. Nevertheless, with a proposed coverage of 180° per reconstruction, the missing wedge artefacts may turn out to be sufficiently weak to be removed by post-processing enhancement, for instance performed by a GAN, even without specialised training.

Another proposition is to perform binning of the projections, meaning that multiple pixels in the detector are summed together. This has two main advantages. Firstly, the signal to noise ratio is increased, as the number of photons per pixel is increased, while the random noise with zero mean will be reduced following a cancellation effect. Secondly, the number of projections necessary for fully sampling of the object is reduced, as the effective width of each pixel is increased. As an example, a 2×2 binning of the projections reduces the number of projections required by a factor of two. The drawback of binning is the reduced spatial resolution, but in cases where the noise due to aliasing is a dominant factor, binning should be one of the first

considerations in order to improve reconstruction quality. In other words, binning allows for a higher temporal resolution. In cases with serious aliasing, binning might actually improve the actual spatial resolution as well, because the aliasing effect, that priorly to binning destroyed the spatial resolution, would be reduced.

10 TEMPORAL CT

10.1 4D-CT OF HOUR GLASS

For the 4D-CT experiment, the hourglass sample was tilted to distort the otherwise cylindrical symmetry, and scanned using the derived golden ratio sampling procedure with 13 projections per revolution. A total of 55 revolutions were performed, resulting in 715 projections in approximately 45 min. As estimated in Section 8.3, the maximal temporal resolution of one full revolution was therefore 49 s. However, a temporal resolution below 1 min came at the cost of significant spatial noise and undersampling artefacts. To compensate for poor spatial resolution, the mentioned GAN was optimised and trained on a diverse dataset of both simulated and experimental reconstructions, using an undersampling factor of 8. The generator found weights that resulted in satisfactory validation loss, structural similarity, and pleasing human perception. This section will be devoted to investigating the dynamical processes of the 4D-CT experiment, and demonstrate the enhancement provided by the trained GAN. Moreover, the absolute limit of temporal resolution will be determined and studied for the hourglass sample, despite an expected increase in undersampling artefacts.

Firstly, Figures 10.1 and 10.2 illustrate the necessary compromise between temporal resolution and spatial resolution. The figures compare the resulting reconstruction from FDK including all 715 projections, and 13 projections of FDK with and without GAN-enhancement, respectively. It is evident that the fully sampled reconstruction suffered from severe dynamical artefacts, while the spatial undersampling artefacts were significant at the hourglass boundaries in the undersampled reconstructions.

Next, the temporal resolution was investigated further. In this case, the fully sampled reconstruction was disregarded entirely. A measure of good temporal resolution in the case of the hourglass sample was the visibility of the sand-air interfaces. The interface, or edge, was highlighted using a Sobel-filter (3.20), representing the first gradient, and a Laplacian-filter (3.21), representing the second gradient. Figure 10.3 corresponds to Figure 10.1, but compares the gradient image of the FDK reconstruction to the FDK-GAN reconstruction. The red box indicates the RoI for the following analysis.

A slight difference may be apparent in the gradient images, but further investigation of a RoI was deemed necessary. Therefore, the sand-air interface was cropped, and is presented in Figures 10.4 and 10.5 in the case of Sobel and Laplacian edge detection, respectively. Moreover, the slice presented represents one slice in a cropped volume, for which the Sobel and Laplacian intensity distributions were plotted in Figure 10.6. With GAN-enhancement, the Sobel-filtered interface was more visible due to a better signal-background contrast. Moreover, less high-frequency noise was present for FDK-GAN, as seen especially in the Laplacian-filtered interface, where the noise dominates the FDK reconstruction. These observations were confirmed by the intensity distributions, since the FDK-GAN Sobel distribution had generally both higher and lower gradient values, as well as a narrower Laplacian distribution.

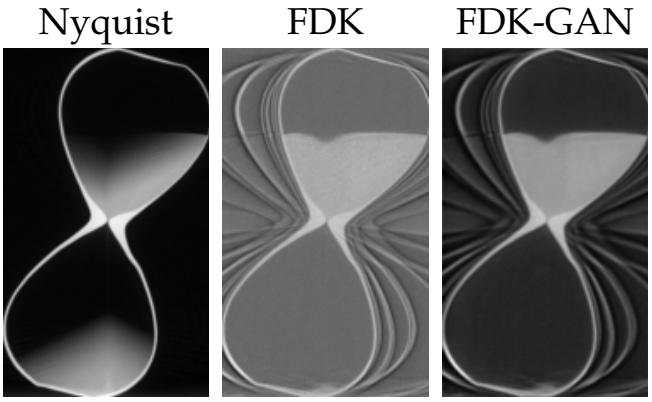


Figure 10.1: An illustration of how undersampled FDK captured the dynamical process, while fully sampled FDK suffered from dynamical artefacts. The GAN-enhanced reconstruction was able to capture the dynamical process with better contrast and less noise, but undersampling artefacts outside the boundaries remained.

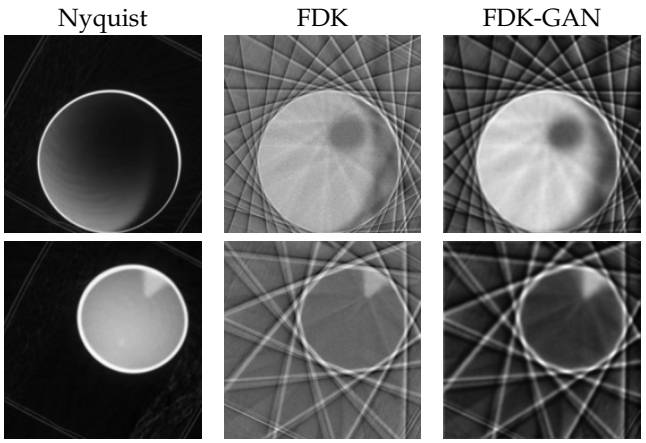


Figure 10.2: Cross section of the hourglass top and bottom early on in the dynamical process. As in Figure 10.1, the undersampled FDK reconstruction captured the dynamical process, while the fully sampled FDK reconstruction suffered from dynamical artefacts.

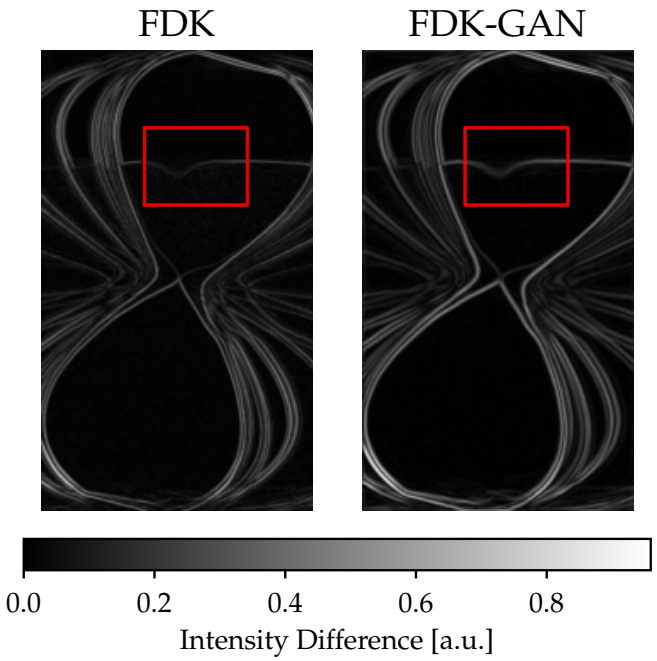


Figure 10.3: A Sobel filter was applied to Figure 10.1 to highlight the sand-air interface. The FDK-GAN reconstruction had a better contrast at the sand-air interface, but also more present undersampling artefacts. The red box indicates the height and width of the RoI for the analysis to follow.

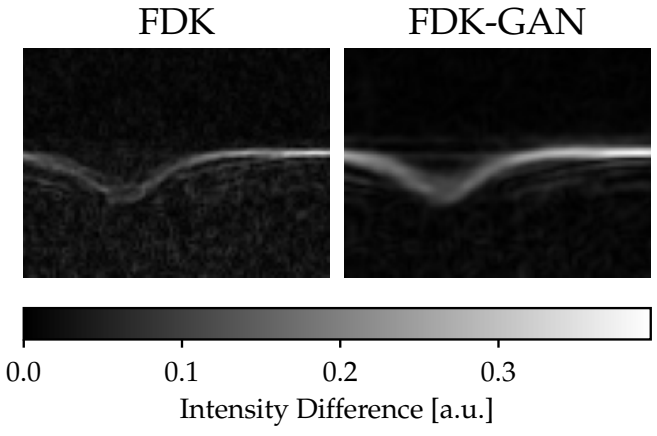


Figure 10.4: From Figure 10.3, the sand-air interface was investigated further. The included slice is a part of a cropped volume surrounding the mentioned interface. The FDK-GAN reconstruction had a better contrast at the sand-air interface and less high-frequency noise.

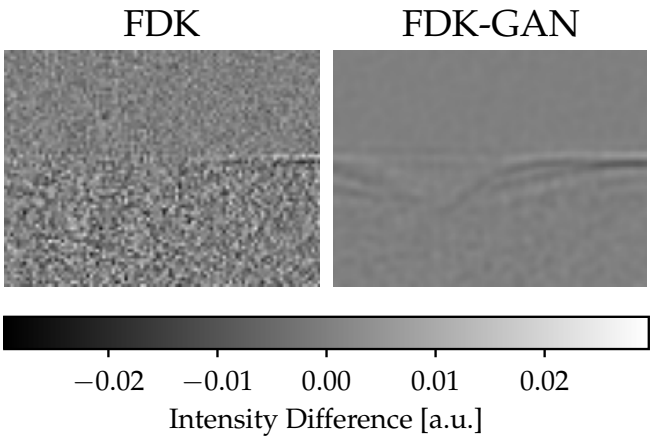


Figure 10.5: Similarly to Figure 10.4, the sand-air interface was investigated, this time applying a Laplacian filter. This filter represents the second derivative, and is therefore more sensitive to noise. Therefore, the observation of less high-frequency noise in the FDK-GAN reconstruction is more apparent, and supports the observation in Figure 10.4.

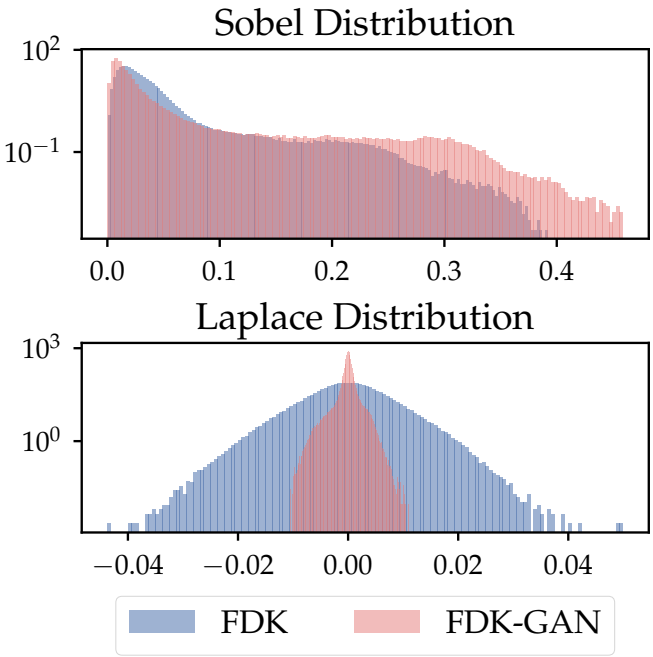


Figure 10.6: The intensity distributions of the cropped sand-air interface volume, illustrated by Figures 10.4 and 10.5. Along the horizontal axis are the respective gradient values, and along the vertical axis are the densities of occurances. With FDK-GAN, the Sobel-filtered interface had a higher zero-gradient peak, as well as a greater dynamical range. This indicates a better signal-background contrast. Meanwhile, the FDK-GAN Laplacian-filtered interface had a narrower distribution, indicating less high-frequency noise, which explains the observation in Figure 10.5.

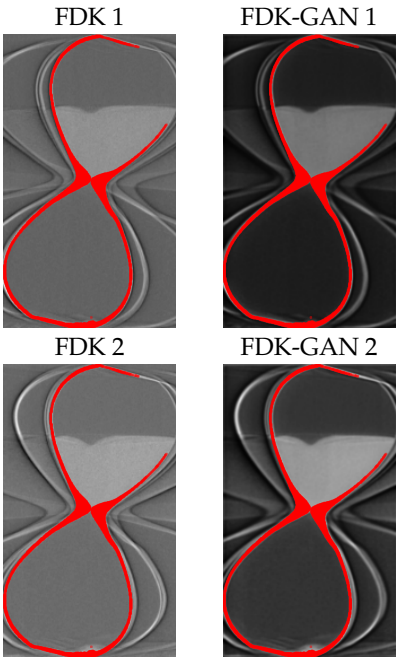


Figure 10.7: Missing wedge reconstructions. One full revolution of projections was divided into two reconstructions, and thereby two time stamps. The red shape marks the real hourglass shape without undersampling artefacts. As seen in all figures, FDK-GAN provided contrast enhancement and noise reduction.

In an effort to further maximise the temporal resolution, missing wedge reconstruction was performed. This was performed regardless of the resulting undersampling artefacts, as the aim was to visualise the sand-air interface with minimal temporal blurring. Due to the reduced scan time, the gradient visibility was expected to be increased. The mentioned reconstruction from one full revolution was divided into two time stamps with 7 projections each, with the middle projection being shared. The inner slices of the mentioned wedge reconstructions are shown in Figure 10.7. Due to the double hourglass edges introduced by the heavy undersampling, the static hourglass shape from the fully sampled reconstruction was drawn in red.

Despite the undersampling, few artefacts were present inside the hourglass, and it was still possible to distinguish the sand-air interface. Figure 10.8 compares the gradient in the different time stamps, both with and without GAN-enhancement. The difference between the time stamps was also of interest, and is included in Figure 10.9. This time the GAN-enhancement was able to increase the gradient visibility, and the difference between the time stamps was more pronounced. It should also be noted that the gradient was more intense in the wedge reconstruction, Figure 10.8, than the full revolution reconstruction, Figure 10.3, which agreed with the proposed hypothesis that enhanced temporal resolution would create a more intense sand-air interface. More importantly, however, was the resolved sand-air interface in the second time stamp, which was not present in the full revolution reconstruction. As opposed to the information in Figure 10.4, where the curved part of the sand-air interface was blurred in time, the missing wedge reconstruction was able to resolve two distinct curvatures, thereby uncovering how the whirlpool in the hourglass propagated.

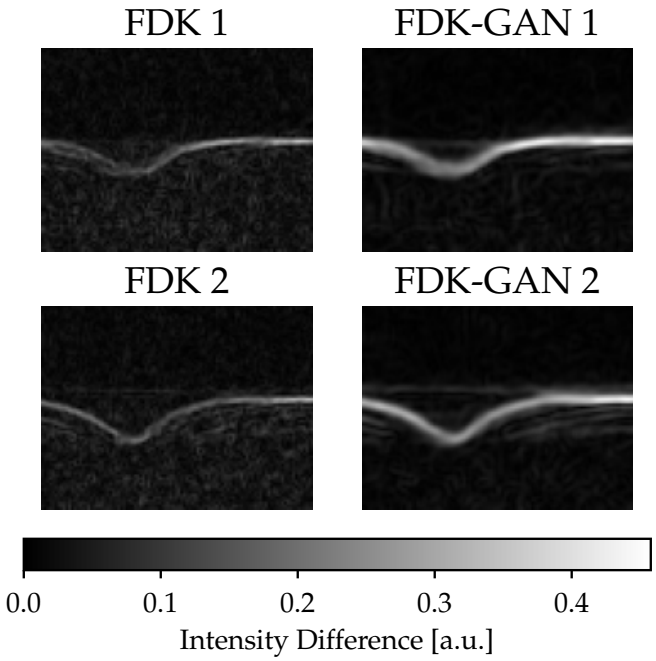


Figure 10.8: The sobel filter was applied to the sand-air interface in the wedge reconstructions. and the dynamical process of the sand-air interface was enhanced. It is clear that the interface has descended in the second time stamp, with an accelerated drop in the centre. The additional temporal resolution also uncovered the sand-air interface presented in the second time stamp, which was not resolved in the full revolution reconstruction.

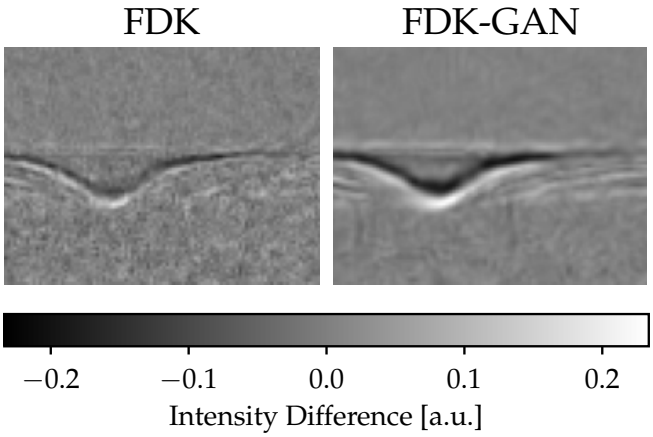


Figure 10.9: To further illustrate the movement of the sand-air interface, the difference between the two time stamps was calculated. The dark line illustrates the presence of a sand-air in the first reconstruction, while the bright line shows where the sand-air interface has descended in the second reconstruction.

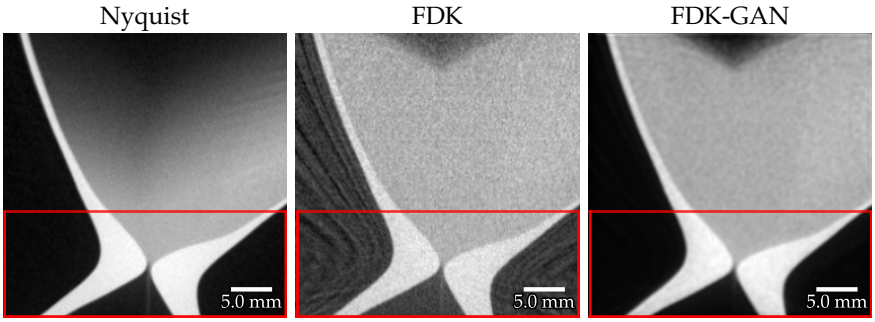


Figure 10.10: Vertical slice of the hourglass neck, reconstructed from FDK on all 715 projections, FDK on 52 projections, and GAN-enhanced FDK on 52 projections, respectively. The upper red line marks the one retrived line profile, and the box marks the region of interest from which quality metrics were calculated.

10.2 RECONSTRUCTION ENHANCEMENT OF HOURGLASS

In addition to the investigation of the dynamical processes of the 4D-CT experiment, the spatial quality of the GAN-enhanced reconstructions was also assessed. The generator found, as mentioned, weights that resulted in satisfactory validation loss, structural similarity, and pleasing human perception. A 256 cubed region of interest was enhanced using the trained generator, and no binning of pixels was performed, unlike Section 10.1.

Figure 10.10 shows a vertical slice of the hourglass neck reconstructed from FDK including all 715 projections, FDK including the first 52 projections, and GAN-enhanced FDK including the first 52 projections, respectively. The given number of projections was effectively an undersampling factor of 14 with respect to the ground truth, or 8 with respect to the width of the region. Since this undersampling factor was the same as the one applied during training, the peak performance of the GAN was expected at this number of projections. The fully sampled slice contained a fade in the upper part of the hourglass, indicating dynamical effects that have not been captured due to the long scan duration. In contrast, the undersampled FDK reconstruction captured the early phase of the dynamical process, but suffered from severe undersampling artefacts, and much high-frequency noise. These artefacts were reduced in the GAN-enhanced FDK reconstruction, as nearly no undersampling artefacts remained. One important aspect of the data post-processing was the normalisation of data. In comparison of slices from the three reconstructions, the GAN-enhanced reconstruction was re-normalised into the range of the fully sampled reconstruction.

When assessing an almost static cross section of the hourglass, the importance of reconstruction enhancement was demonstrated. As shown in Figure 10.11, the trained GAN removed undersampling artefacts, high-frequency noise, and increased the MSSIM from 0.37 to 0.80. Further visualisation of the elimination of high-frequency noise due to GAN-enhancement is shown in Figure 10.12, where the line profile of the neck is plotted for the three reconstructions. The enhancement displayed a lower variance in the line profile, especially in the low-intensity region, at the same time as the contrast was enhanced. To evaluate the independent pixel-wise loss, the re-

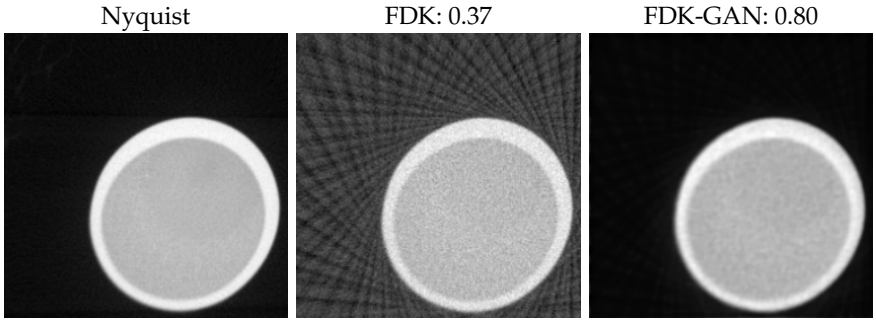


Figure 10.11: A static cross section in the early phase of the 4D-CT scan. Because this region remained static, the fully sampled Nyquist reconstruction may function as a ground truth. The GAN ensured an increase in MSSIM from 0.37 to 0.80 for the FDK reconstruction consisting of 52 projections.

spective difference maps of the static cross sections are shown in Figure 10.13. The difference maps subtracts the undersampled reconstructions from the ground truth, and the metric included is the peak signal-to-noise ratio. Here, the GAN-enhanced reconstruction displayed a only slighter lower pixel-wise loss than the undersampled reconstruction. The intensity values were generally slightly low compared to the ground truth, which may be due to the fact that the FDK-GAN applied some blurring, thereby reducing the max intensities, which would affect the renormalisation of the GAN-enhanced reconstruction.

As mentioned, the enhancements presented had the same undersampling factor as the training dataset. However, it would be beneficial to investigate the performance of the GAN when applied to a higher and lower undersampling factors. Therefore, the MSSIM, global SSIM, and PSNR were calculated before and after enhancement for undersampled reconstructions ranging from 1 to 8 revolutions. The global SSIM in the volume marked by the red box in Figure 10.10 is plotted in Figure 10.14. Additionally, the MSSIM score was averaged over all slices in the volume. To illustrate the increase in performance, Figure 10.15 relates the SSIM, PSNR, and MSSIM of the GAN-enhanced reconstruction to the FDK reconstruction by calculating the relative increase in the respective metrics.

Finally, an analysis of the frequency content of the reconstructions was performed. The Fourier Sheell Crosscorrelation (FSC) (3.19) was calculated for all mentioned number of projections before and after GAN-enhancement. The left subplot in Figure 10.16 presents the FSC curves for the undersampled FDK reconstruction, and the right subplot presents the FSC curves for the GAN-enhanced reconstruction.

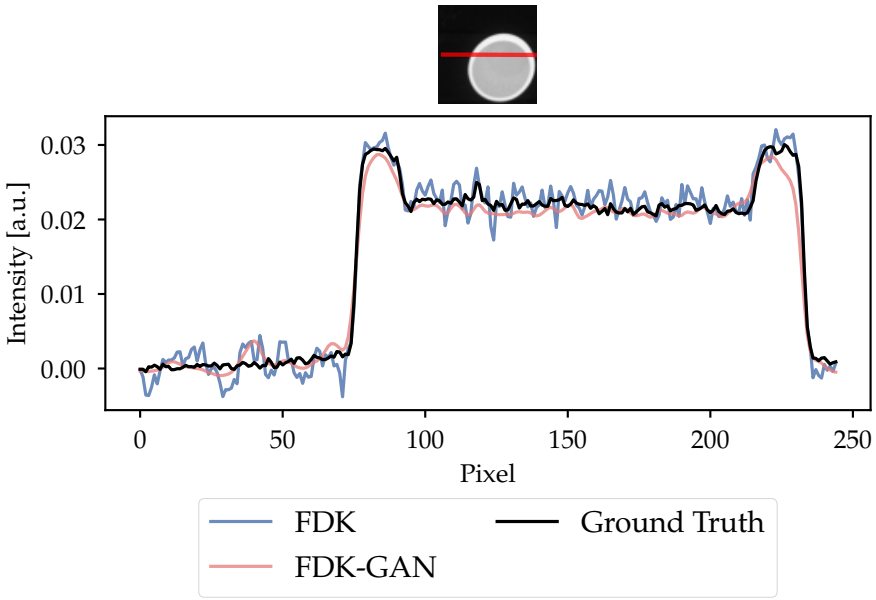


Figure 10.12: A line profile across the static cross section of the hourglass neck, as shown in Figure 10.11. 52 projections were used for the undersampled reconstructions, compared to 715 projections for the fully sampled reconstruction. Enhanced contrast and blurring of outliers were observed in the GAN-enhanced reconstruction. Therefore, features and high-frequent noise were in most cases not the same order of magnitude, which was the case for the undersampled reconstruction.

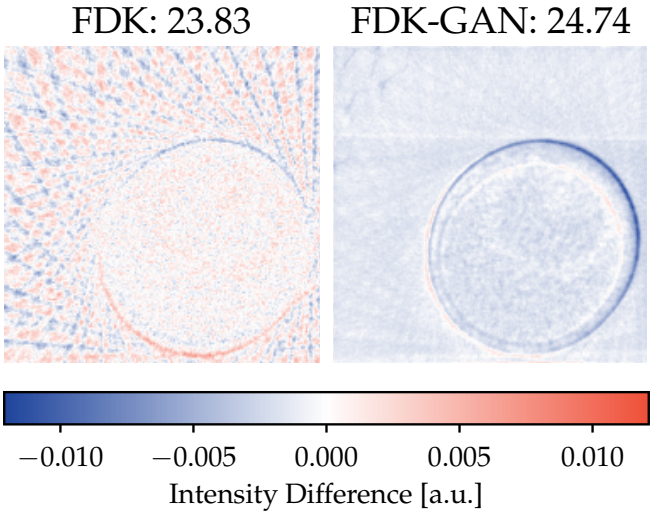


Figure 10.13: PSNR illustrated by a difference map of the pixel-wise loss between the undersampled reconstructions and the ground truth, as shown in Figure 10.11. As also illustrated in Figure 10.12, the GAN-enhanced reconstruction ensured a higher PSNR than the undersampled reconstruction, but the blurring effect resulted in a generally lower pixel intensity than the ground truth.

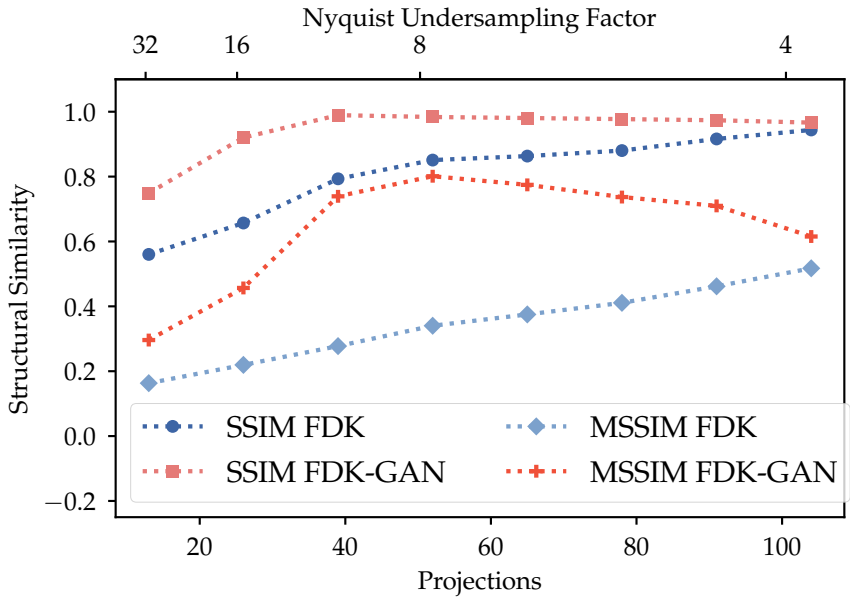


Figure 10.14: The global SSIM was calculated for the volume marked by the red box in Figure 10.10. For each cross section in this volume, the MSSIM was calculated, and the average MSSIM across all slices is shown. The optimal performance was observed close to the training undersampling factor of 8.

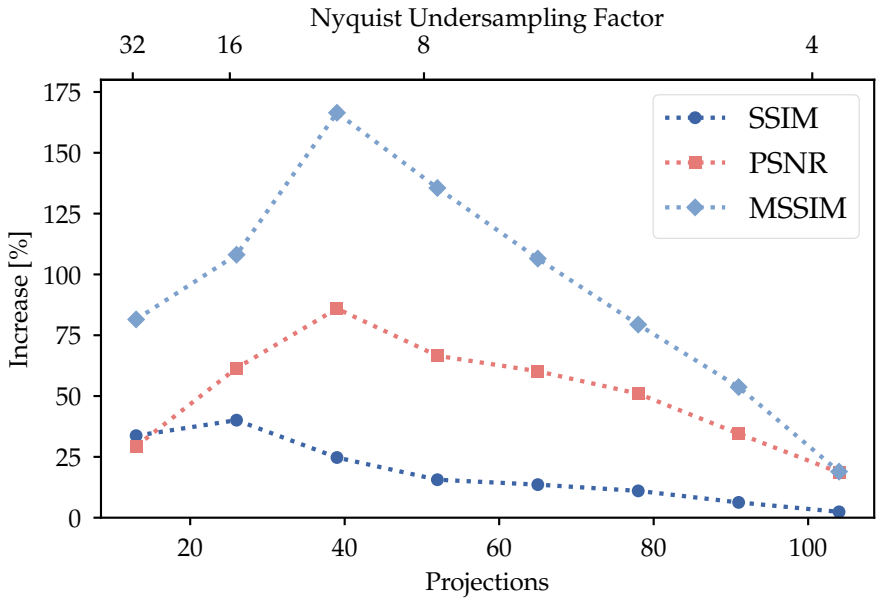


Figure 10.15:
The global PSNR was calculated for the volume marked by the red box in Figure 10.10. The scores for SSIM and MSSIM were calculated as described in Figure 10.14. From the data, the respective performance increases were calculated, and are shown. Again, the optimal performance was observed around the training undersampling factor of 8.

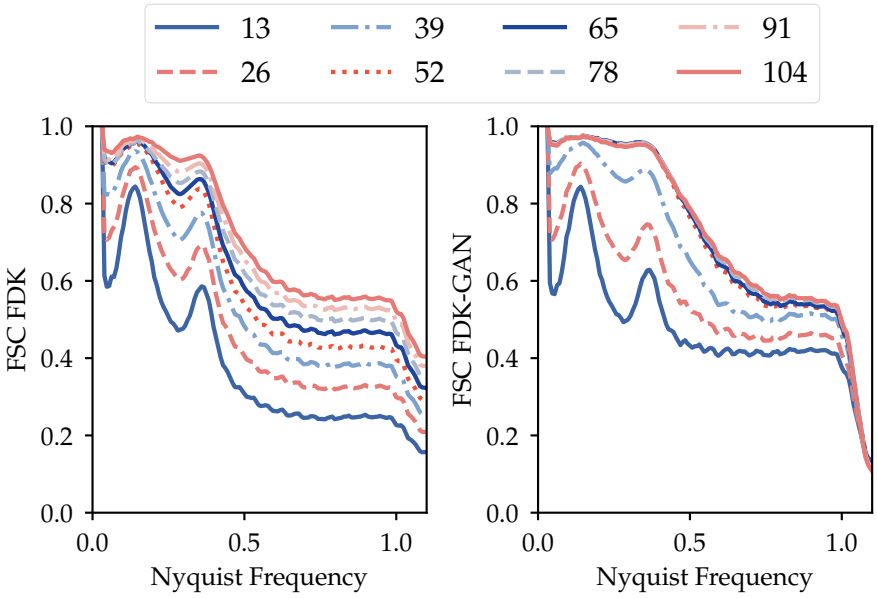


Figure 10.16: FSC calculated for different number of projections before and after GAN-enhancement. The volume represented by the red box in Figure 10.10 was used for the analysis. Processing steps included three-dimensional Fourier transform, shifting the zero-frequency point to origo, and calculating the FSC for each unique integer spatial frequency. Moreover, a floating mean of the FSC was plotted, using the 4 prior and posterior frequencies at each point. This was done to reduce the noise in the plot. FDK-GAN reconstructions had generally higher correlation at lower frequencies up to the Nyquist frequency, where the spatial frequencies suddenly became uncorrelated.

Table 10.1: The SIRT algortihm managed to produce reconstructions with the following SSIM, MSSIM, and PSNR scores. The algortihm was limited to 100 iterations, which corresponded to a computational time of less than 20 min using GPU-acceleration. The volume represented by the red box in Figure 10.17 was used for the analysis.

	SIRT		
Projections	**SSIM**	**MSSIM**	**PSNR**
26	0.94	0.36	22.23
52	0.96	0.39	23.44

10.3 HOUR GLASS ITERATIVELY RECONSTRUCTED

An alternative to analytical reconstruction is iterative reconstruction, which is known to produce less noise and undersampling artefacts when the undersampling is high. Therefore, a similar analysis to the one performed in Section 10.2 was performed on reconstructions of the hourglass using 100 iterations of the SIRT algorithm for 26 and 52 projections, respectively. Initially, only 50 iterations of were performed, but the reconstructions were still early in the convergence process after 10 minutes of computation. The reconstructions with 100 iterations were completed in less than 20 min using GPU-acceleration, which was deemed an acceptable computational time. Correspondingly to Figures 10.10, 10.11, 10.12, and 10.13, the iterative reconstruction data are shown in Figures 10.17, 10.18, 10.19, and 10.20, respectively. The upper red line in Figure 10.17 indicates the cross section shown in Figure 10.18. Furthermore, the encapsulated region within the red box illustrates again the height and width of the volume which is considered static, and thus was used to calculate the global SSIM, PSNR, and MSSIM in Table 10.1, where the latter metric is an average over all two-dimensional slices in the volume. Generally, the iterative reconstruction performed better than the analytical reconstruction, as expected. This came at the cost of increased computational time, less contrast and blurred edges. The latter is visualised both at the sharp interface between the hourglass and the background in Figure 10.19, and in the difference map in Figure 10.20, where there was a positive difference right outside the hourglass boundaries, and a negative difference inside the hourglass.

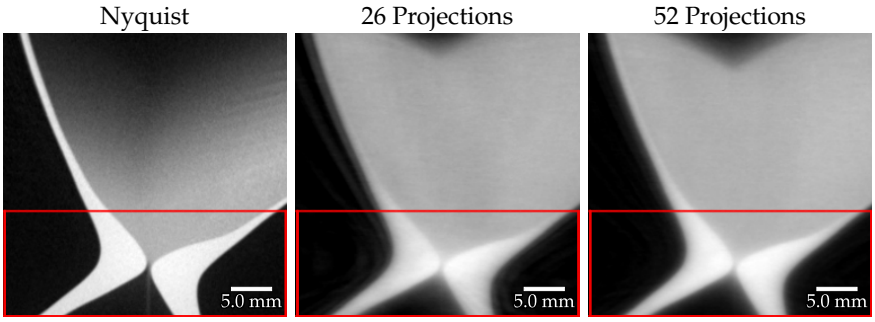


Figure 10.17: Vertical slice of the hourglass neck, reconstructed from FDK on all 715 projections, SIRT on 26 projections, and SIRT on 52 projections, respectively. It is important to remember that blurring is typically included in iterative reconstructions, as shown in this figure. Only 100 iterations were performed in order to complete the reconstructions within 20 minutes each.

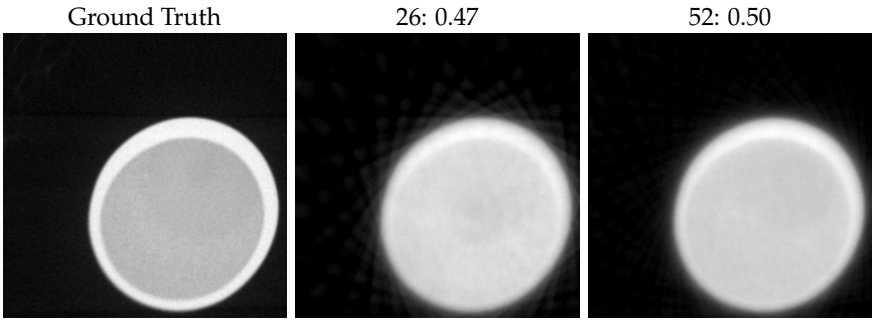


Figure 10.18: A static cross section in the early phase of the 4D-CT scan. Because this region remained static, the fully sampled Nyquist reconstruction may function as a ground truth. The MSSIM values with respect to the ground truth are listed. Including more projections resulted in a more accurate reconstruction, as shown in the figure. However, there is a probability that too few iterations were included in the SIRT reconstructions, but only 100 iterations were performed in order to complete the reconstructions within 20 minutes each.

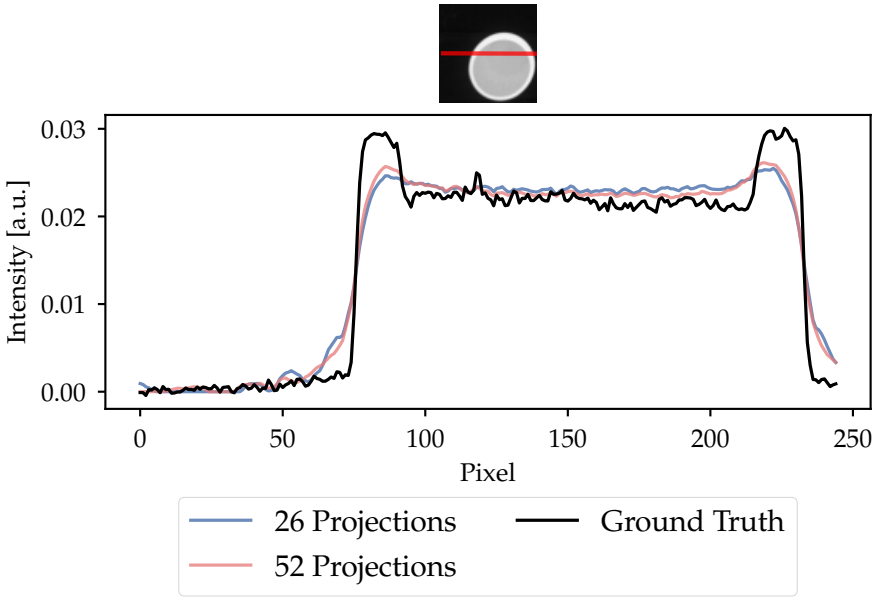


Figure 10.19: A line profile through the static cross section shown in Figure 10.18. Both reconstructions suffered from blurring and smooth edges rather than sharp transitions and good contrast. However, the reconstruction with 52 projections was better in this regard than the reconstruction with 26 projections. 100 iterations were performed for the SIRT reconstructions.

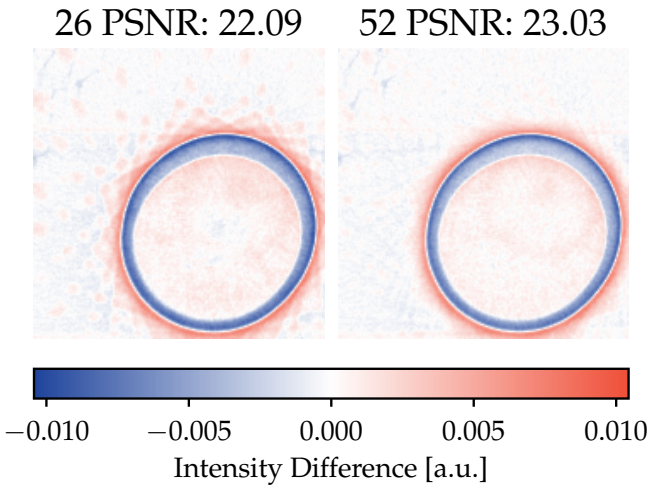


Figure 10.20: The difference map between the fully sampled Nyquist reconstruction and the SIRT reconstructions. The smoothed out edges became more apparent in this figure, as the region outside the hourglass had a positive difference, while the region immediately inside the hourglass had a negative difference.

10.4 ENHANCEMENT OF UNDERSAMPLED SANDSTONE RECONSTRUCTION

The enhancement ability of the GAN trained on a diverse set of simulated and experimental data was assessed by applying it to a static CT reconstruction of the sandstone sample. It is important to understand that the weights of the generator were the same in this section as in Sections 10.1 and 10.2. The ground truth was obtained by reconstructing the full dataset, 935 projections divided into 55 revolutions of 17 projections, utilising the analytical FDK-algorithm. Figure 10.21 compares the ground truth to the undersampled and enhanced reconstructions, respectively. Added to the comparison is a reconstruction using the iterative SIRT algorithm with 150 iterations. A higher number of iterations than in Section 10.3 was deemed necessary because of the complexity of the sandstone sample relative to the hourglass sample. The number beside the reconstruction technique indicates the MSSIM compared to the ground truth, and it becomes clear that the SIRT algorithm benefitted from the increased number of iterations, as it achieved the highest MSSIM. As in Section 10.2, the GAN outputted a reconstruction with better contrast, less high-frequency noise, and better visual perception than the undersampled FDK reconstruction.

The effectiveness of the GAN at different undersampling factors was also investigated. The result is shown in Figure 10.22, where the SSIM and the MSSIM values for reconstructions with varying number of projections are plotted. Interestingly, the span in which the GAN was effective was much narrower for the sandstone sample than the hourglass sample. It provided an enhancement of 25 % to 60 % when the undersampling factor was above 8, while the reconstruction actually worsened for lower undersampling factors.

Figures 10.23, 10.24, and 10.25 further exemplify the noise-canceling ability of the GAN by comparing the line profiles of the ground truth, the undersampled reconstruction, the iterative reconstruction, and the enhanced reconstruction for different lines of interest. It is evident that the SIRT reconstruction also reduced the noise, but at the cost of blurring the image, while contrast-enhancement was achieved by the GAN. At the same time, the GAN had a tendency to enhance outliers of the FDK reconstruction as well, making the line profile somewhat unstable. The Iterative reconstruction, on the other hand, correctly reconstructed most parts of the sandstone, but the intensity drop within cracks was not nearly as sharp as in the ground truth.

This instability is apparent in Figure 10.26, which illustrates the pixel-wise loss between the ground truth and the undersampled reconstructions. As shown in Figures 10.23, 10.24, and 10.25, the SIRT algorithm managed to minimise the pixel-wise loss, only struggling in region of sudden change, while the two other algorithms had a higher noise-level.

Other training sessions were also performed, where the GAN found other weights that produced better results for the sandstone sample. However, this came at the cost of the peak performance for the hourglass sample presented in Section 10.2. These training sessions are presented in Appendix C, and include training exclusively on another CT scan of the sandstone sample, reconstructed from 50 projections. Moreover, a training session using spatial dimensions 64^3 with batchsize 4, as apposed to the 128^3 and batchsize 1, is also included. These results will be elaborated upon in the mentioned appendix, but shortly summarised it was confirmed that satisfactory enhancement was possible for the sandstone sample, as well.

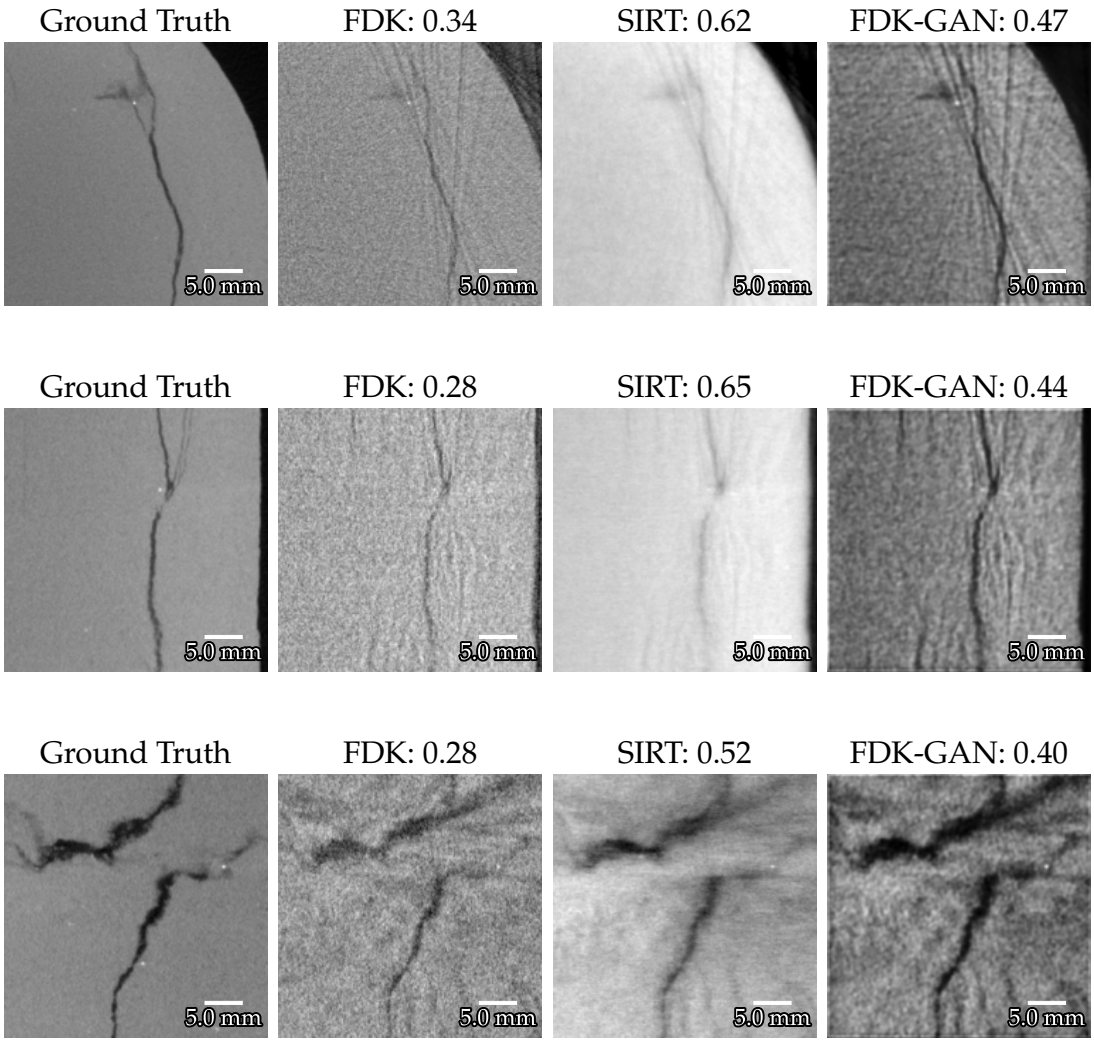


Figure 10.21: The figure compares the fully sampled FDK reconstruction to the reconstructions obtained from 34 projections utilising the FDK algorithm, the SIRT algorithm, and the FDK-GAN procedure. As shown, the GAN was able to reduce some noise, but has enhanced streaking artefacts as well as vital details. After 150 iterations, the SIRT reconstruction was able to reduce noise at the cost of blurring the image, and achieved the highest MSSIM, which is the value displayed together with the reconstruction method.

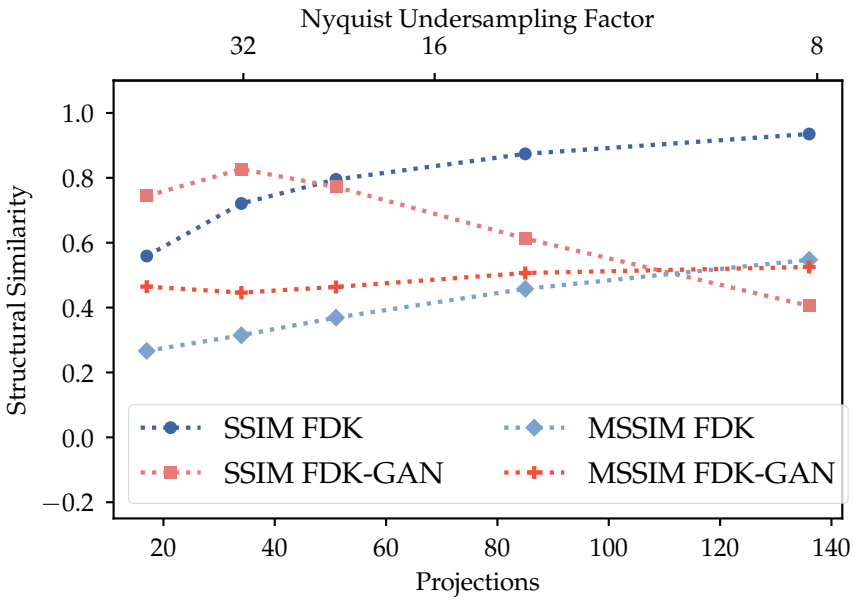


Figure 10.22: The global SSIM score of the region of interest, as well as the MSSIM score from convolving with a 11×11 Gaussian kernel. Enhancement was actually only observed for undersampling factors above 8, while the reconstruction worsened for lower undersampling factors. In terms of MSSIM, the GAN provided an output MSSIM score of approximately 0.46 to 0.50 regardless of the quality of the input. This tendency is in stark contrast to the hourglass sample. However, GAN weights that provided better results for the sandstone sample were also obtained, and will be presented in Appendix C.

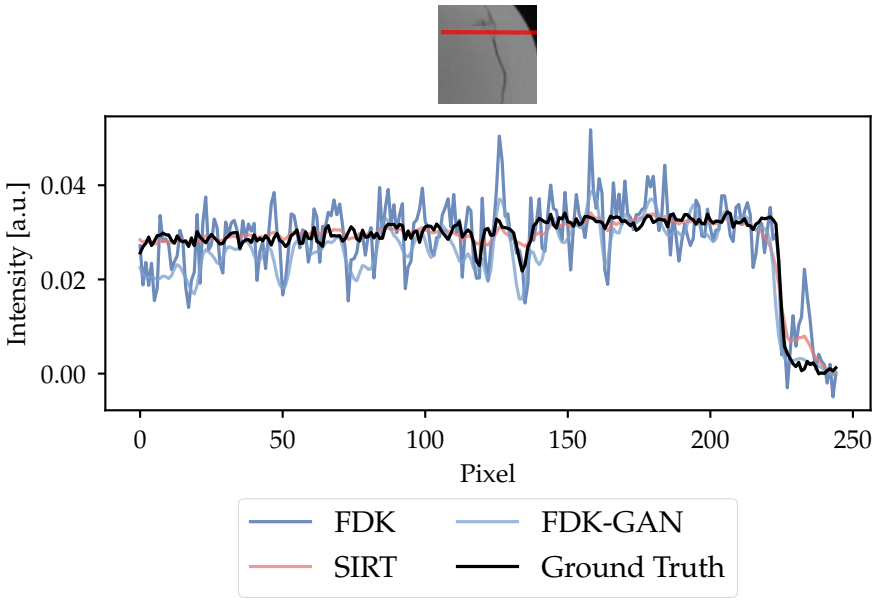


Figure 10.23: The GAN was able to provide some noise-cancelling, but not in the same way as the SIRT algorithm. However, the GAN was able to enhance the contrast of the undersampled reconstruction, even though it also enhanced some of the outliers. FDK, SIRT, and FDK-GAN were all reconstructed from 34 projections.

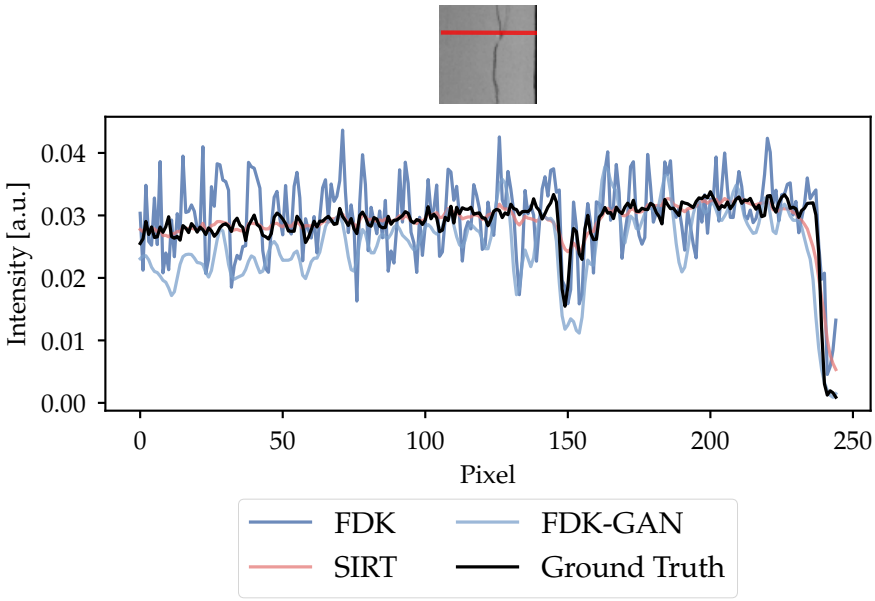


Figure 10.24: The GAN was able to provide some noise-cancelling, but not in the same way as the SIRT algorithm. In this instance, elements of the effects of the GAN outperformed the SIRT algorithm, like enhancing the cracks and presenting the rough texture of the sandstone. This was not captured by the SIRT algorithm, which instead blurred the image. FDK, SIRT, and FDK-GAN were all reconstructed from 34 projections.

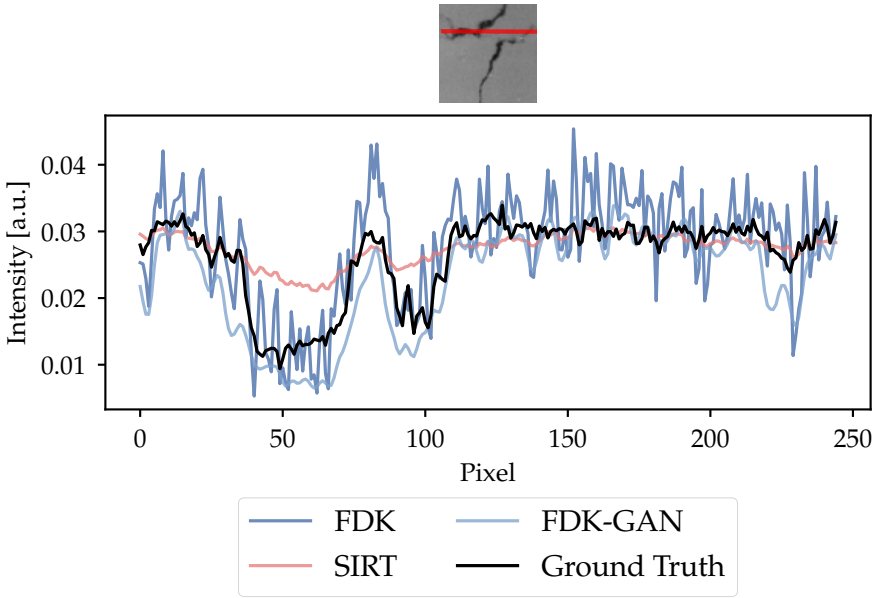


Figure 10.25: A line profile that exposes the weakness of the SIRT algorithm. Even though the line profile of the ground truth varied rapidly, the SIRT algorithm reconstructed a smooth line profile with much less variance. The GAN followed the ground truth more closely, but also had a higher noise-level. FDK, SIRT, and FDK-GAN were all reconstructed from 34 projections.

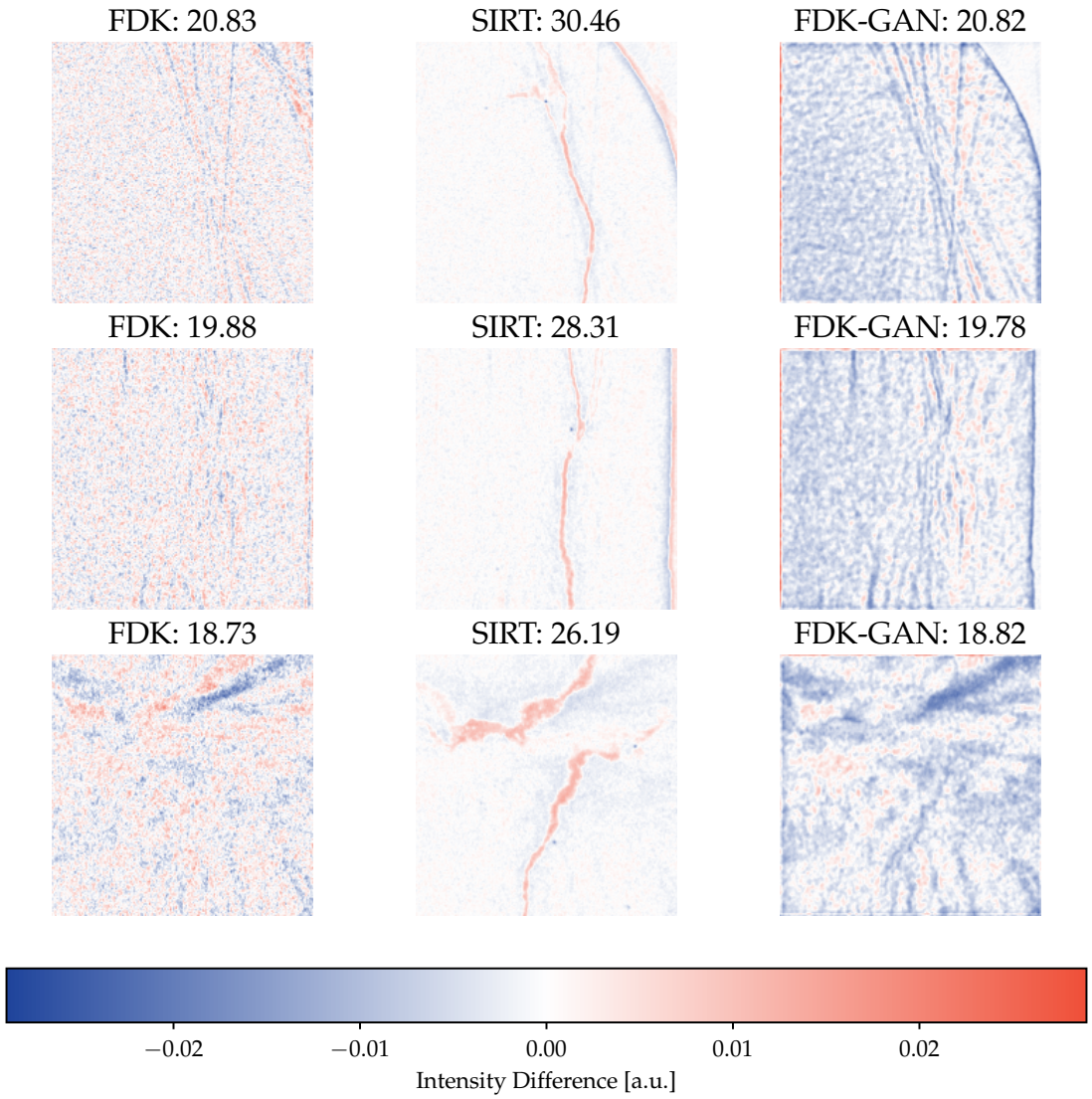


Figure 10.26: A difference map where the intensity values of the ground truth have been subtracted from the undersampled reconstruction intensities. Naturally, the SIRT algorithm had the lowest pixel-wise loss, displayed through PSNR. The GAN, on the other hand, was generally slightly below the ground truth in terms of intensity, which may be an effect of the renormalisation of the data.

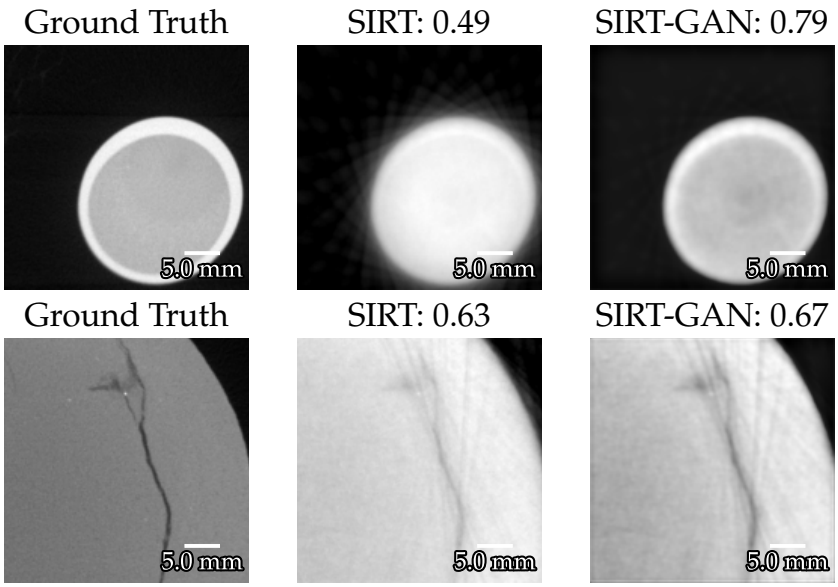


Figure 10.27: From left to right: the ground truth, an undersampled SIRT reconstruction, and the GAN-enhanced SIRT reconstruction, where both the training data and the input data were SIRT reconstructions. The upper cross sections are from the hourglass sample, reconstructed from 100 iterations using 26 projections. The lower cross sections are from the sandstone sample, reconstructed from 150 iterations using 34 projections. The SIRT-GAN provided contrast enhancement, and reduced blurring, but failed to remove the worst streaking artefacts.

10.5 ITERATIVE RECONSTRUCTIONS WITH GAN-ENHANCEMENT

It was noticed that, as the literature suggests, that the SIRT algorithm was more robust at handling undersampled data than the FDK algorithm. However, the SIRT algorithm was also more computationally expensive, requiring 150 iterations to achieve a result better than the FDK-GAN. At the same time, the GAN had difficulties separating the noise from the actual signal in heavily undersampled FDK-data. These circumstances led to the idea of training the GAN on undersampled and unconverged SIRT reconstructions, which proved successful. Figure 10.27 illustrates the result of applying the GAN to an undersampled SIRT reconstruction. It recovered the air-glass-sand interface for the hourglass cross section, and produced enhanced contrast between the sandstone and the cracks in the sandstone sample. These features are further exemplified in the line profiles included in Figures 10.28 and 10.29

However, in the case of the sandstone sample, streaking artefacts from the SIRT reconstruction remained after enhancement, as seen in the extension of the crack in Figure 10.27. Moreover, the structural similarity only increased slightly. Therefore, the spatial frequencies of the reconstructions were assessed by FSC, similarly to the analysis that resulted in Figure 10.16. In contrast to the mentioned analysis, the FSC was this time calculated for both samples, and compared SIRT reconstructions with SIRT-GAN reconstructions. Additionally, FDK reconstructions and FDK-GAN reconstructions were included for reference. The FSC curves of the reconstructions produced fundamentally by the SIRT algorithm are situated in the left subplot of Figure 10.30, while the FDK reconstructions are situated in the right subplot. From the figure, it was confirmed that the SIRT algorithm is more robust at handling undersampled data

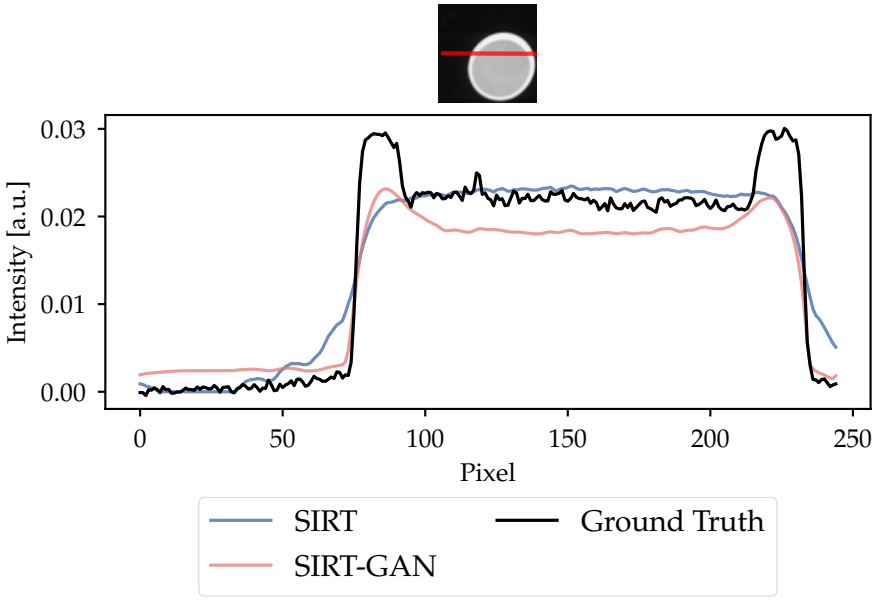


Figure 10.28: The line profile was retrieved from a static cross section of the hourglass sample. 100 iterations were performed using 26 projections. It is clear that the sharp edges of the hourglass were recovered by the SIRT-GAN. However, the intensity range remained lower than the ground truth, as the highest intensities were not recovered. The SIRT-GAN's intensities were slightly above the ground truth in the background, and slightly below the ground truth inside the hourglass.

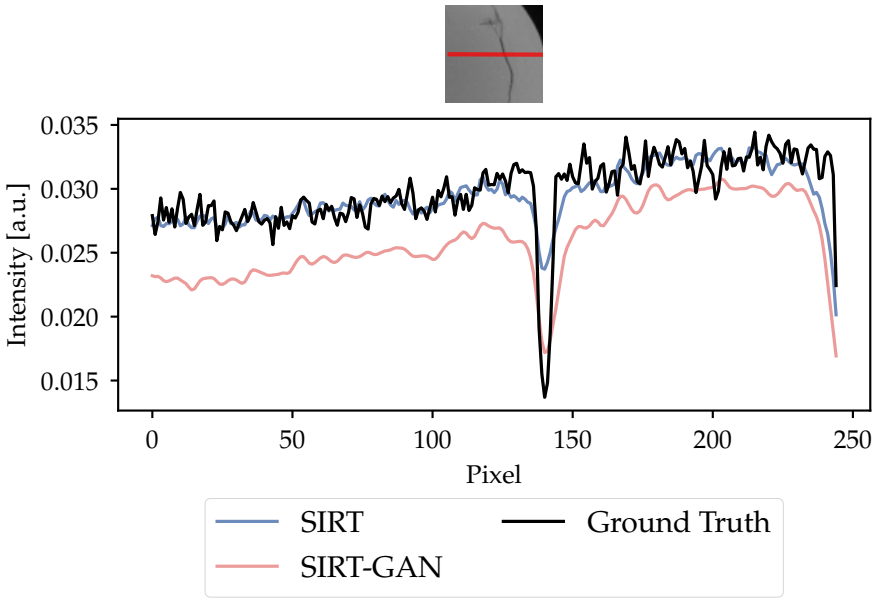


Figure 10.29: The line profile was retrieved a static cross section of the sandstone sample. 150 iterations were performed using 34 projections. The SIRT-GAN provided contrast enhancement of the crack. It also provided the main low-level features of the ground truth. However, the mean intensity was lower than the ground truth, which may be a result of the renormalisation. Because the line depicted retrieved intensities from both the centre and the edge of the sample, it is possible to see the effect of beam hardening.

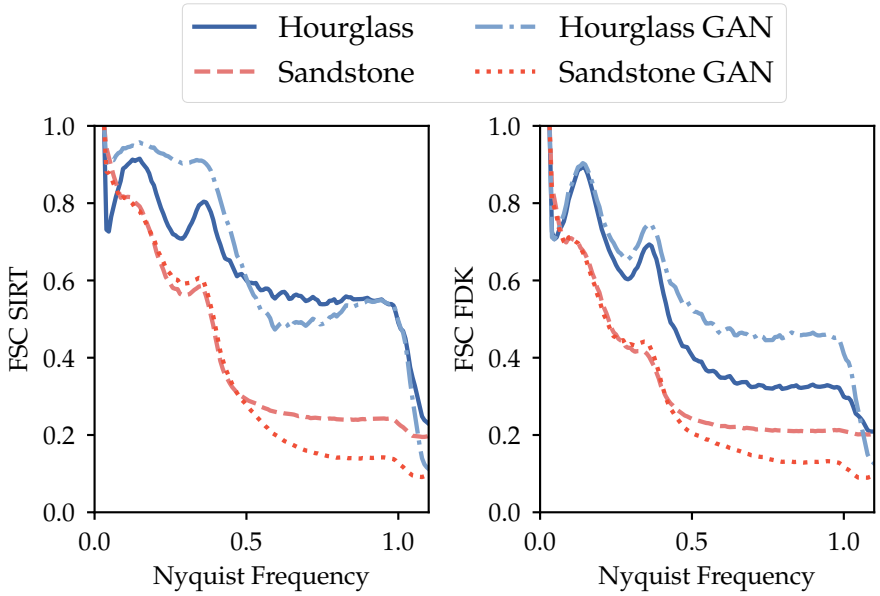


Figure 10.30: A comparison of the FSC of different samples and different methods is included in this figure. The number of projections was kept constant at 26 for the hourglass sample and 34 for the sandstone sample. This corresponded to two full revolutions, respectively. The FSC curves of the respective SIRT reconstructions are shown in the left subplot, with the corresponding FSC curves of FDK reconstructions situated on the right. GAN-enhancement resulted in a slight improvement at lower spatial frequencies, but a collapse at the Nyquist frequency. This collapse was greater with GAN-enhancement than without. For the sandstone sample, the GAN improvement was negligible. The SIRT-GAN reconstruction of the hourglass had a correlation above 0.9 up till a Nyquist frequency fraction of approximately 0.4.

than the FDK algorithm, as the correlation was generally higher in the left subplot. The same tendency as in Figure 10.16 was observed, namely that the GAN provided an enhancement for lower spatial frequencies, but collapsed for higher spatial frequencies. Another interesting tendency appeared for the hourglass sample beyond a spatial frequency of one half Nyquist frequency. Enhancement correlation was better than its respective input in the case of FDK, but worse in the case of the SIRT. The GAN-enhanced sandstone reconstructions, on the other hand, underperformed beyond one half Nyquist frequency in both cases, and only a positive vertical shift in the correlation differentiated the SIRT-GAN from the FDK-GAN.

11 TEMPORAL CT PERFORMANCE

11.1 TEMPORAL RESOLUTION

First and foremost, one key to satisfactory temporal resolution in the temporal CT experiment was the golden ratio V3 scanning procedure. The procedure was nicely situated between standard CT and original golden angle sampling in terms of scan time, as seen in Figure 8.8. With this minimal increase in scan time, the number of projections included in a given time stamp could easily be increased or reduced by integer multiples of the number of collected projections per revolution. This allowed for investigating the maximal temporal resolution, as performed in Section 10.1, or to reconstruct with the lowest number of projections that the GAN could handle, as performed in Section 10.2. Finally, despite being designated for undersampled CT, the derived scanning procedure was also applicable for CT following the Nyquist criterion, as each new revolution would provide unique projections that would contribute to new information in the reconstruction. One of the drawbacks with the procedure was, however, the discretised reconstruction flexibility, and that one had to carefully choose the number of projections per revolution. Ultimately, this became a compromise between the maximal number of projections per time and the best flexibility *a posteriori*.

The latter was chosen for the temporal CT experiment of the hourglass. To enable great flexibility in the data analysis, the number of projections per revolution was chosen to be 13, resulting in an appropriate temporal resolution of 49 s. It was also *a priori* assumed to be the absolute lower limit of projections that would provide working reconstructions. Anyway, because of the low number of projections per revolution, it was important that an odd number of projections were obtained per revolution. To elaborate, it was preferred that the projections were not directly opposing each other. Even though directly opposing projections are supposed to provide new information when performing cone-beam CT, it was observed that more unique information and better reconstructions were acquired when using an odd number of projections per revolution. As shown in Figure 10.2, for instance, the coverage of the sample was surprisingly good with only 13 projections. This result was part due to the binning of pixels in the projections, which was performed to increase the temporal resolution, and to enable the FDK-GAN to enhance the entire hourglass in one forward pass. As discussed in Section 9.3, the performed binning effectively halved the sampling requirements, thereby reducing the undersampling artefacts of the resulting reconstruction.

The maximal temporal resolution was then achieved, as seen in Figure 10.7, by splitting the revolution into two parts with 7 projections each, and reconstructing based on wedges with 180° coverage each. As expected, the reconstructions suffered from heavy undersampling artefacts, but a temporal resolution of approximately 26 s was achieved. Theoretically, the temporal resolution in the case of 13 projections per reconstruction could also be enhanced by utilising this missing wedge strategy. If one were to consider a CT scan with 25 projections per revolution, the scan would

then have a temporal resolution of 69 s, as shown in Figure 8.6. However, reconstructions with a temporal resolution of 36 s could be realised with the mentioned missing wedge strategy, which is an improvement of 48 %. Here, any introduction of missing wedge artefacts are assumed to be negligible relative to the extensive aliasing already present. With this assumption, the flexibility in the post-processing would not be affected, as it would simply be possible to add and subtract half revolutions to a given time stamp.

As demonstrated by the investigation of the sand-air interface in Section 10.1, the golden ratio V3 scanning procedure enabled reconstructions with increasingly high temporal resolution. The same strategy is applicable for studies of flow in porous media. For instance, an initial reconstruction with satisfactory spatial resolution could be performed to identify the regions of interest with dynamical artefacts. Then, a series of smaller time stamps with lower spatial resolution could be reconstructed to better resolve the flow dynamics in these regions of interest.

11.2 SPATIAL ENHANCEMENT OF HOURGLASS

The second essential key to a successful temporal CT scan was the FDK-GAN. Figures 10.4, 10.5, 10.6, 10.8, and 10.9 demonstrated how the FDK-GAN enhanced the dynamical process while at the same time removing pixel noise. The qualitative assessment presented was supported by the intensity distributions in Figure 10.6, since the FDK-GAN resulted in a greater dynamical range of gradient intensities than the FDK reconstruction, and a narrower Laplacian distribution without prevalent and predominant high-frequent noise.

However, the reconstructions in Section 10.1 observed little aliasing removal in comparison to the results presented in Section 10.2. Instead, the FDK-GAN had a tendency to enhance the undersampling artefacts. One probable reason for this was the discrepancy between the undersampling factor used in the binning experiment relative to the training undersampling factor. The effective undersampling factors were 32 and 59 for the binned hourglass before and after the wedge reconstructions, respectively. Evidently, the magnitude of the intensity of the undersampling artefacts scaled with the undersampling factor. Therefore, it would be difficult for the GAN to distinguish between the undersampling artefacts and the actual signal, especially since it was trained on a lower undersampling factor. This tendency was also observed in Figure 10.15, as the trained GAN was particularly effective at enhancing the hourglass sample reconstructed from 52 projections, which corresponds to the training undersampling factor of 8. This is further visualised in the frequency domain in Figure 10.16. As a result of the GAN-enhancement, all hourglass reconstructions ranging from 52 to 104 projections were enhanced to a similar level. With further elevated undersampling factors, on the other hand, the GAN provided enhancement, but the FSC curves of these reconstructions showed a generally lower correlation with the ground truth. This tendency also seems to support the fact that the GAN in these cases failed to distinguish between undersampling artefacts and true signals.

It is also important to realise that the relation between the chosen temporal resolution and the corresponding spatial resolution is highly sample-dependent. The hourglass sample mainly consists of air, sand, and glass. The fact that these materials had nice contrast to one another, and were distributed in separate regions, is one prob-

able explanation to why it was manageable to reconstruct and assess the sand-air interface without more undersampling artefacts. Moreover, the GAN was probably particularly adapted to handle the simple and homogeneous hourglass sample. As shown in Figure 5.1, half of the data, namely the simulated data, were very similar homogeneous objects with bright contrast. More homogeneous regions result in less undersampling artefacts, as sharp edges correspond to high frequencies in Fourier space, which cannot be accurately described with a too low sampling frequency [13].

In order to reconstruct with fewer projections, the SIRT algorithm was utilised. The line profiles in Figures 10.19 and 10.28 were, however, clear indications that the iterative reconstruction resulted in loss of signal, because of smoothed out transitions together with the loss of contrast. These features are common artefacts associated with iterative reconstruction techniques, and were therefore no surprise [29]. However, these artefacts can be reduced by increasing the number of iterations, but this would also increase the computational cost.

Instead, a GAN trained at elevated undersampling factors using the FDK algorithm could be a better alternative for this sample. However, this would include an enhanced difficulty for the generator part of the GAN, since the undersampling artefacts would be more severe, while the discriminator would have an easier task. Subsequently, the outcome of the training could more easily be mode collapse, as the generator would fail to produce realistic reconstructions. This would therefore require careful reconsideration of the model and training procedure, which was not prioritised. With that being said, if the magnitude of the undersampling artefacts are equal to the magnitude of the actual signal, the GAN would not be able to distinguish between the two. Essentially, the trained GAN is just a sequence of optimised filters convolved with the input volume. These filters are optimised to perform feature extraction [59], and in the case of the U-net generator, also feature enhancement and concatenation. An extremely simplified example of the learned filters of the generator could be the following. If the GAN has learned to feature extract and enhance the signals with the highest intensity, because the undersampling artefacts intensities were only a fraction of the signal intensities, noise in an evaluated reconstruction with near-signal magnitude would end up also being enhanced. Likewise, true signals with low intensities would be suppressed, as the GAN would have learned that these signals are noise. These two drawbacks, though simply described, are considered the ultimate limitations of convolutional neural networks, and is why attention and transformers, mentioned in Section 3.4, are considered the next step within this topic [52].

Therefore, the amount of undisturbed information provided to the GAN is essential for its performance, as suggested by the SIRT-GAN versus the FDK-GAN for the sandstone sample, illustrated by the FSC curves in Figure 10.30. Because the input contains more information, the enhancement becomes more accurate. However, one would like to avoid the use of iterative reconstruction methods in the first place, as they are computationally expensive. The alternative to better input reconstructions could be to provide additional information, such as the sinogram of the CT scan. This was performed by [48], and enabled enhancement with more information present. A more elaborated discussion on the possibilities of such an approach is, however, reserved for the last section of this chapter.

11.3 SPATIAL ENHANCEMENT OF SANDSTONE

As presented in Section 10.4, the sandstone sample was more challenging than the hourglass sample. From the line profiles in Figures 10.23, 10.24, and 10.25, it is evident that the FDK-GAN reduced the influence of outliers from the original FDK reconstruction, and that the main signal, the crack, was contrasted from the background. One should also notice that the FDK-GAN reconstruction for the most part seemed to correlate with the ground truth, but due to an input signal rich in outliers, the FDK-GAN reconstruction tended to overestimate the variation of the signal.

In contrast, the background variance presented by the SIRT algorithm resembled the ground truth more closely, while the cracks were heavily underestimated. With the GAN trained using the dataset of undersampled SIRT reconstructions in Figure 5.5, most of the main background features were captured. The SIRT-GAN also enhanced the cracks, but not as much as the FDK-GAN. The general impression of these results is either that the training dataset was too diverse, resulting in too careful enhancement.

Alternatively, the updates of weights were too noisy when performing online learning, as explained in Section 2.2, which would be an indication to increase the batch size. Tests were performed for these hypotheses, with findings presented in Appendix C. To shortly summarise the results in Appendix C, model weights with improved performance for the sandstone sample were found in both cases. However, these models underperformed relative to the original model for the hourglass sample. Moreover, undersampling artefacts were still present in the reconstructions, but these became less prominent for 85 projections. As already mentioned in Section 11.1, this may be because of the training undersampling factor in relation to the testing undersampling factor. It should be noted that the region of interest for the sandstone sample includes the edge of the sample, which means that a region with more undersampling than the hourglass sample has been depicted. Nevertheless, the tests demonstrated that enhancement of the sandstone sample was possible, even with the original dataset.

Considering the performance variance of the derived FDK-GANs, there is a possibility that too few training iterations were performed, resulting in no performed final fine-tuning of the generator weights. This will, however, be further investigated in the following chapter.

11.4 COMPARISON TO THE LITERATURE

A comparison between the developed three-dimensional GAN and similar architectures in the literature is of great relevance in order to assess the spatial enhancement ability of the model. It would be natural to compare with the TomoGAN architecture [33], as much inspiration was retrieved from this work. A comparison between the two is especially relevant in order to uncover any weaknesses or strengths in the transition to a 3D convolutional network.

The most relevant results to compare would be the enhancement of the hourglass presented in this work with the enhancement of the simulated data presented in [33]. In the paper, one randomly created simulated dataset of circles was used for training, while the test dataset was randomly created from a different seed. As mentioned, the hourglass reconstruction was enhanced by a model trained using a mixture of

experimental and simulated data. In this regard, it can be argued that the GAN in this work obtained a more general enhancement ability than the TomoGAN. Moreover, by comparing the listed slice SSIM values, it is clear that the performance of the two models were fairly similar. Both models produced reconstructions with SSIM values of approximately 0.9, which dropped to approximately 0.8 for undersampling factors of 25 to 32.

It is also interesting how the TomoGAN elimated the undersampling artefacts at elevated undersampling factors. No undersampling artefacts were observed at an undersampling factor of 25, corresponding to 64 projections, for the 1024×1024 input slices. At this level of undersampling, the FDK-GAN considered undersampling artefacts as signal, and enhanced them. Furthermore, by comparing the TomoGAN line profiles to the FDK-GAN line profile of the hourglass, it is clear that the TomoGAN trained on simulated data learned to only enhance the most prominent signals, and consider the rest as noise. The FDK-GAN was much more careful in this regard. Consequently, there is reason to believe that the TomoGAN would struggle to enhance data where a third signal intensity were present. Hence, the TomoGAN trained on simulation data seems to be somewhat over-adapted.

Nonetheless, the so-called CTTR model developed in [49] was trained on more complex medical data, and was able to recover high-quality reconstructions. It should be noted that the CTTR architecture differed from the mentioned GANs by being an ordinary convolutional neural network without a discriminator. Instead, the sinogram was added as part of the network input, making it a transformer with the ability to exploit *attention* [52]. Attention allows the network to assign significance to each input feature. With this strategy, a global SSIM of 0.86 was achieved for an undersampling factor of 25, corresponding to 30 projections for 512 wide slices. This transformer architecture should be considered in future work. To elaborate, there is definitely a possibility to replace the generator of the GAN with the convolutional transformer network presented in [49].

The transformation to 3D with this implementation, however, would not be trivial, as the entire reconstruction slice and corresponding sinogram would have to be provided to the transformer. Due to memory limitations, this would probably not be possible without binning and a limited height of the volume. For the purpose of studying dynamical processes in 3D, it is imperative for the continuity in the reconstructed volume to utilise three-dimensional convolutional neural networks instead of two-dimensional [56]. The flow in dynamical experiments tends to go from top to bottom, deeming it necessary to not just convolve in the XY-plane, but also along the Z-axis. For this reason, it is important to notice the fact that the enhanced reconstructions presented in Figure 10.21, for instance, exhibited equally good quality along the vertical axis as in the horizontal plane. Here, the power of three-dimensional convolutional neural networks was demonstrated, as discontinuities in the Z-direction can be a typical weakness of two-dimensional convolutional neural networks, while three-dimensional neural networks ensure continuity in the third dimension [56].

Consequently, typical reconstructions with a width of 1024 pixels would have to be binned to at least a width of 512 pixels. Then, to compensate for the feature increase from having 512×512 slices, the length of the third dimension would theoretically have to be reduced to 8. However, with memory optimisation and discarding of the perceptual loss term, which up to this point has not been used in this work, the third dimension could be prolongated.

Finally, it should be noted that the desired generalisation emphasised in this work exceeds what was presented in [49], as the CTTR model was solemnly trained on lung data. Moreover, the splitting of training and testing data was performed across patients, which means that very similar slices were present in both the training and the testing dataset, as lung reconstructions should be fairly similar across different patients. This could also be a challenge in the theoretical 3D implementation of the proposed CTTR-GAN, as the training dataset size would drastically be reduced. The idea of providing unaltered sinogram information to the enhancement network is nevertheless interesting, because of the shown dependency on input data quality. This natural next step will be considered in future work.

12 3DCONVGAN

For the sake of emphasising the three-dimensional feature extraction of the developed enhancement model, the model is named *3DConvGAN*.

12.1 INITIAL TRAINING ON SIMULATED DATA

As an initial test of the derived enhancement model, the model was trained on simulated data reconstructed from 100 projections, effectively having an undersampling factor of 4. No perceptual loss was included in the loss function, and the model was trained for 500 epochs. This allowed for fast training, and each elapse of the training took approximately 0.7 s. Considering evaluation of 10 samples in each epoch, the training took approximately 1 h. The MSE loss was weighted so that it was approximately equal to the adversarial loss initially. However, this loss term was relatively quickly reduced, making the adversarial loss the dominant loss term. The discriminator loss stagnated at approximately 1.39, caused by providing a score of 0, effectively guessing on both real and fake samples.

Moreover, one instance of augmented ground truth, undersampled reconstruction, and enhanced validation reconstruction are included in Figure 12.1. As shown in the figure, the model was able to enhance the contrast of the undersampled reconstruction. Consequently, SSIM was increased from 0.385 to 0.859, while the pixel-wise loss was slightly increased from a PSNR of 26.0 to 26.6. In terms of MSSIM with a kernel size of 11, the score increased from 0.14 to 0.23.

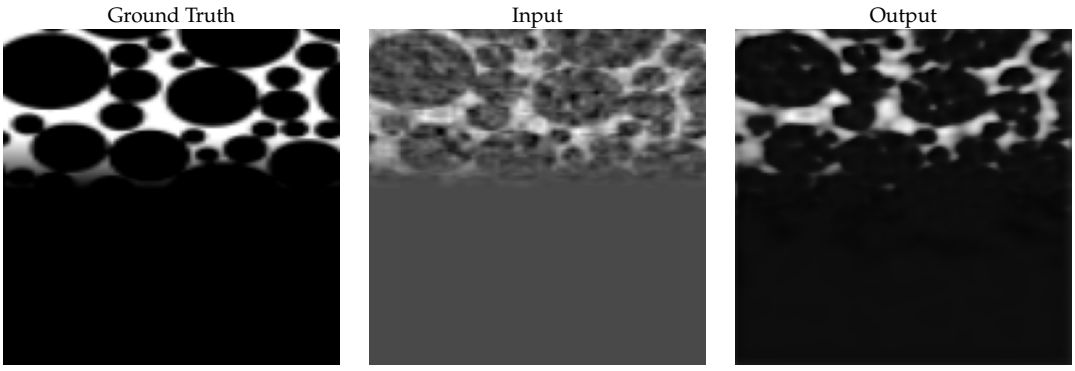


Figure 12.1: The effect of the model after initial training for 500 epochs on the simulated dataset. Enhancement of the augmented validation sample in the simulated dataset resulted in a great increase in global SSIM, from 0.385 to 0.859, while the pixel-wise loss and the MSSIM increased slightly from 26.0 to 26.6 and 0.14 to 0.23, respectively.

For the purpose of further optimisation, the effect of the model was manually evaluated by comparing a central 256 cubed region of a fully sampled reconstruction with the corresponding undersampled reconstruction before and after GAN-enhancement. A comparison of the ground truth, undersampled reconstruction, and enhanced re-

construction for 26 and 78 projections, respectively, is shown in Figure 12.2. Like the validation enhancement in Figure 12.1, the model managed to remove undersampling artefacts from the least undersampled reconstruction, while much remained for the reconstruction containing only 26 projections. In both cases, the GAN increased the contrast between the interfaces, and reduced the high-frequency noise. This intermediate result indicated that representative data had been produced, and that the architecture was able to learn its designated task. However, as seen in the enhancement at 78 projections, the weak strip of sand in the neck of the hourglass was interpreted as noise, and was therefore removed. The model therefore had to be trained further to avoid this behaviour.

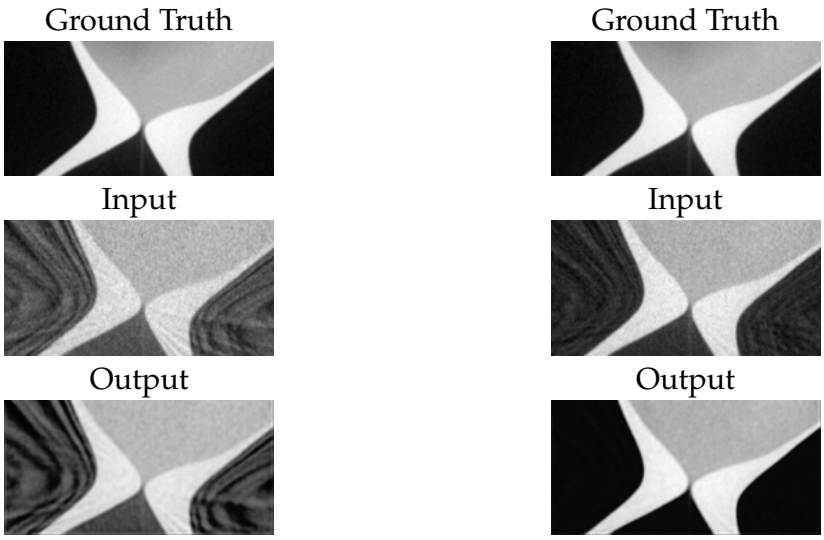


Figure 12.2: The effect of the model after initial training for 500 epochs on the simulated dataset. The left column presents the centremost neck region of the hourglass reconstructed from 26 projections, while the right column is retrieved from the same region, but reconstructed from 78 projections. These number of projections correspond to an undersampling factor of approximately 15 and 5, respectively. With respect to the listed ground truth, the relative undersample factors are approximately 27.5 and 9, respectively.

12.2 TRAINING OF OPTIMISED MODEL

As explained in Section 6.4, transfer learning was used to train optimised models. The model mentioned in Section 12.1 was used as initialisation for further training using both the FDK 5.2 and SIRT 5.5 datasets. Two instances of transfer learning are shown in Figure 12.3, where the loss curve of the initial training is included in a). The resulting model was then used as initialisation for training on the FDK dataset 5.2, for which the generator loss is included in b), and the discriminator loss is included in c). Likewise, the generator loss and discriminator loss were included for training on the SIRT dataset 5.5 in d) and e), respectively. A general increase in generator loss was observed for both datasets, as well as a general decrease in discriminator loss. This was an indication that the discriminator catched up with the pre-trained generator. To further investigate the dynamics of the discriminator, the discriminator evaluation scores of real and fake reconstructions were plotted for both datasets in

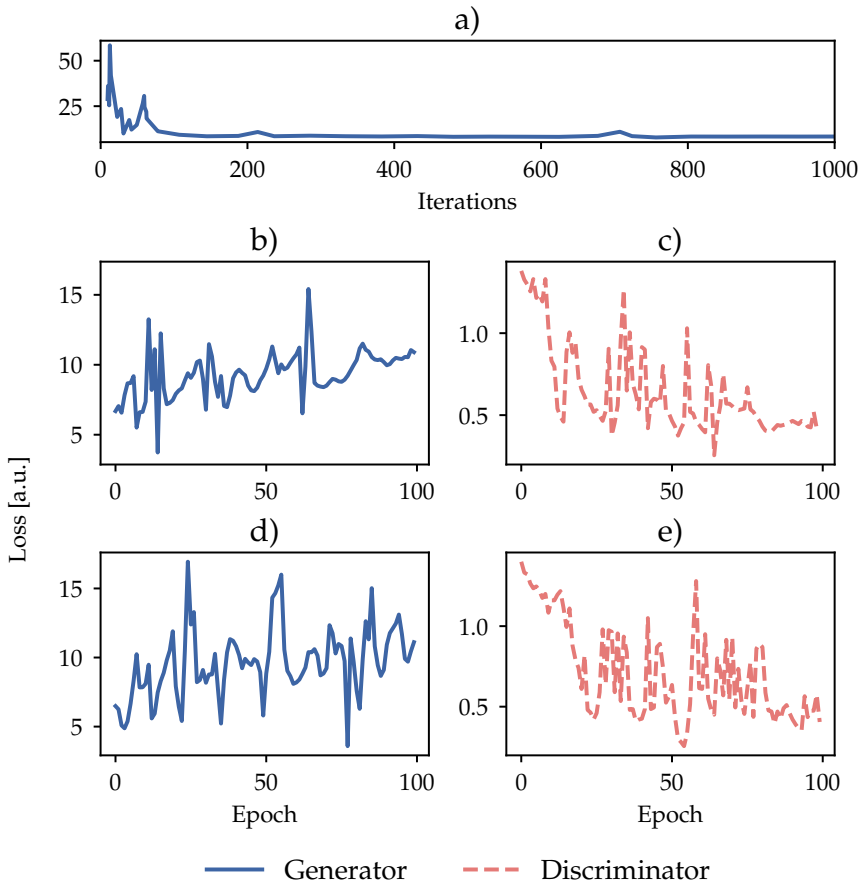


Figure 12.3: a): The validation loss curve of the initial training of the model on simulated data.
b): The development of the generator loss when the trained model was used as initialisation for training of the GAN on undersampled FDK reconstructions.
c): The development of the discriminator loss during the same training as in b).
d): The development of the generator loss when the trained model was used as initialisation for training of the GAN on undersampled SIRT reconstructions.
e): The development of the discriminator loss during the same training as in d).

Figure 12.4. Included in the figure are the average scores during the first and last epoch of training, respectively. The trend of the discriminator score was similar for both datasets, where the score of the real reconstructions increased, while the score of the fake reconstructions slightly decreased. A score close to zero is an indication of a coin flip, while a high positive score is an indication of confidence in the realness of the input.

12.3 EXPERIENCES FROM INITIAL TRAINING

The results of the initial training of the model were promising. Without much hyperparameter tuning, the model was able to enhance the validation data, as seen in Figure 12.1. However, some weaknesses were also discovered. The most significant weakness was the lack of generalisation of the model. This became apparent when the model was tested on the hourglass and the sandstone samples. Even though

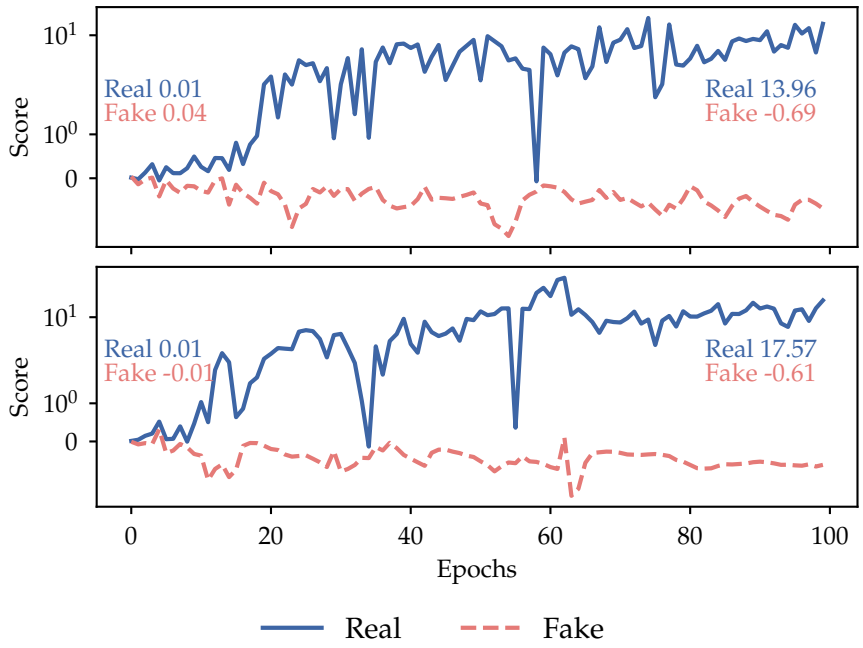


Figure 12.4: From top to bottom, the discriminator score during training of the GAN on undersampled FDK reconstructions and undersampled SIRT reconstructions, respectively. For each loop of updates, the last iteration of the generator and discriminator updates were stored, respectively. The average of the collected values within each epoch were then plotted.

the enhancement of the hourglass sample was satisfactory, the enhancement of the sandstone sample was worse than the original undersampled reconstruction. When comparing the training data to the test samples, one probable cause for this was the lack of intensity and contrast variation in the training data. Because the training data was mostly black or white, it could better enhance the abrupt edges of the hourglass than the dark gray cracks of the sandstone. The significant difference between global SSIM and MSSIM was also noted during validation. Globally, it was observed that the model increased the contrast between the background and white features of the validation data, and that the correlation coefficient was high, because the lower half of the image was mostly dark, while the local correlation in the upper part of the image was lower due to imperfectly reconstructed circles, for instance. Therefore, the resulting global SSIM score was high, while the MSSIM score was much lower. The distortions in the undersampled images would be more significant in the view of a 11×11 gaussian kernel than in a global evaluation.

With that being said, the low MSSIM score could also be a result of interpolation artefacts from the rotation performed in the data augmentation. A visual inspection confirmed that interpolation artefacts were present during training, causing blurring and imperfect edges. However, with regard to the limited size of the dataset, the rotation augmentation was still necessary to increase the effective dataset size.

Moreover, from assessing the training process in the log data, illustrated in Figure 12.3 a), it was supposed that the discriminator was somewhat weak, as it often guessed randomly. Therefore, the number of output channels in the final convolutional layer of the discrmiminator was increased from 4, which was the case in the

original TomoGAN [33], to 16. A compression from 256 to 4 channels seemed to be an aggressive feature selection with the potential of limiting the training capability of the GAN, as explained in Section 2.6. Another reason for increasing the number of output channels was the fact that training was performed using twice the number of input features relative to the original TomoGAN. Adjusting the importance of the adversarial loss term relative to the MSE loss term was also necessary to prevent the generator from blurring the output. The model trained initially was actually trained with a too high MSE loss term weight, essentially making it an ordinary neural network. With this setting, there was a high probability that the generator would end up completely blurring the output after some epochs, an indication of mode collapse, described in Section 2.6. A stronger discriminator together with a lower MSE loss term weight was therefore expected to improve the training process.

Additionally, it was believed that a scheduling of the learning rate could be beneficial, hoping for a more fine-tuned training of the model in the last parts of the training. A cosine annealing schedule was therefore implemented, where the learning rate would essentially drop like a cosine function from a top learning rate at epoch 0 to the lowest designated learning rate at the final epoch [40].

More importantly, however, was the relation between generator updates and discriminator updates. Based on the fact that the discriminator was essentially incapable of distinguishing between real and fake data after only 200 iterations in Figure 12.3 a), a more dynamical training algorithm was deemed necessary. Instead of trying different numbers of generator updates per discriminator update, a condition for which the generator updates would continue, with a corresponding condition applying to the discriminator, was implemented. To prevent infinite loops, this would reset after iterating through the entire dataset. Moreover, two hyperparameters were introduced, indicating the maximal number of generator and discriminator updates in a row, respectively. To elaborate, the generator should perform updates as long as the adversarial loss was above 0.69, which is the mark of random guessing in BCE loss (2.5). Since the discriminator loss consisted of two BCE terms, its threshold was set to 1.38. Based on the desired loss curve oscillations in Figure 12.3 b) and c), the implementation was deemed beneficial for the training process.

Finally, training was resumed on data that was more heavily undersampled, to further adapt the model to even more undersampling artefacts. Because the TomoGAN paper [33] demonstrated the abilities of perceptual loss, a perceptual loss term limited to the first few layers of the ResNet50 neural network was added to the loss function [25]. The perceptual feature extractor was limited to the first few layers of the ResNet50 neural network due to memory limitations. Unfortunately, this loss term inhibited the training process, and was disregarded. A discussion of this is included in Section 12.6.

12.4 EVALUATION OF OPTIMISED WEIGHTS

There is reason to believe that optimal and achievable generalisation across numerous samples with unique characteristics has yet to be found. The FDK-GAN model used in Chapter 10 showed a bias towards being optimised for the hourglass sample, while another FDK-GAN model trained using batchsize 4 and input dimensions of 64^3 voxels, included in Appendix C, showed a bias towards the sandstone sam-

ple. The aspect of broad generalisation is therefore still a challenge, which is not frequently mentioned in the literature, as the supervised machine learning models for CT imaging found in the literature are usually trained on a dataset of very similar samples. For instance, the model might be two-dimensional, and may therefore be trained using slices of only one reconstruction of the same type of sample [33]. In other cases the task of the model is to segment a special organ or tissue, and the model is therefore trained on multiple similar samples of the same organ or tissue [49].

12.5 HYPERPARAMETERS

The fine-tuning of hyperparameters is a vital part of training a machine learning model, and sufficient time should be allocated to this task [41]. However, due to the many hyperparameters of the model, and the limited time of the project, the hyperparameters have not been thoroughly optimised, as this would require a significant amount of time. Not only would it be time-consuming to perform the hyperparameter search, it could be a manual and tedious challenge to evaluate what settings produced the best results, because of the amount of output data together with the need for human visual perception to evaluate the enhancement. Strategies of hyperparameter optimisation include grid search and stochastic search [11]. The first is to create many combinations of hyperparameters, and train the model on each combination to find a suitable combination. The latter is similar to grid search, but the combinations are randomly selected.

Instead, most of the hyperparameters, such as adjustments to the model architecture, choice of optimiser, learning rate scheduler, cropping variance, reasonable data augmentation, choice of activation function, and choice of loss functions have been chosen based on experience, intuition, and what is found to be effective in the literature [6]. The hyperparameters that have been attempted optimised were the loss function weights and the strength relation between the generator and the discriminator. As mentioned in Section 12.3, undesirable effects occurred when the loss function weights were not properly tuned. For instance, a too high weight on the MSE loss resulted in blurring of the enhanced reconstruction, because outputting the average pixel value is one way to minimise MSE loss. Conversely, not including the MSE loss resulted in output not resembling the input reconstruction at all.

Nevertheless, an optimal combination of hyperparameters have probably not been found, and the model or the training process could likely be improved by a different set of hyperparameters. In such a hyperparameter search, the most important parameters to tune would be the loss function weights, which is the major disadvantage of combining multiple loss functions. For instance, a stable training process with a perceptual loss was never found. However, as demonstrated in the literature, the perceptual loss function contributed to the performance of the model [33]. With the 3DConvGAN, however, it consistently resulted in mode collapse by blurring the output, which will be further discussed in the next section.

12.6 ARCHITECTURE

Even though the utilised architecture was heavily inspired by the architecture of the TomoGAN [33], the many adaptations to the model have to be addressed. Some were a matter of preference and intuition, while others were necessary to successfully make the transition from 2D to 3D.

The first difference between the two models was the choice of activation function. Here, the ordinary ReLU activation function was replaced by the parametric ReLU variant. The latter has the ability to adjust the slope of the activation function for input values $x < 0$ at the cost of one additional learnable parameter. The variant was proposed by [24], and was found to provide a slim performance increase in the original paper. In the case of 3DConvGAN, it was chosen to ensure that the dying ReLU problem would not occur. For neurons with negative output, the ReLU activation function will output zero, and the neuron will therefore no longer contribute to the network [19]. A slim increase in computational cost was considered a reasonable trade-off for insurance against this issue.

Another adaptation of minor importance was an adaptive average pooling layer in the connection between the convolutional layer and the fully connected layer of the discriminator. With this implementation, the GAN should in theory be able to perform training on input data of any size divisible by 8. However, the effects of this layer were not thoroughly evaluated in the cases of heavily increased or decreased input sizes. This is nevertheless of minor importance, as the number of input features to the fully connected layer were tuned to be reasonable for the training input size of $128 \times 128 \times 128$.

Secondly, an increase in number of output channels in the final convolutional layer of the discriminator was made based on interpretation of the initial training loss curve in Figure 12.3 a). This increase accounted for both the increased number of input features of 3DConvGAN relative to TomoGAN, together with a strengthening of the discriminator, as described in Section 12.3. It was, however, never determined whether this change contributed to an enhanced performance of the model.

On the other hand, it was very difficult to train the GAN from scratch after having implemented this change, as the discriminator would outperform the generator. Moreover, the discriminator loss curve in Figure 12.3 c) and e) would suggest that the discriminator was quite strong, as the loss quickly decreased despite the pre-training head start of the generator.

However, this tendency is not necessarily an issue. As shown in Figure 12.4, the decrease in discriminator loss was predominantly caused by an increase in the score for real samples. A hypothetical interpretation of the discriminator based on this observation, would be that it was confident in its ability to recognise real samples, possibly because of the complete absence of undersampling artefacts.

Nevertheless, when presented with fake samples, it was not overly confident, which could be because of the generator enhancement. Here, a discriminator score on unenhanced samples would be valuable for comparison. Anyway, this relation between the generator and the discriminator can be interpreted as a sign of a progressing adversarial training process, where there is still room for generator improvement [20]. It is also important to note that only the loss calculated before switching between generator and discriminator was calculated, which would typically imply that only the optimal performance of the respective networks has been presented. Therefore,

the mean and variance of the loss curves would generally be lower than if all loss calculations were saved. Anyway, observations of printing to screen during training supported generally the tendencies presented here, and there seems to be enough balance between the generator and the discriminator to ensure a stable training process.

Thirdly and most importantly, the limitations associated with the transition from 2D to 3D have to be addressed. As a consequence of this transition, the memory consumption, the number of learnable parameters, and the computational time all increased significantly. Since it was decided to train the model on a single GPU, the memory consumption was a major concern, as 24 GB of memory was the absolute maximum available GPU memory.

The first constraints from the limited GPU memory was that online learning with a batch size of 1 was the only viable option. As mentioned in Section 2.2, this enhances training speed, which is beneficial given the computational time of 3DConvGAN, at the cost of noisy updates and a vulnerability to outliers. It should be noted that a batch size of 4 was tested, and is presented in Appendix C. To enable this, the input dimensions were halved along each axis. With this setting, the generator found weights that resulted in a satisfactory enhancement for 85 projections of the sandstone sample.

Another issue resulting from the memory constraint was the perceptual loss term. Perceptual loss requires a deep pre-trained neural network with many million of parameters, which involves a significantly extended computational graph, and therefore a large memory consumption. Moreover, considering that the input size of each size had to be increased to $224 \times 224 \times 3$, it was soon realised that a full worthy perceptual loss term was not feasible. Instead, only the first few layers of the pre-trained ResNet50 network was utilised [25]. Consequently, it was difficult to evaluate how accurate the loss term was, and how it contributed to learning. Anyway, it seemed to correlate with the MSE loss term. The result of this implemented perceptual loss term was blurring during training if the weighting was not close to zero. This undesired effect could imply that the weighting of loss terms involving independent pixel-wise loss, as mentioned in Section 12.3, was too high. Therefore, this loss term was discarded entirely.

A proposed solution to the problem is firstly to choose the smallest pre-trained network possible, but include all feature extraction layers to ensure the quality of the extracted features. Next, a batch of 8 to 16 slices should be evaluated in the perceptual loss. In this way, at least some valuable perceptual information could be fed to the generator, without the immense memory consumption.

13 CONCLUSION

13.1 CONCLUSION

In conclusion, a laboratory 4D-CT technique consisting of golden ratio sampling and a generative adversarial network reconstruction enhancer was developed. This technique allowed for satisfactory reconstructions of the hourglass sample using only 52 projections, corresponding to an undersampling factor of 8, the same as the training undersampling factor of the GAN. The experimental temporal resolution of the conducted experiment with this number of projections was 196 s, but by applying the discussed propositions for optimisation, the temporal resolution could be improved to approximately 120 s for the same number of projections by reconstructing twice per revolution, so that a sampling frequency of 103 projections per revolution could be chosen.

If flexibility in the post-processing were prioritised, one would instead choose a lower number of projections per revolution at the cost of a longer scan time. With such a choice, a proposed data analysis strategy where an initial reconstruction with good spatial resolution were to be performed, following by a series of increasingly undersampled reconstructions to resolve the fastest dynamics, would be possible.

When accepting severe undersampling, a theoretically achievable temporal resolution of 26 s was found to be possible by obtaining reconstructions using 7 projections covering 180°. With this degree of undersampling, specifically fewer than 51 projections when reconstructing the sandstone sample, it was shown that an increase in exposure time would greatly increase the exposure ratio, which seemed to be somewhat linked to the quality metrics used to assess reconstructions; namely SSIM, PSNR, and MSSIM.

The developed GAN was important for enhancing contrasts, reducing noise, and if the undersampling factor did not exceed 8, removing streaking artefacts. More importantly, its three-dimensional symmetry preserved continuity in all three dimensions, which is considered a necessary requirement when studying flow along the Z-axis, which is typically the case for such 4D-CT experiments.

To put things into perspective, instead of studying dynamical processes using only 2D projection data with temporal resolution in the order of seconds, or only studying slow processes in volumes due to Nyquist sampling requirements, the developed technique has provided the flexibility to study dynamical processes in volumes with an adjustable temporal resolution in the order of minutes. Consequently, both slow and fast processes can be studied from the same scan, and the location and direction of the process can be determined in all three dimensions.

13.2 FUTURE WORK

The GAN-enhancement is still in need of further optimisation to uncover how to best find network weights for the removal of undersampling artefacts, particularly streaking artefacts. Suggestions for future work has been discussed in the previous chapters, but the main aspects shortly summarised are the following.

Firstly, training with elevated undersampling factors is required to improve the GAN performance for heavy undersampling. For this purpose, another dataset has to be acquired where the reconstruction quality is ensured to be uniformly poor. Adaptations to the training process and GAN architecture should be performed to account for the shifted relation between the generator and discriminator, as an increased undersampling factor will result in a more challenging task for the generator.

Secondly, an increased dataset size will increase the probability that the dataset probability distribution will represent the probability distribution of all samples encountered during application of the GAN. With a larger dataset, possibly together with minibatch training if applicable, there is also an increased probability that a more universal compromise of enhancement is reached.

Thirdly, a set of hyperparameters that maximises the capacity of the generator and discriminator, respectively, and enables fine-tuning of weights at the end of the training procedure should be found through stochastic search or grid search.

Finally, given the restriced GPU memory at hand, the GAN architecture should be optimised to prioritise the most important components for performance. For instance, a choice should be made in terms of perceptual loss versus batch size. It is also worth deciding whether the enhancement procedure is in need of additional input, such as the sinogram data of the scan, because it has been implied that peak enhancement performance is related to the information provided by the input data. It was discussed whether investigations concerning the advantages of *attention* would be fruitful in the development of a more robust enhancer [52]. One proposition mentioned was to replace the U-net architecture with a transformer architecture that could exploit both the undersampled reconstruction and the measured sinogram.

Despite the need for further optimisation, a natural next step would be to apply the developed technique to an advanced 4D-CT experiment. In this way, an extended understanding of the technique's capabilities and limitations could be obtained.

Furthermore, with an achieved optimised architecture and training procedure, only small adaptations should be required to solve entirely new tasks. These tasks could all provide valueable insights, and should be investigated. For instance, the most relatable task would be reconstruction segmentation, where the GAN would be trained to pixel-wise segment the reconstruction into the different phases present [63]. With such an analysis tool, the dynamics could more easily be quantified. The disadvantage of this approach is the immense amount of manually labelled data required for training.

Finally, it was experienced that improvements in terms of quantitative image quality metrics that resemble human visual perception would be beneficial. Even though it would be a side quest in the development of optimal 4D-CT, it would not only be interesting, but in fact important, to investigate how to best quantify image quality in 3D and 4D. Currently, weaknesses in all metrics used have been exposed. For instance, PSNR is not sensitive to blurring, a distortion that humans tend to notice [27]. Additionally, there is the aspect of how to best quantify the quality of a 4D re-

construction, which is a task that has not yet been addressed. Advances in the field of deep learning could be applicable, for instance transfer learning of a pretrained network to derive a loss term more resembling human perception, a perceptual loss term [33]. Experiments with different loss terms at the end of such a feature extraction, for instance MSE, SSIM, or a correlation metric, could further improve this deep image quality assessment. To include the time domain into this metric, a pre-trained sequence based neural network, such as a recurrent neural network or a transformer, could possibly be used to capture this feature, given the current complexity of networks such as ChatGPT [39].

BIBLIOGRAPHY

[1] Serhat Akin and A. Kovscek. 'Computed tomography in petroleum engineering research.' In: *Geological Society, London, Special Publications* 215 (Jan. 2003), pp. 23–38. DOI: 10.1144/GSL.SP.2003.215.01.03.

[2] Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. 'Automatic differentiation in machine learning: a survey.' In: *Journal of Machine Learning Research* 18 (2018), pp. 1–43.

[3] Ander Biguri, Manjit Dosanjh, Steven Hancock, and Manuchehr Soleimani. 'TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction.' In: *Biomedical Physics & Engineering Express* 2.5 (2016), p. 055010.

[4] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*. Vol. 4. 4. Springer, 2006.

[5] Mikhail Arnol'dovich Blokhin. *The Physics of X-rays*. Vol. 4502. United States Atomic Energy Commission, Office of Technical Information, 1961, p. 12.

[6] Léon Bottou. 'Stochastic gradient descent tricks.' In: *Neural Networks: Tricks of the Trade: Second Edition* (2012), pp. 421–436.

[7] Thorsten M Buzug. *Computed tomography: from photon statistics to modern cone-beam CT*. Soc Nuclear Med, 2009, pp. 15–46.

[8] Thorsten M Buzug. *Computed tomography*. Springer, 2011.

[9] Francesco De Carlo et al. 'Tomobank: a tomographic data repository for computational x-ray science.' In: *Measurement Science and Technology* 29.3 (2018), p. 034004. URL: http://stacks.iop.org/0957-0233/29/i=3/a=034004.

[10] Guang-Hong Chen, Jie Tang, and Shuai Leng. 'Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets.' In: *Medical physics* 35.2 (2008), pp. 660–663.

[11] Marc Claesen and Bart De Moor. 'Hyperparameter search in machine learning.' In: *arXiv preprint arXiv:1502.02127* (2015).

[12] Richard A Dunlap. *The golden ratio and Fibonacci numbers*. World Scientific, 1997.

[13] Nicholas Dwork. *The Nyquisy Theorem and Sinc Interpolation*. https://nicholasdwork.com/tutorials/nyquistInterp.pdf. 2016.

[14] I Education. *What is Machine Learning*. 2021. URL: https://www.ibm.com/cloud/learn/machine-learning.

[15] Lee A Feldkamp, Lloyd C Davis, and James W Kress. 'Practical cone-beam algorithm.' In: *Josa a* 1.6 (1984), pp. 612–619.

[16] Li Feng. 'Golden-angle radial MRI: basics, advances, and applications.' In: *Journal of Magnetic Resonance Imaging* 56.1 (2022), pp. 45–62.

[17] Hana Ben Fredj, Mouna Ltaif, Anis Ammar, and Chokri Souani. 'Parallel implementation of Sobel filter using CUDA.' In: *2017 International Conference on Control, Automation and Diagnosis (ICCAD)*. IEEE. 2017, pp. 209–212.

[18] Rafael C Gonzalez and Richard E Woods. 'Digital image processing 4th edition, global edition.' In: (2018).

[19] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 'Generative adversarial networks.' In: *Communications of the ACM* 63.11 (2020), pp. 139–144.

[21] Jiuxiang Gu et al. 'Recent advances in convolutional neural networks.' In: *Pattern Recognition* 77 (2018), pp. 354–377. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2017.10.013. URL: https://www.sciencedirect.com/science/article/pii/S0031320317304120.

[22] Zhen Guo, Jung Ki Song, George Barbastathis, Michael E. Glinsky, Courtenay T. Vaughan, Kurt W. Larson, Bradley K. Alpert, and Zachary H. Levine. 'Physics-assisted generative adversarial network for X-ray tomography.' In: *Opt. Express* 30.13 (2022), pp. 23238–23259. DOI: 10.1364/OE.460208. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-13-23238.

[23] George Harauz and Marin van Heel. 'Exact filters for general geometry three dimensional reconstruction.' In: *Optik.* 73.4 (1986), pp. 146–156.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 'Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.' In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1026–1034.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 'Deep residual learning for image recognition.' In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[26] Yaoshiang Ho and Samuel Wookey. 'The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling.' In: *IEEE Access* 8 (2020), pp. 4806–4813. DOI: 10.1109/ACCESS.2019.2962617.

[27] Alain Hore and Djemel Ziou. 'Image quality metrics: PSNR vs. SSIM.' In: *2010 20th international conference on pattern recognition*. IEEE. 2010, pp. 2366–2369.

[28] Stefan Iglauer, Adriana Paluszny, Christopher H Pentland, and Martin J Blunt. 'Residual CO2 imaged with X-ray micro-tomography.' In: *Geophysical Research Letters* 38.21 (2011).

[29] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001.

[30] Daniil Kazantsev, Valery Pickalov, Srikanth Nagella, Edoardo Pasca, and Philip J. Withers. 'TomoPhantom, a software package to generate 2D–4D analytical phantoms for CT image reconstruction algorithm benchmarks.' In: *SoftwareX* 7 (2018), pp. 150–155. ISSN: 2352-7110. DOI: https://doi.org/10.1016/j.softx.2018.05.003. URL: https://www.sciencedirect.com/science/article/pii/S2352711018300335.

[31] Zoran Kotevski and Pece Mitrevski. 'Experimental Comparison of PSNR and SSIM Metrics for Video Quality Estimation.' In: *ICT Innovations 2009*. Ed. by Danco Davcev and Jorge Marx Gómez. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 357–366. ISBN: 978-3-642-10781-8.

[32] Yann LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. 'Efficient backprop.' In: *Neural networks: Tricks of the trade*. Springer, 2002, pp. 9–50.

[33] Zhengchun Liu, Tekin Bicer, Rajkumar Kettimuthu, Doga Gursoy, Francesco De Carlo, and Ian Foster. 'TomoGAN: low-dose synchrotron x-ray tomography with generative adversarial networks: discussion.' In: *JOSA A* 37.3 (2020), pp. 422–434.

[34] Mario Livio. *The golden ratio: The story of phi, the world's most astonishing number*. Crown, 2008.

[35] Des McMorrow and Jens Als-Nielsen. *Elements of modern X-ray physics*. John Wiley & Sons, 2011.

[36] Mehdi Mirza and Simon Osindero. 'Conditional generative adversarial nets.' In: *arXiv preprint arXiv:1411.1784* (2014).

[37] Grégoire Montavon, Geneviève Orr, and Klaus-Robert Müller. *Neural networks: tricks of the trade*. Vol. 7700. springer, 2012.

[38] Ilia Nouretdinov, Volodya Vovk, Michael Vyugin, Alex Gammerman, et al. 'Pattern recognition and density estimation under the general iid assumption.' In: *Lecture notes in computer science* 2111 (2001), pp. 337–353.

[39] OpenAI. *ChatGPT: Language Model*. Retrieved from https://openai.com/research/chatgpt. 2021.

[40] Adam Paszke et al. 'PyTorch: An Imperative Style, High-Performance Deep Learning Library.' In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[41] Tony Plate. 'Controlling the hyperparameter search in MacKay's Bayesian neural network framework.' In: *Neural networks: Tricks of the trade* (1998), pp. 93–112.

[42] Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin. 'TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning.' In: *Computer Methods and Programs in Biomedicine* 208 (2021), p. 106236. ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2021.106236. URL: https://www.sciencedirect.com/science/article/pii/S0169260721003102.

[43] Nina Rohringer. 'Introduction to the theory of x-ray matter interaction.' In: *Physics of and Science with X-Ray Free-Electron Lasers* 199 (2020), p. 71.

[44] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 'U-net: Convolutional networks for biomedical image segmentation.' In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241.

[45] Sebastian Ruder. 'An overview of gradient descent optimization algorithms.' In: *arXiv preprint arXiv:1609.04747* (2016).

[46] Robin Santra. 'Concepts in x-ray physics.' In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 42.2 (2008), p. 023001.

[47] Jonathan Richard Shewchuk et al. *An introduction to the conjugate gradient method without the agonizing pain*. 1994.

[48] Changrong Shi, Yongshun Xiao, and Zhiqiang Chen. 'Dual-domain sparse-view CT reconstruction with Transformers.' In: *Physica Medica* 101 (2022), pp. 1–7.

[49] Changrong Shi, Yongshun Xiao, and Zhiqiang Chen. 'Dual-domain sparse-view CT reconstruction with Transformers.' In: *Physica Medica* 101 (2022), pp. 1–7. ISSN: 1120-1797. DOI: https://doi.org/10.1016/j.ejmp.2022.07.001. URL: https://www.sciencedirect.com/science/article/pii/S1120179722020154.

[50] Jan Malte Slowik. 'Quantum effects in nonresonant x-ray scattering.' PhD thesis. Universität Hamburg Hamburg, 2015.

[51] Vera W de Spinadel. 'The family of metallic means.' In: *Vis. Math* 1.3 (1999), pp. 1–16.

[52] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 'Attention is All you Need.' In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[53] Pauli Virtanen et al. 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.' In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[54] Toru Wakasa, Yuzuru Higuchi, Miki Hisatomi, Hideki Aiga, Yasutoshi Honda, and Kanji Kishi. 'Application of dynamic CT for various diseases in the oral and maxillofacial region.' In: *European Journal of Radiology* 44.1 (2002), pp. 10–15. ISSN: 0720-048X. DOI: https://doi.org/10.1016/S0720-048X(01)00369-2. URL: https://www.sciencedirect.com/science/article/pii/S0720048X01003692.

[55] Kunfeng Wang, Chao Gou, Yanjie Duan, Yilun Lin, Xinhu Zheng, and Fei-Yue Wang. 'Generative adversarial networks: introduction and outlook.' In: *IEEE/-CAA Journal of Automatica Sinica* 4.4 (2017), pp. 588–598.

[56] Yukai Wang, Qizhi Teng, Xiaohai He, Junxi Feng, and Tingrong Zhang. 'CT-image of rock samples super resolution using 3D convolutional neural network.' In: *Computers & Geosciences* 133 (2019), p. 104314.

[57] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 'Image quality assessment: from error visibility to structural similarity.' In: *IEEE transactions on image processing* 13.4 (2004), pp. 600–612.

[58] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. 'A survey of transfer learning.' In: *Journal of Big data* 3.1 (2016), pp. 1–40.

[59] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. 'Convolutional neural networks: an overview and application in radiology.' In: *Insights into imaging* 9 (2018), pp. 611–629.

[60] Xiaogang Yang, Maik Kahnt, Dennis Brückner, Andreas Schropp, Yakub Fam, Johannes Becher, J-D Grunwaldt, Thomas L Sheppard, and Christian G Schroer. 'Tomographic reconstruction with a generative adversarial network.' In: *Journal of synchrotron radiation* 27.2 (2020), pp. 486–493.

[61] Marek Zemek, Jakub Šalplachta, Tomáš Zikmund, Kazuhiko Omote, Yoshihiro Takeda, Peter Oberta, and Jozef Kaiser. 'Automatic marker-free estimation methods for the axis of rotation in sub-micron X-ray computed tomography.' In: *Tomography of Materials and Structures* 1 (2023), p. 100002. ISSN: 2949-673X. DOI: https://doi.org/10.1016/j.tmater.2022.100002. URL: https://www.sciencedirect.com/science/article/pii/S2949673X2200002X.

[62] Gengsheng Lawrence Zeng. *Medical image reconstruction: a conceptual tutorial.* Springer, 2010, pp. 10–24.

[63] Hai-Miao Zhang and Bin Dong. 'A review on deep learning in medical image reconstruction.' In: *Journal of the Operations Research Society of China* 8 (2020), pp. 311–340.

APPENDIX A

A.1 THE QUANTUM MECHANICS OF X-RAY ABSORPTION

For the most complete description of the role of X-ray radiation in CT, a quantum mechanical description of X-ray absorption is required. Quantum Mechanics introduces the quantisation of light, in the form of photons, which allows for a more intuitive description. However, a sufficient description of this evolves around several lenghty derivations, well explained by Robin Santra [46], Nina Rohringer [43], and Jan Malte Slowik [50]. Therefore, big leaps between the essential aspects are instead included in this section. This derivation is also heavily based upon the quantum mechanical explanation used during the project thesis on Small Angle X-ray Scattering Tensor Tomography (SAXSTT). Logically, the focus of this derivation will be X-ray absorption, as this is the most relevant for CT, while X-ray scattering was more relevant in the project thesis.

A quantum mechanical system must be described by a Hamiltonian. The system consists of X-rays, matter, and interactions between them. The total Hamiltonian $\hat{H}$ is therefore a sum of a matter Hamiltonian $\hat{H}_{\text{part}}$, a radiation field Hamiltonian $\hat{H}_{\text{rad}}$, and an interaction Hamiltonian $\hat{H}_{\text{int}}$,

$$\hat{H} = \hat{H}_{\text{part}} + \hat{H}_{\text{rad}} + \hat{H}_{\text{int}}. \tag{A.1}$$

The Hamiltonian of the electromagnetic field, $\hat{H}_{\text{rad}}$, follows from the quantisation of the vector potential $\widehat{\mathbf{A}}$, adopting the Coulumb gauge, $\boldsymbol{\nabla} \cdot \widehat{\mathbf{A}} = 0$. Applying the assumptions of free, coherent, and monochromatic X-rays, the vector potential is expressed as

$$\widehat{\mathbf{A}}(\mathbf{r}) = \sum_{\mathbf{k},\lambda} \sqrt{\frac{2\pi}{V\omega_{\mathbf{k}}\alpha^2}} \left(\hat{a}_{\mathbf{k},\lambda} \boldsymbol{\epsilon}_{\mathbf{k},\lambda} e^{i\mathbf{k}\cdot\mathbf{r}} + \hat{a}^{\dagger}_{\mathbf{k},\lambda} \boldsymbol{\epsilon}^{*}_{\mathbf{k},\lambda} e^{-i\mathbf{k}\cdot\mathbf{r}} \right). \tag{A.2}$$

The essential parameters from Equation (A.2) are the creation and annihilation operators, $\hat{a}^{\dagger}_{\mathbf{k},\lambda}$ and $\hat{a}_{\mathbf{k},\lambda}$, respectively. These operators are used to describe absorption as an annihilation of a photon. Moreover, the normalisation constant reveals the electromagnetic field to be similar to the well-known "Harmonic Oscillator" [35]. Consequently, H_{rad} becomes

$$\hat{H}_{\text{rad}} = \sum_{\mathbf{k},\lambda} \left(\hbar\omega_{\mathbf{k}} \hat{a}^{\dagger}_{\mathbf{k},\lambda} \hat{a}_{\mathbf{k},\lambda} \right). \tag{A.3}$$

The Hamiltonian of the matter, $\hat{H}_{\text{part}}$, is the sum of the kinetic energy operator, potential energy operator, and the electronic Hamiltonian operator:

$$\hat{H}_{\text{part}} = \hat{T}_N + \hat{V}_{NN} + \hat{H}_{\text{el}}. \tag{A.4}$$

These are, though, of no particular interest concerning the process of absorption, and are therefore not elaborated upon.

The interaction Hamiltonian, $\hat{H}_{\text{int}}$ accounts for the interaction between the electromagnetic field and the matter. It can be shown that interactions between an electromagnetic field and charges q may be accounted for by performing the substitution

$\mathbf{p} \to \mathbf{p} - q\mathbf{A}$ [35]. Eventually, the interaction Hamiltonian is revealed to have a term linear and a term quadratic in the vector potential $\widehat{\mathbf{A}}$:

$$\hat{H}_{int} \sim \widehat{\mathbf{p}} \cdot \widehat{\mathbf{A}} + \widehat{\mathbf{A}}^2. \tag{A.5}$$

As the first term is only linear in $\widehat{\mathbf{A}}$, it can only account for absorption of X-rays through annihilation of photons. The second term is quadratic, and therefore accounts for annihilation and subsequent creation of photons. This follows from squaring Equation (A.2), which gives terms on the form $a^\dagger_{\mathbf{k},\lambda} a_{\mathbf{k},\lambda}$.

Furthermore, second quantisation is more fitting for describing the electronic Hamiltonian and the interaction Hamiltonian when the system consists of many electrons, as the Hamiltonian in second quantisation is independent of the number of electrons [43]. Here, Fock states are introduced, and the annihilation and creation operators are replaced by annihilation and creation field operators. The electronic density operator $\hat{n}(\mathbf{r})$ may be expressed in terms of these field operators as

$$\hat{n}(\mathbf{r}) = \hat{\psi}^\dagger(\mathbf{r})\hat{\psi}(\mathbf{r}). \tag{A.6}$$

Next, the interaction Hamiltonian is considered a small time-dependent perturbation. From the *Time-Dependent Perturbation Theory* in the interaction picture, the transition rate Γ_{IF} is derived from the transition amplitude squared. It is given as

$$\Gamma_{IF} = 2\pi\delta\left(E_F - E_I\right)\left|\left\langle F\left|\hat{H}_{int}\right|I\right\rangle + \sum_M \frac{\left\langle F\left|\hat{H}_{int}\right|M\right\rangle\left\langle M\left|\hat{H}_{int}\right|I\right\rangle}{E_I - E_M + i\epsilon} + ...\right|^2, \tag{A.7}$$

where ϵ is a small positive number, and M is a state in between the initial and final state.

With the transition rate derived from time-dependent perturbation of the interaction Hamiltonian, X-ray absorption can be investigated. X-ray absorption will in this case be defined as the process where the final state of the photon field is $N_{EM} - 1$, where N_{EM} is the Fock state of the photon field before the process. Put plainly, this means that absorption is, as mentioned, annihilation of one photon [46]. If $\Psi_F^{N_{el}}$ is the electronic state of the system after the process, one describes the absorption process by evaluating the transition rate (A.7) with the final state expressed as

$$F = \left|\Psi_F^{N_{el}}\right\rangle |N_{EM} - 1\rangle. \tag{A.8}$$

By skipping the intermediate calculations, one arrives directly at the absorption cross section, σ_F, which is given as

$$\begin{aligned}\sigma_F(\mathbf{k}_{in}, \lambda_{in}) = {} & \frac{4\pi^2}{\omega_{in}}\alpha\delta\left(E_F^{N_{el}} - E_I^{N_{el}} - \omega_{in}\right) \\ & \times \left|\left\langle \Psi_F^{N_{el}} \left| \int d^3x \hat{\psi}^\dagger(\mathbf{x}) e^{i\mathbf{k}_{in}\cdot\mathbf{x}} \boldsymbol{\epsilon}_{\mathbf{k}_{in},\lambda_{in}} \cdot \frac{\boldsymbol{\nabla}}{i} \hat{\psi}(\mathbf{x}) \right| \Psi_I^{N_{el}} \right\rangle\right|^2,\end{aligned} \tag{A.9}$$

with the essential parameters being α, $\mathbf{k}_{in}$, and $\boldsymbol{\epsilon}$. They denote the fine-structure constant, the incoming wave vector, and the polarisation of the wave, respectively.

This cross section expression may be simplified by introducing orthonormal spin orbitals φ. By these definitions, the absorption cross section may be written as

$$\begin{aligned}\sigma_F(\mathbf{k}_{in}, \lambda_{in}) = \frac{4\pi^2}{\omega_{in}} \alpha \delta \left(E_F^{N_{el}} - E_I^{N_{el}} - \omega_{in} \right) \\ \times \left| \sum_{p,q} \left\langle \varphi_p \left| e^{i\mathbf{k}_{in}\cdot\mathbf{x}} \boldsymbol{\epsilon}_{\mathbf{k}_{in},\lambda_{in}} \cdot \frac{\boldsymbol{\nabla}}{i} \right| \varphi_q \right\rangle \left\langle \Psi_F^{N_{el}} \left| \hat{c}_p^\dagger \hat{c}_q \right| \Psi_I^{N_{el}} \right\rangle \right|^2 ,\end{aligned} \tag{A.10}$$

where $\hat{c}_p^\dagger$ and $\hat{c}_q$ are the annihilation and creation operators for the spin orbitals, respectively.

A.2 INTRODUCTION TO GITHUB REPOSITORY

All code can be found, in its raw and unedited form, in the GitHub repository `https://github.com/RubenDragland/4D_CT`. A short illustration of the workflow associated with the repository is given in the file *workflow.ipynb*. This section is a short description of the different folders and files in the repository. Most of the necessary packages and libraries for the Python scripts can be found in the respective *requirements.txt* files.

3DTomoGAN

This folder contains code for the three-dimensional GAN, and *model.py* is where the generator and discriminator are defined using Pytorch [40]. Included in this folder is also code for loading the data during training, including data augmentation techniques. The actual scripts for training, with prefix *gym*, can also be found; one general version and one for training on a specific dataset. Additionally, scripts for training on distributed GPUs are denoted with the prefix *ddp*. The *utils.py* file holds useful functions for the training process and the evaluation process.

Analysis

The data analysis can be found in folders that include the word *analysis*, and is written using Jupyter Notebooks.

CT

The folder named *CT* holds everything involving the TIGRE toolbox, and therefore anything related to projections and reconstructions. The files with *preprocessing* as prefix were used for preprocessing of CT measurements. Processed projections were utilised further in *data_making.py* where training datasets and 4D reconstructions can be made. The used CT geometries were also defined in this file.

B

HOURGLASS 3D-MODEL AND TIME SERIES

As a final demonstration of the capabilities of 4D-CT using golden ratio sampling and deep neural network enhancement, the four dimensions of the hourglass scan are included in this appendix. A three-dimensional model of the sand-air interface in the hourglass sample is included in Figure B.1 to provide a three-dimensional visualisation of the fascinating flow pattern, where some sort of whirlpool is formed. The movement of the sand-air interface is also illustrated in a time series of the 4D-CT scan, presented in Figure B.2. For the purpose of simplicity, the time series consists of 2D slices instead of entire 3D reconstructions. Finally, if you as the reader have not yet scanned the QR codes on the front page of this thesis, you are encouraged to do so now. Enjoy!

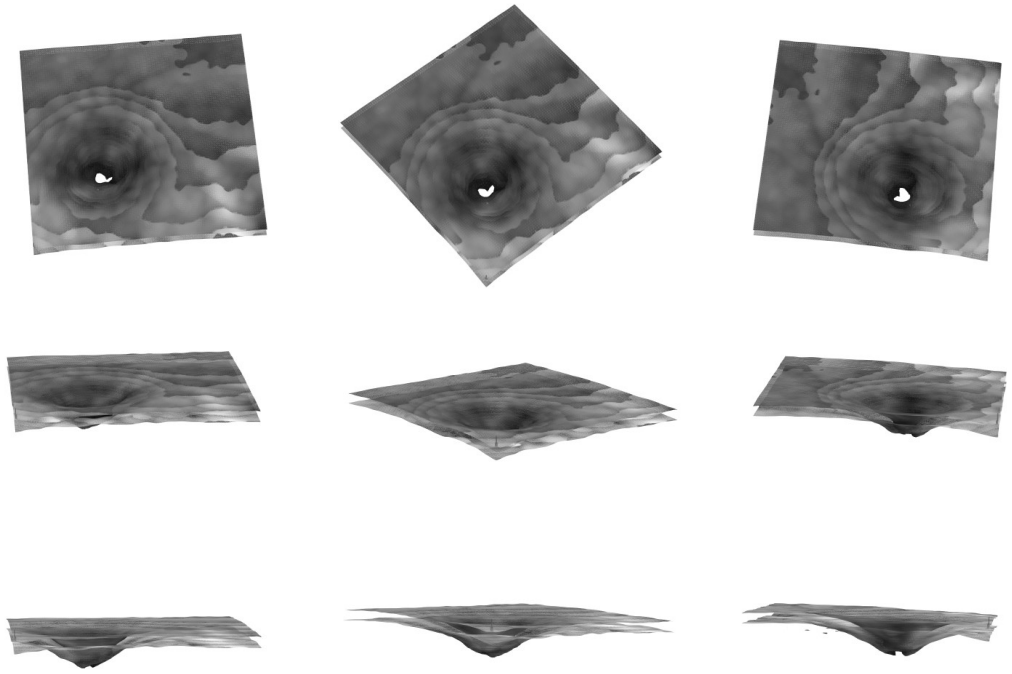


Figure B.1: A three-dimensional model of the sand-air interface in the hourglass sample is shown. The interface is shown from different azimuthal and polar angles. Note that the surface detection algorithm clearly struggled somewhat due to noise and dynamical artefacts in the undersampled reconstruction. The model is a visualisation of the flow pattern in the hourglass sample, where some sort of whirlpool is formed.

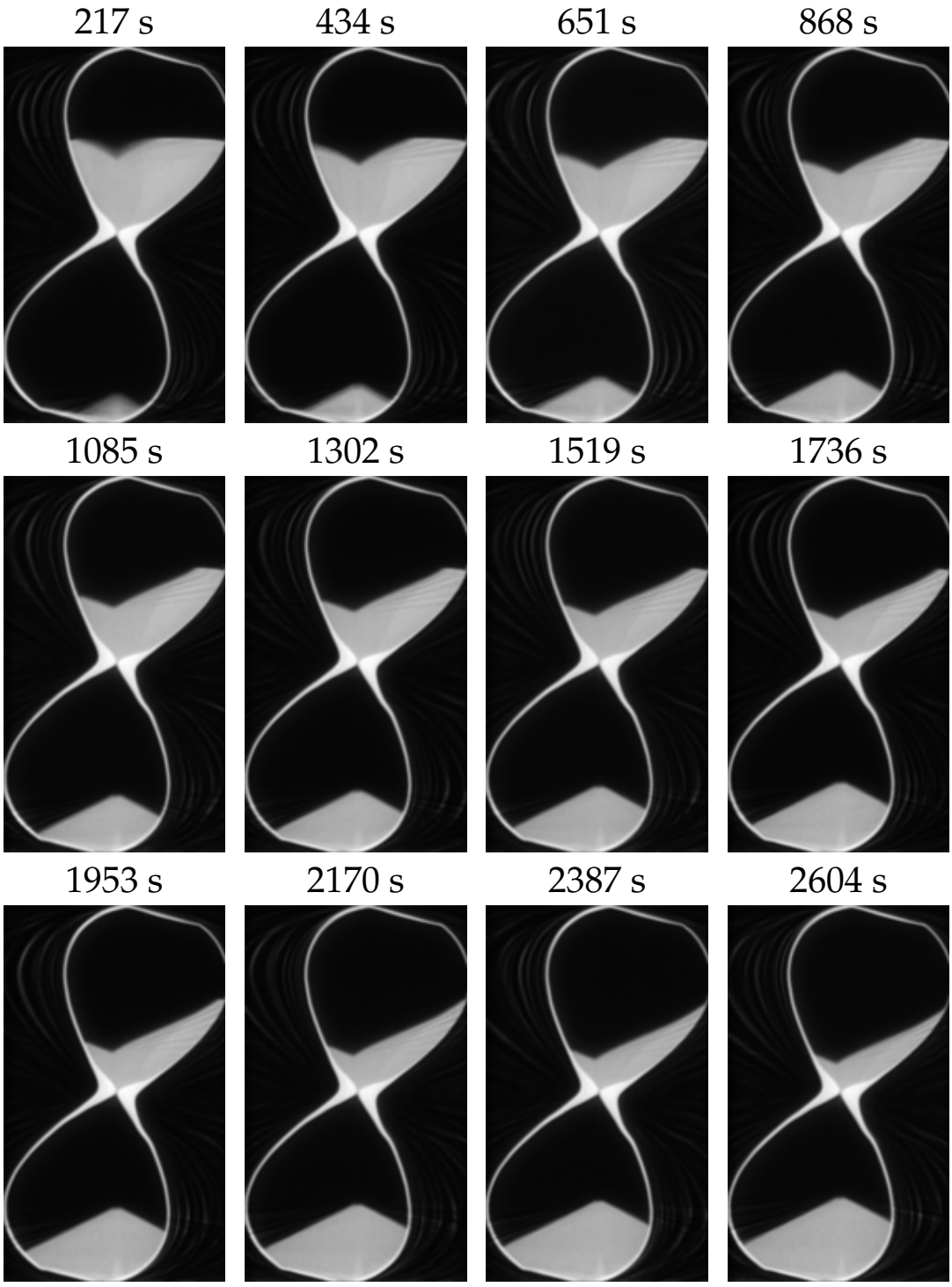


Figure B.2: A time-resolved illustration of the dynamics within the hourglass sample is shown. The middlemost slice of the 3D reconstruction is shown for each time step. 52 projections were used in each time step, as this gave the optimal compromise between temporal and spatial resolution. Note, however, that 13 projections were collected per revolution for the benefit of flexibility in the post-processing. Therefore, the observed temporal resolution was 217 s, and undersampling artefacts were almost completely diminished by the GAN. The time series revealed that the flow pattern of the sand was some sort of whirlpool.

C

ADDITIONAL GAN TRAINING AND TESTING

C.1 GAN WITH PRIOR KNOWLEDGE OF STATIC CONTENT

To address the challenge of achieving proper reconstruction enhancement with regard to the sandstone sample, inspiration was drawn from the literature. One reconstruction algorithm for undersampled dynamical CT reconstructions with sparsity in the time domain is the PICCS algorithm [10]. This algorithm combines a high-quality static reconstruction of the matrix and a low-quality reconstruction of the dynamical process. These constraints can be assumed and applied to the application of the GAN to better describe the complex probability distribution of the sandstone sample. In a final effort to maximise the performance of the GAN, it was trained solemnly on the sandstone sample. The ground truth was a FDK reconstruction from the 1000 golden ratio sampled projections obtained in an initial testing of the derived sampling procedure. The input data was a undersampled SIRT reconstruction using 150 iterations and 34 projections, effectively a relative undersampling factor of 29. Additionally, the GAN was also re-trained using only the sandstone sample reconstructed from 50 projections using FDK. Table C.1 provides a quantitative comparison of the different GAN enhancements in the entire 256^3 volume. It has to be noted that the pixel-wise loss was very high compared to the previous results, indicating the presence of outliers during renormalisation. This would also affect the other metrics, as both structural similarity indexes used compare the mean intensities. Moreover, cross sections in the XY-, XZ-, and YZ-planes are shown in Figure C.1 for the fully sampled ground truth, the FDK-GAN, and the SIRT-GAN, respectively. These cross sections confirm that the features of the fully sampled ground truth were mostly recovered by the GANs, even though some noise and blurring were still present.

Table C.1: The global SSIM, MSSIM, and PSNR of the GAN enhancement of the sandstone sample are shown. This time, the GAN was trained only using prior knowledge of the sandstone sample from another CT scan.

Prior Knowledge GAN			
Projections	**SSIM**	**MSSIM**	**PSNR**
85 **FDK-GAN**	0.76	0.64	15.15
34 **SIRT-GAN**	0.54	0.63	11.45

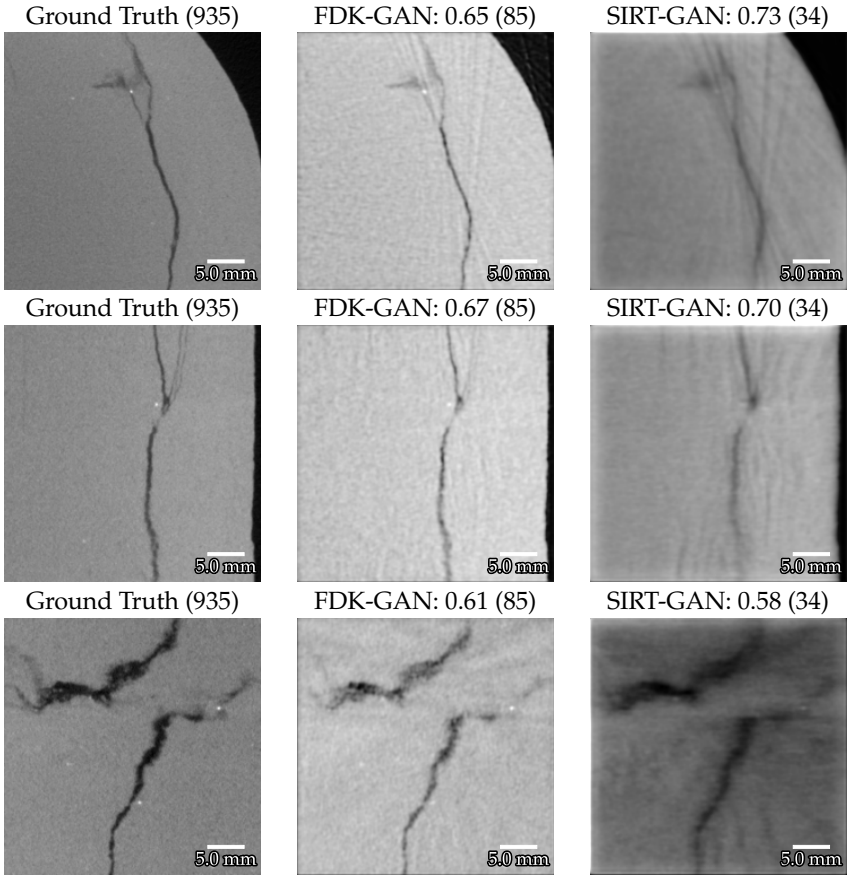


Figure C.1: The ground truth, the FDK-GAN trained on the sandstone sample, and the SIRT-GAN trained on the sandstone sample are shown from left to right, respectively. MSSIM values are listed above each slice, together with the number of projections used for the reconstruction in parenthesis. This time the number of projections used for the FDK-GAN was 85, while the number of projections used for the SIRT-GAN was 34. The FDK-GAN managed to resemble the ground truth with minor streaking artefacts with an undersampling factor of 13 at the sample edge, while the SIRT-GAN provided the highest enhancement to be recorded in this project given the severe undersampling factor of 32 at the sample edge. As has been the case when presenting the sandstone sample, the XY-plane is shown in the first row, while the XZ-plane and YZ-plane are shown in the second row and third row, respectively.

C.2 SANDSTONE ENHANCEMENT AFTER MINIBATCH TRAINING

One additional training session similar to the one presented in Section 10.2 was performed. The motivation behind the test was to observe enhanced training stability, and to find a model that would perform generally better for both the hourglass and sandstone sample. Therefore, a natural adjustment was to utilise a minibatch size of 4 instead of 1. Due to memory limitations, this meant spatial dimensions of 64^3 instead of 128^3 had to be used.

After only 15 epochs of training in addition to the pre-training performed on simulated data, a set with weights was saved because of promising enhancement of the validation data. For the sandstone sample, the MSSIM after enhancement improved compared to the previous results in Section 10.21. On the other hand, streaking artefacts remained for all number of projections used, but were less pronounced for the reconstruction from 85 projections, as shown in Figure C.2. However, the hourglass sample was not enhanced as well as the previous results in Section 10.2. This is an indication that the training of the GAN in most cases has been ended prematurely, and that whether the stored weights were optimal for the hourglass or sandstone sample was a matter of coincidence. To elaborate, it is possible that the training progress of the GAN would alternate between converging for the hourglass sample and the sandstone sample. However, convergence towards a compromise between all samples is desired. Therefore, as mentioned in Section 12.6, more emphasis on the fined-tuned training of the GAN, exemplified through increased batch size, is necessary in future work.

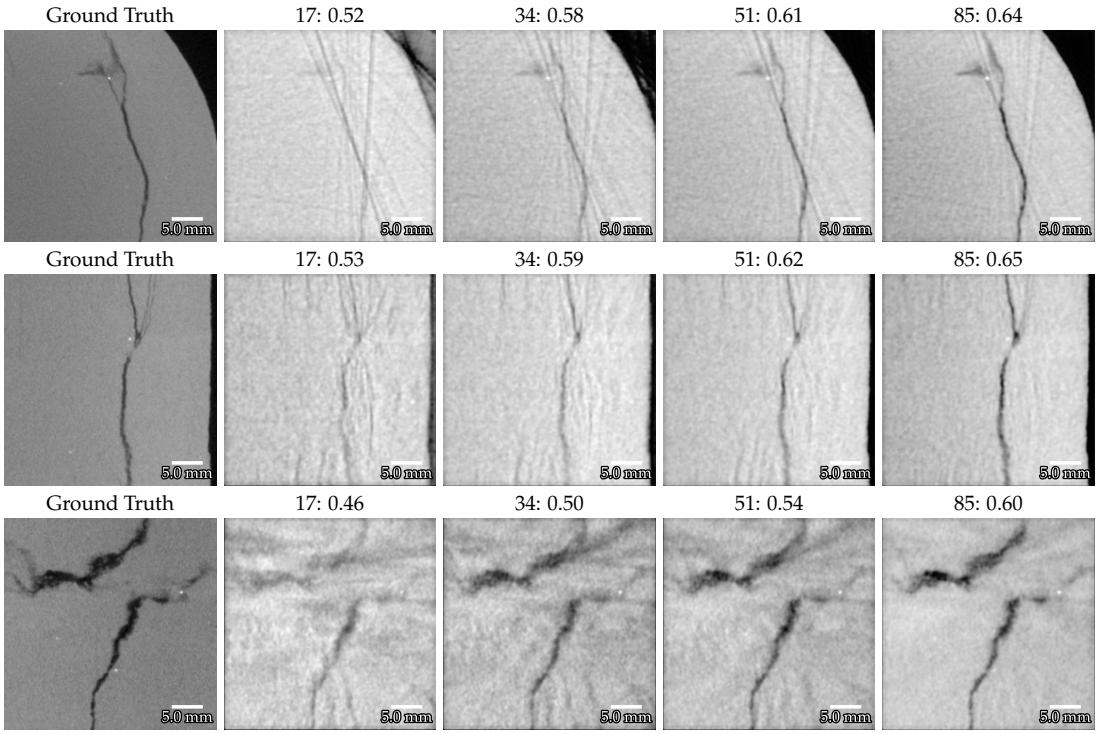
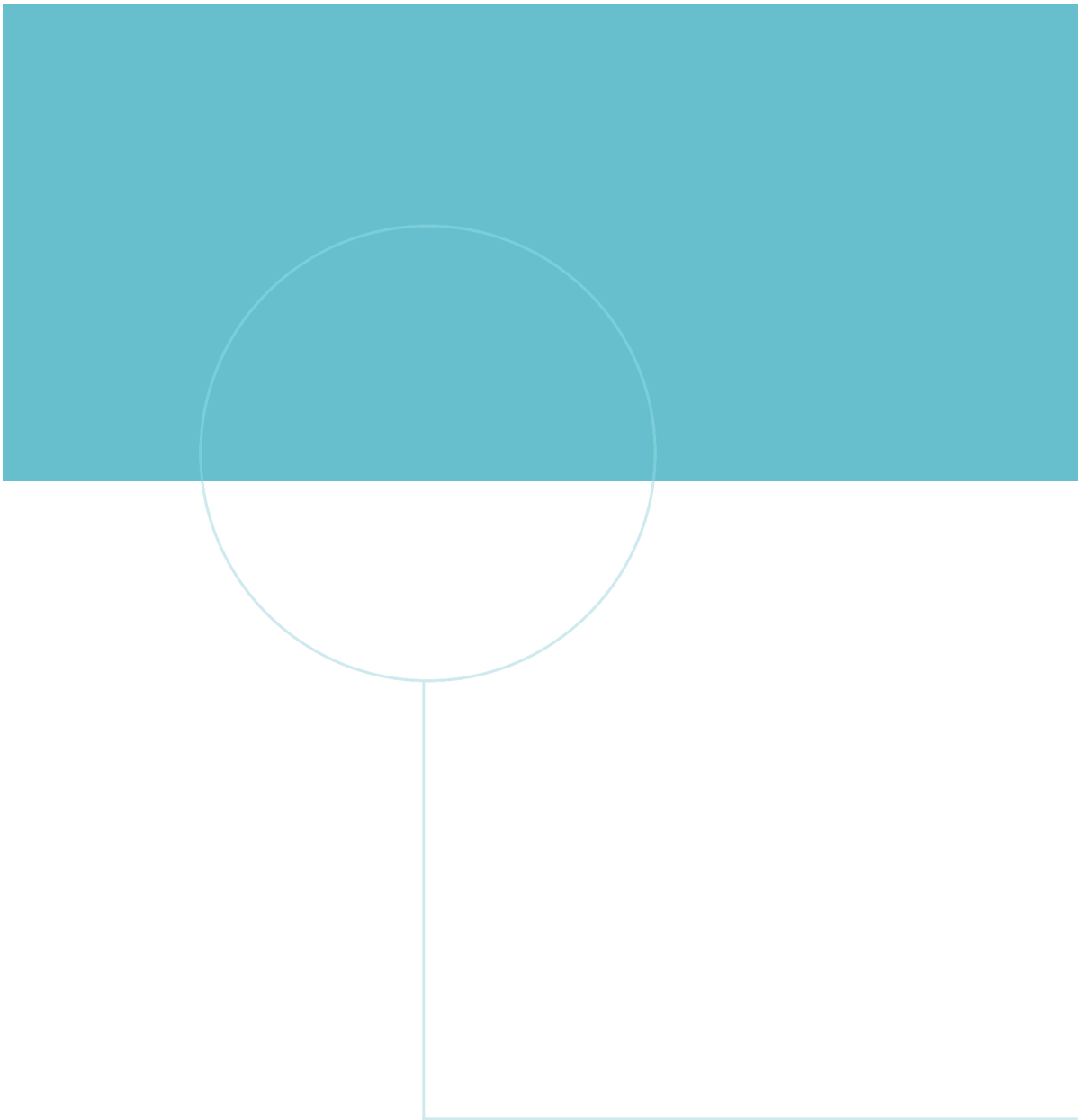


Figure C.2: The ground truth compared to the GAN-enhanced reconstructions of the sandstone sample for increasing number of projections. Each slice is described with the following format: *projections: MSSIM*. The GAN was trained on the complete dataset 5.2 using a minibatch size of 4 instead of 1. For this reason, the spatial dimensions of the input and output data were reduced to 64^3 instead of 128^3. This resulted in improved results compared to the previous results in Figure 10.21. With 85 projections, the undersampling given the width of the sample was 13, the obtained MSSIM was 0.64, and streaking artefacts were somewhat reduced by the GAN.



NTNU
Norwegian University of
Science and Technology

equinor