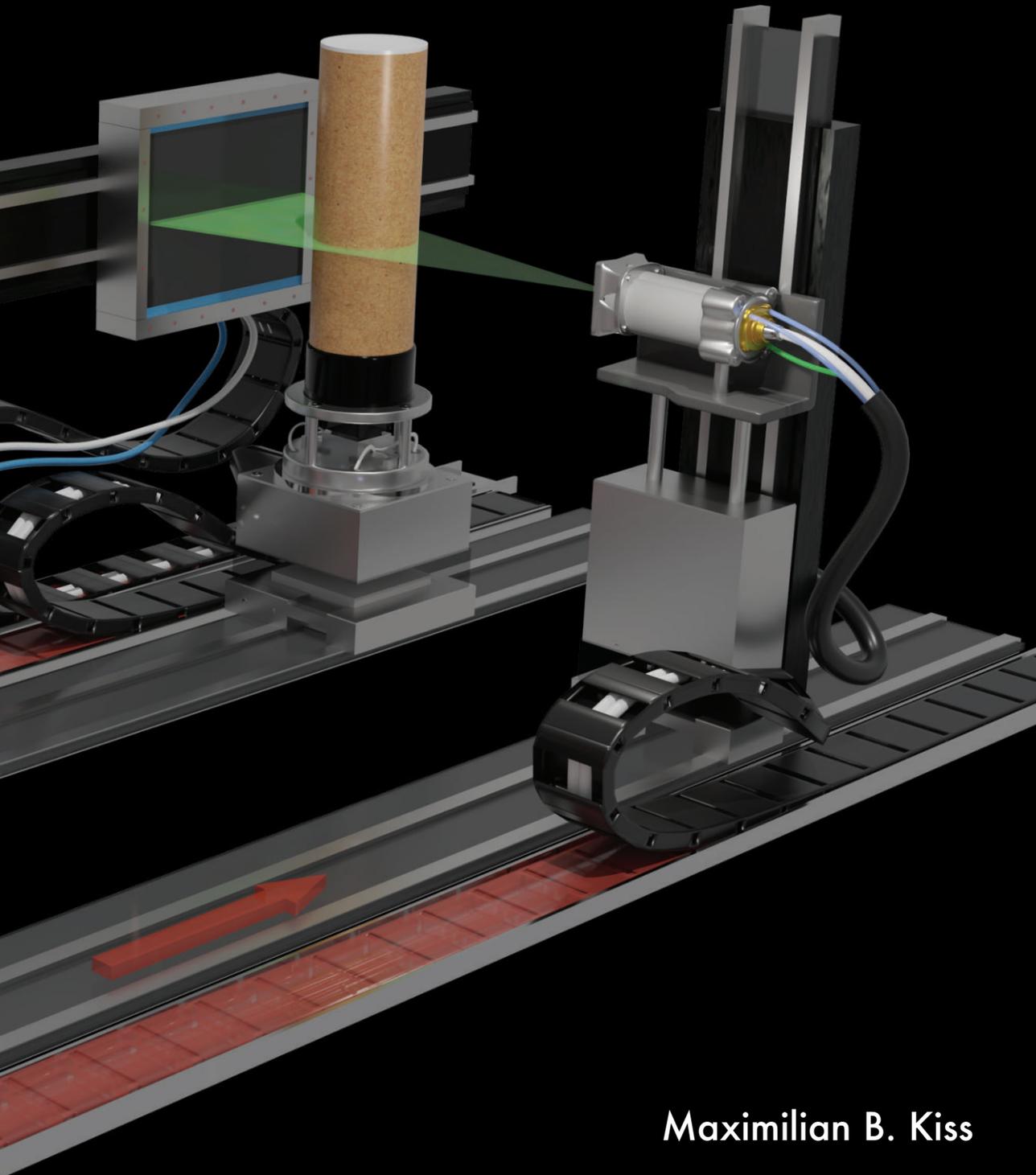


Advancing learned algorithms for 2D X-ray computed tomography



Maximilian B. Kiss

Advancing learned algorithms for 2D X-ray computed tomography

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op vrijdag 7 november 2025
klokke 13.00 uur

door

Maximilian Benjamin Kiss

geboren te Regensburg, Duitsland
in 1995

Promotores:

Prof.dr. K. J. Batenburg

Prof.dr. T. van Leeuwen

Co-promotor:

Dr. F. Lucka

Promotiecommissie:

Prof.dr. M. M. Bonsangue

Prof.dr.ir. F. J. Verbeek

Prof.dr. C. Brune (Universiteit Twente)

Prof.dr. A. Hauptmann (University of Oulu)

Prof.dr. C.-B. Schönlieb (University of Cambridge)

The research presented in this dissertation was carried out at the Centrum Wiskunde & Informatica (CWI) in Amsterdam and in collaboration with the Cambridge Image Analysis (CIA) group at the Department for Applied Mathematics and Theoretical Physics (DAMTP) of the University of Cambridge (chapters 5 and 6).

Financial support was provided by the Dutch Research Council (NWO), project number OCENW.KLEIN.285. The FleX-ray Laboratory is supported by the Dutch Research Council, project number 639.073.506.

Cover design by Paul van Laar and Johannes Krauß.

Printing by Ridderprint.

© 2025 Maximilian B. Kiss

Contents

1	Introduction	1
1.1	Computed tomography	1
1.1.1	General introduction	1
1.1.2	Tomographic reconstruction as an ill-posed inverse problem	3
1.2	Challenges for computed tomography	5
1.2.1	Objects	5
1.2.2	Imaging	5
1.3	Machine learning	12
1.4	Machine learning for computed tomography	15
1.5	Research questions	18
2	Tailoring CT Acquisitions in the FleX-ray Lab	25
2.1	Functionalities of the FleX-ray scanner	25
2.2	Extensions to the FleX-ray scanner	26
2.2.1	Sample stage disks	26
2.2.2	Sample stage cylinders	27
2.2.3	Beam filtration	30
2.3	Guideline for choosing suitable acquisition parameters	33
3	Beam filtration for object-tailored X-ray CT of multi-material cultural heritage objects	37
3.1	Background	40
3.1.1	Polychromatic X-ray spectrum and metal artifacts	40
3.1.2	Contrast between materials	41
3.1.3	Noise and detector saturation	41
3.2	Methods	42
3.2.1	Concept of beam filtration	42
3.2.2	Hard- and software	43
3.3	Experiments and results	45
3.3.1	The case studies	46
3.3.2	General concept of steps for an object-tailored CT scan	56
3.4	Discussion and conclusion	57

4	2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning	61
4.1	Methods	64
4.1.1	X-ray computed tomography scanner	64
4.1.2	Experimental design	64
4.1.3	Data acquisition	71
4.1.4	Computational processing	72
4.2	Data records	76
4.3	Technical validation	77
4.4	Usage notes	77
4.4.1	Raw projection data	77
4.4.2	Reconstructions and segmentations	77
4.4.3	Expansion possibilities	77
4.4.4	Further usage	78
4.5	Code availability	78
5	Learned denoising with simulated and experimental low-dose CT data	81
5.1	Methods	83
5.1.1	Related work	83
5.1.2	Noise simulation	84
5.1.3	Training data	86
5.1.4	Method development	89
5.1.5	Comparison metrics	89
5.1.6	Computational experiments	90
5.2	Results and discussion	91
5.2.1	Empirical selection of noise level	91
5.2.2	Sinogram denoising	92
5.2.3	Optimization in the reconstruction domain: mapping directly from Sinogram to Reconstruction	96
5.3	Discussion and conclusions	96
5.4	Code availability	99
6	Benchmarking Learned Algorithms for Computed Tomography Image Reconstruction Tasks	101
6.1	Related work	103
6.2	Data-driven CT reconstruction	104
6.2.1	Tomographic reconstruction as a linear inverse problem	104
6.2.2	Data-driven methods for tomographic reconstruction	105
6.3	Benchmark design	107
6.3.1	CT Image reconstruction tasks	109
6.3.2	Pipeline	110
6.3.3	Performance metrics	111
6.4	Numerical experiments	112
6.4.1	Pre-processing	112
6.4.2	Evaluated methods	112

6.4.3	Training details	115
6.5	Results and discussion	117
6.5.1	Relevance and difficulty of CT image reconstruction tasks . . .	117
6.5.2	Quantitative and qualitative analysis of the evaluated methods	123
6.5.3	Limitations and broader impact	124
6.5.4	Code and data availability	126
6.6	Conclusion	126
7	Conclusion and outlook	129
7.1	Contributions and limitations	129
7.2	Outlook	132
	Bibliography	135
	List of publications	155
	Samenvatting	157
	Curriculum Vitae	165
	Acknowledgements	167

1

Introduction

The contents of this thesis revolve around four general topics that will be introduced in the following: the technique of computed tomography (CT), the challenges associated with acquiring CT scans, the possibilities to tailor this acquisition, and the use of machine learning for CT reconstruction. In the first section of this introduction the technique of computed tomography is introduced as well as its application areas and mathematical background. Furthermore, important concepts and technical terms needed for the rest of this thesis will be explained. In the second section, challenges for computed tomography imposed by either the objects to be scanned or by the circumstance of the imaging process are discussed. In the third section, the possibilities to tailor CT acquisitions to overcome these challenges are presented. In particular, the functionalities of the FleX-ray scanner at the Centrum Wiskunde & Informatica and the extensions introduced to it in the course of this PhD thesis. In the fourth section, the field of machine learning is generally introduced and the importance and role of open-access datasets are discussed. In the two remaining sections of this introduction, research at the intersection of the first two areas, computed tomography and machine learning, is presented and the research questions of this thesis are introduced.

1.1 Computed tomography

1.1.1 General introduction

Computed tomography is a non-invasive X-ray absorption contrast technique that has been used in a range of fields, including medicine, materials science, the food industry, and the manufacturing industry. While in the medical sector the geometry and acquisition protocols of state-of-the-art CT scanners have been optimized to the general characteristics of a patient's body, the other fields commonly utilize more versatile laboratory equipment called micro-CT scanners [83] (cf. Figure 1.1).



Figure 1.1: Comparison of a medical CT scanner (on the left, photo credit [44]) and a laboratory micro-CT scanner (on the right).

Basic operating principles of micro-CT scanners The setup of such a micro-CT scanner typically consists of an X-ray source, a detector and a rotation stage on which the object under investigation is mounted. Usually, these components can be adjusted individually and enable not only changing the acquisition geometry but also parameters such as the tube voltage, the number of projections, or the exposure time. To acquire a single X-ray image, also called *radiograph* or *projection*, X-rays are projected onto the detector, with the object in between. Part of the radiation is attenuated by the object, based on the properties of the material. The resulting image on the detector is a composite projection of the various materials encountered along the path of the X-rays. For a CT scan, the object is usually rotated by 360° , while images are taken at small angular intervals. In many micro-CT scanners the X-ray beam is cone-shaped and the detector is a two-dimensional array of pixels. The resulting data of the CT acquisition in such a *cone-beam CT scanner* therefore is a large number of *2D X-ray projections* from the full angular range. In a simplified version, the X-ray beam is fan-shaped or only the middle detector line of a cone-beam CT setup is used. This way only *1D line projections* are acquired over the full angular range. This large number of line projections can be combined into a 2D image, a so-called *sinogram*.

From measurement to reconstruction To get from these X-ray projections or sinograms to cross-sectional images computer algorithms are then applied to this data to obtain a so-called *CT reconstruction*. This is a 3D representation of the interior of the object, a gray-scale map where the pixel/CT values represent the relative densities of materials in the object. This 3D image can be viewed, for example, as a stack of 2D images, called slices, which show a virtual cut through the object [79]. The CT reconstructions from 2D sinograms can be interpreted straightforward as a cross-sectional image, a 2D representation of the intersection of the *fan-beam* and the scanned object (cf. Figure 1.2).

Physically, each of the projections represents the cumulative X-ray attenuation along straight lines through the sample, which can be mathematically expressed as a line

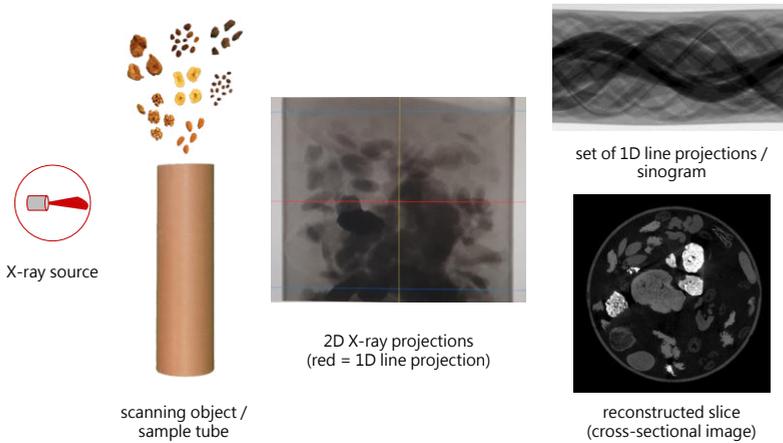


Figure 1.2: Basic terminology of (2D) Computed Tomography.

integral. Given an angle and position, the initial intensity I_0 of the X-rays at the source undergoes absorption along this line l determined by the material dependent coefficients $\mu(z)$ and reaches the detector with an intensity $I(\ell) = I_0 e^{\int_{\ell} -\mu(z) dz}$ according to Beer-Lambert's law [79]. In principle, this is a simplified version assuming that all X-rays have the same photon energy. However, micro-CT scanners commonly operate with polychromatic X-ray spectra. This means that the X-rays emitted from the X-ray source have multiple photon energies. To accurately describe the physical process with respect to the Beer-Lambert's law, an energy-dependent non-linear integral model would be necessary: $I(\ell) = \int I_0(E) e^{\int_{\ell} -\mu(E,z) dz} dE$ [66]. Usually, this energy-dependence is neglected though and an effective absorption coefficient $\mu_{eff}(z)$ is assumed for the CT reconstruction which creates image artifacts. These can be either mitigated through pre-processing the projection data or by physically filtering out parts of the X-ray spectrum as described in the section "Beam filtration".

Another common practice to pre-process the projection data, consisting of raw photon counts per detector pixel, is the so-called dark- and flat-field correction. The dark-fields represent the offset counts of the detector system and the flat-fields are the values measured when irradiating the detector without an object present between the X-ray source and the detector. These two additional measurements are usually acquired before and/or after the acquisition of the 360° projections and used to remove the dark currents of the detector and to normalize its pixel-dependent sensitivities.

1.1.2 Tomographic reconstruction as an ill-posed inverse problem

The tomographic reconstruction problem in 2D is an inverse problem which means it involves deducing an unknown quantity from observable measurements. It can be

described as an image recovery problem given some measurements after application of the Radon transform [163, 164] as $y(\ell) = \int_{\ell} x(z) dz, \ell \in \mathcal{L}$, where \mathcal{L} represents the lines in \mathbb{R}^2 from the X-ray source to each detector element, defined by the scanner geometry and rotation. To compute the CT reconstruction the continuous Radon transform must be discretized, i.e. both the continuous function and the integral operations are approximated leading to a discrete version of the problem, which can be expressed as:

$$Ax = y + \tilde{e} \quad (1.1)$$

where A represents the so-called *forward operator* which encapsulates the integral computations over these lines. Here, A is a matrix where each row corresponds to a line integral over the pixel grid of the object. In this context, x is a vector representing the pixel values of the image, y is a vector representing the measured sinogram values, and \tilde{e} accounts for the noise or error, which may arise from the measurements themselves or from the linearization of the forward operator.

Since the Radon transform specifies through the process of projection how an image is mapped to a sinogram, an intuitive attempt in inverting this process, i.e. to recover an image from its sinogram, is to perform a back-projection. But this is not the true inverse of the Radon transform, it is only its formal adjoint. Using the Fourier slice theorem one can obtain an inverse by first filtering the projection data and then back-projecting it [79]. This simple reconstruction method is called Filtered Back-Projection and was first introduced by Bracewell and Riddle [28] for astronomy and rediscovered for electron tomography by Ramachandran and Lakshminarayanan [165].

According to Hadamard's criteria the tomographic reconstruction problem in Equation 1.1 is ill-posed, mainly because of the instability of the solution. The severity of the inverse problem depends on whether the stability of the inverse problem can be established. For tomography problems, when measurements are taken well-sampled from all angles around an object, the problem is usually only mildly ill-posed. Conversely, in limited-data scenarios where measurements are restricted to a specific angular range or are badly-sampled it becomes severely ill-posed [79].

To solve the inverse problem in Equation 1.1 in a robust manner, a variational regularization approach [55, 176] can be employed. The reconstruction is defined by the following minimization problem:

$$\hat{x} = \arg \min_x \mathcal{D}(y, Ax) + \mathcal{R}(x), \quad (1.2)$$

where \mathcal{D} measures the data fidelity between the measurement and the reconstructed image (most commonly the L^2 -distance in CT) and \mathcal{R} is a regularization function that promotes images of desired properties. The fidelity term \mathcal{D} is usually chosen according to the noise distribution, and a good choice of regularizer \mathcal{R} is important for achieving accurate results. Traditionally, regularization functionals were hand-crafted to encourage the reconstruction x to have structures known to be realistic.

In practice, Equation 1.2 is then solved using iterative optimization schemes, and the quality of the reconstructions depends heavily on the choice of \mathcal{R} . Many methods have been proposed in the optimization literature to solve Equation 1.2, given particular choices of \mathcal{D} and \mathcal{R} , which are oftentimes assumed to be convex. In certain cases, these methods result in better reconstructions compared to the standard filtered backprojection (FBP) [59, 150], when appropriate functions and parameters are chosen.

While convexity of \mathcal{R} may be analytically desirable to provide efficient optimization schemes with various guarantees, in practice, reconstruction quality is significantly enhanced for non-convex regularizers. This, however, comes at a cost: finding global minima becomes impossible in general and sometimes even finding stationary points cannot be guaranteed [182].

1.2 Challenges for computed tomography

Acquiring a CT scan of an object can be a challenging task. Although micro-CT scanners offer many degrees of freedom in adjusting, for example, the geometry or the acquisition parameters such as the tube voltage, tube current, or exposure time, they usually have restrictions in the size of the objects that they can image. This can be caused, e.g. by the limited dimensions of the scanning cabinet, the rotation stage that can only host objects of a certain weight or size, or the detector's dimensions that limit the flexibility of the acquisition geometry or object size. Two principal sources for the challenges of acquiring a CT scan, therefore, are usually either the object itself or the imaging circumstances.

1.2.1 Objects

The objects scanned with CT imaging can be as diverse as the areas of applications the technique has been used in. The objects exhibit not only different atomic compositions and densities but also vary in their size. They show features of multiple scales and also the object's thickness can range from a few millimeters to tens of centimeters. This multi-material and multi-scale nature of scanning objects presents challenges for CT imaging that can lead to reduced image quality and errors in the perception or representation of information called image artifacts. These image artifacts (cf. Figure 1.3) can inhibit effective reconstruction and visualization of the acquired data but also limit the information or disrupt its interpretation. The challenges are, for example, due to areas with metals or other very dense materials, or due to the sizes of the objects and their respective areas of interest requiring partial or tiled CT scans.

1.2.2 Imaging

One of the first considerations when designing a CT scan usually is the desired scanning resolution, a concept that encompasses a range of interpretations. Its definition can vary based on whether it is defined in terms of the optics of the system — referring

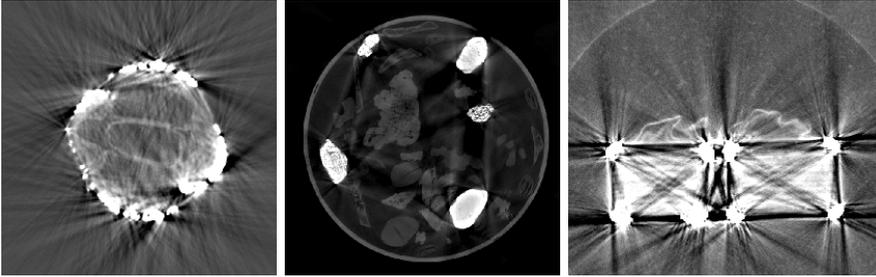


Figure 1.3: Image artifacts in CT imaging of various multi-material and multi-scale objects. From left to right: a velvet knife holder with gold and silver embroidery causing metal artifacts; a cardboard tube filled with dried fruits and nuts, coffee powder, and lava stones causing beam hardening artifacts, a purse adorned with a drawstring of braided gold thread with silver core causing metal artifacts.

to the minimum size of features that can be distinguished — or in terms of pixel size, which relates to the inherent resolution of the imaging detector. In this work, we will consider resolution as the physical length within the scanned object that is resolved by one detector pixel. For micro-CT scanning, this length is usually in the range of micrometers. For example, for a scan with $60\ \mu\text{m}$ resolution, features of the object that have a size of one pixel are $60\ \mu\text{m}$ big. But if the size of the object is too big to realize the desired resolution, one experiences geometric constraints. They result in so-called limited-view artifacts which are caused by a practical inability to measure all the projection data that theoretically would be needed to obtain a complete dataset. These artifacts can also appear in region-of-interest (ROI) scans where one is only interested in a certain region of the object and not the whole object is fitted on the detector. Since for certain projections parts of the objects are visible on the detector but for other projections they are not, this leads to artifacts within the reconstructions.

Artifacts can also occur for cone-beam CT scans when the object is moved very close to the source. This results in a high beam angle at the upper and lower edge of the field-of-view and the reconstructions show streaks at the boundaries of materials near the edge of the detector. This problem can also occur when one conducts a tiled scan for a big object at a larger source-to-object distance (SOD). For such a tiled scan the detector is placed at multiple positions behind the object at the same source-to-detector-distance (SDD) for the acquisition of the projections during a full rotation. This is done to virtually create a bigger detector to cover the whole object and is possible in both horizontal and vertical direction. Usually, the multiple projections are stitched together and the reconstructions are computed with the virtually extended detector. For these reconstructions of a circular-orbit cone-beam geometry, an FBP-type reconstruction technique called FDK (Feldkamp-Davis-Kress) [59] can be used. In the case of vertical tiling, one can also move both source and detector and stitch together the reconstructed volumes. For horizontal tiling, however, moving both source and detector necessitates an iterative solver for the reconstruction (cf. Figure 1.4).

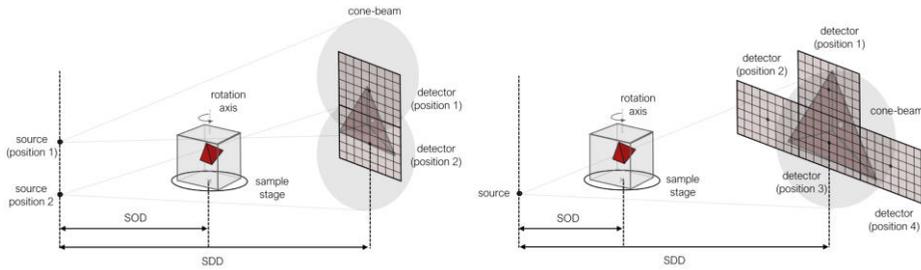


Figure 1.4: Schematic representation of a CT setup illustrating two options for tiling with different magnification factors $\frac{SDD}{SOD}$; left: vertical tiling with multiple source and detector positions; right: tiling with fixed source position and multiple detector positions to virtually create a bigger detector (adopted from [23]).

Generally speaking, acquiring CT images is a finite measurement. In an ideal setting, we would have continuous information and optimal signal from all angles, but this is not always the case. In fact, we are limited by the nature of the measurement process and introduce artifacts to our X-ray projections and reconstructions by that. These occur due to the ill-posedness of the inverse problem which is getting stronger with badly-sampled measurement data as described in section 1.1.2. Subsequently, we introduce different causes for a suboptimal measurement process:

Limited-angle

Another cause for limited-view artifacts are limited-angle scans. They become relevant, for example, when the object cannot be fully rotated within the scanner cabinet which could be the case for e.g. big paintings. Because the rotation of the object is limited, projections from different angles are not available and undersampling artifacts occur within the reconstruction. When limiting the angular range for fan-beam CT scans to e.g. 180° or less, the missing information causes visible image artifacts such as streaking, elongation, ghost tail, and missing boundaries. In the following overview (cf. Figure 1.5), we present the artifacts occurring in an FBP reconstruction of projection data with limited angles of 180° , 150° , 120° , 90° , or 60° and show the corresponding data fully sampled FBP reconstruction. FBP reconstruction of projection data from a limited angle 60° are already dominated by artifacts and a further sub-sampling is omitted. Generally, the difficulty of a limited-angle reconstruction increases with decreasing available angular range.

Sparse-angle

Next to limited-angle artifacts there are also so-called under-sampling errors [89]. When taking a CT scan of an object, a sampling process takes place in both the number of pixels the object covers on the detector and the number of angles from which projections are taken. The number of projections and their angular distribution

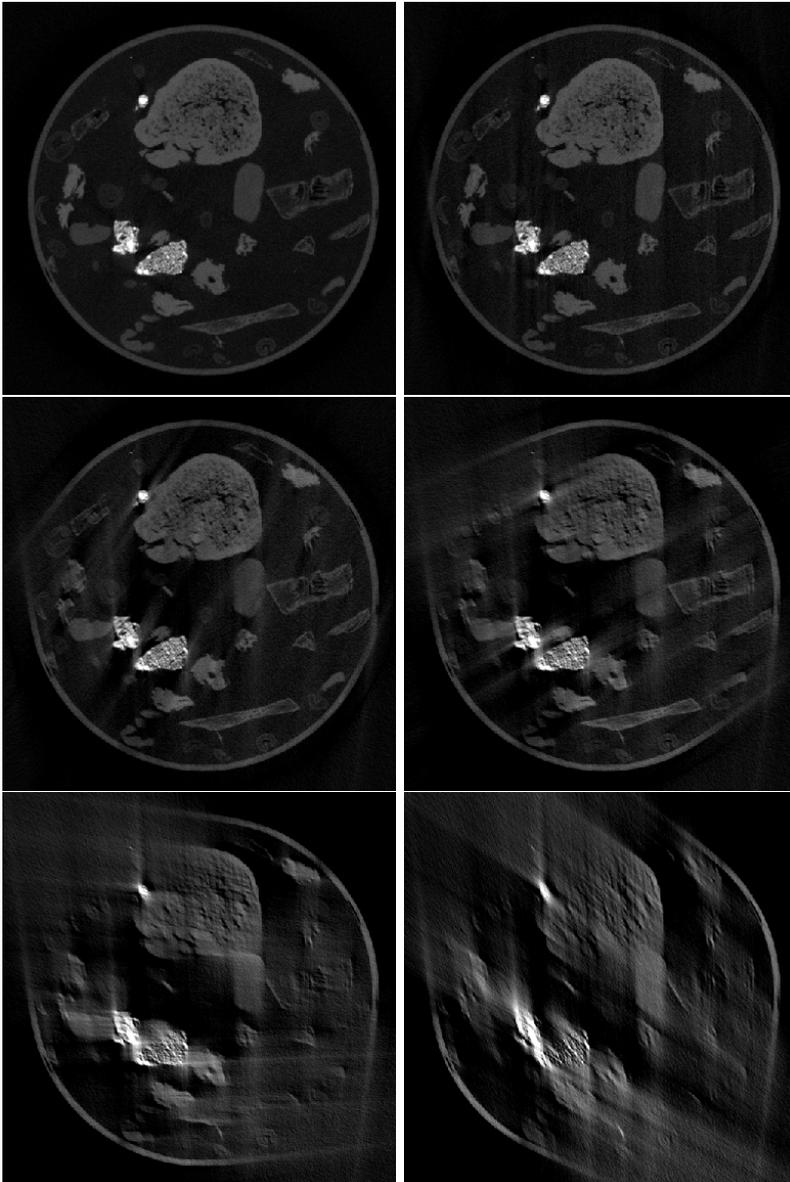


Figure 1.5: From top to bottom, left to right: FBP reconstructions of a limited-angle CT acquisition based on projection data with a limited angle of 180° , 150° , 120° , 90° , or 60° and a fully sampled CT acquisition.

plays an important role, particularly for calculating the reconstruction. Ideally, measurements are available continuously around 360 degrees and earlier research showed though that with the Nyquist-Shannon theorem one is able to determine the minimal number of projection angles to yield a unique reconstruction. Applied to the field of computed tomography, the theorem yields a relationship between the number of scanned points S in one projection line and the number of necessary projections P : $P \geq \frac{\pi}{2}S$ [94]. These projections have to be equally distributed around 360 degrees.

For the experimental setup of the 2DeteCT dataset presented in chapter 4 a minimal number of $\sim 3,000$ projections is required for sufficient sampling. Noticeable differences, however, only occur for under-sampling by factors of five or more. We tested specifically under-samplings based on 720, 360, 120, 90, and 60 projections and show the severity of the artifacts in Figure 1.6. The difficulty of sparse-angle reconstruction increases with decreasing number of available projections.

Low-dose

Another origin for artifacts caused by finite measurement is related to the average photon fluence and the signal of the attenuated photons in the detector. The average fluence, defined as the product of flux and exposure time, is inherently constrained by factors such as radiation safety limits and practical time restrictions. These limitations becomes particularly significant when considering the quantum nature of photons, which results in a discrete number of detected photons that exhibit random fluctuations proportional to the average fluence. Consequently, in scenarios with low photon counts, the signal-to-noise ratio (SNR) tends to be low, making it challenging to discern meaningful signals from background noise. Additionally, when digitizing these signals, considerations regarding dynamic range become crucial. For a good measurement all detector pixels should measure some photons and should not measure close to full capacity. For the measurement of the photons, we are limited by the number of counts the computer system is able to store. Usually, this is given by the bit-length which limits the measurement to e.g. the range of 0 and 65,535 for a 16-bit unsigned integer. For a signal that would cause higher counts the detector saturates in the projection images and this yields artifacts in the reconstructions. But also at the small end of the count range one can run into problems. If the measured signal from the attenuated photons yields counts that are in the range of the offset counts (“dark currents”) of the detector, the pre-processing of the data can involve negative values or divisions by zero which generate artifacts as well. Furthermore, a low detector signal creates noisy projection images and subsequently noisy reconstructions.

In medical imaging, a low-dose setting is typically chosen to achieve images of adequate quality for clinical purposes while minimizing radiation dose to the patient. The “tube current”-“exposure time” products usually range from 50 to 400 mAs in clinical practice. For the 2DeteCT dataset presented in chapter 4 the high-dose acquisition has a “tube current”-“exposure time” product of 18.0 mAs, while the low-dose acquisition has a product of 0.6 mAs. The low-dose setting presented in Figure 1.7 uses a 1/30 tube current compared to the high-dose setting next to it which has been optimized

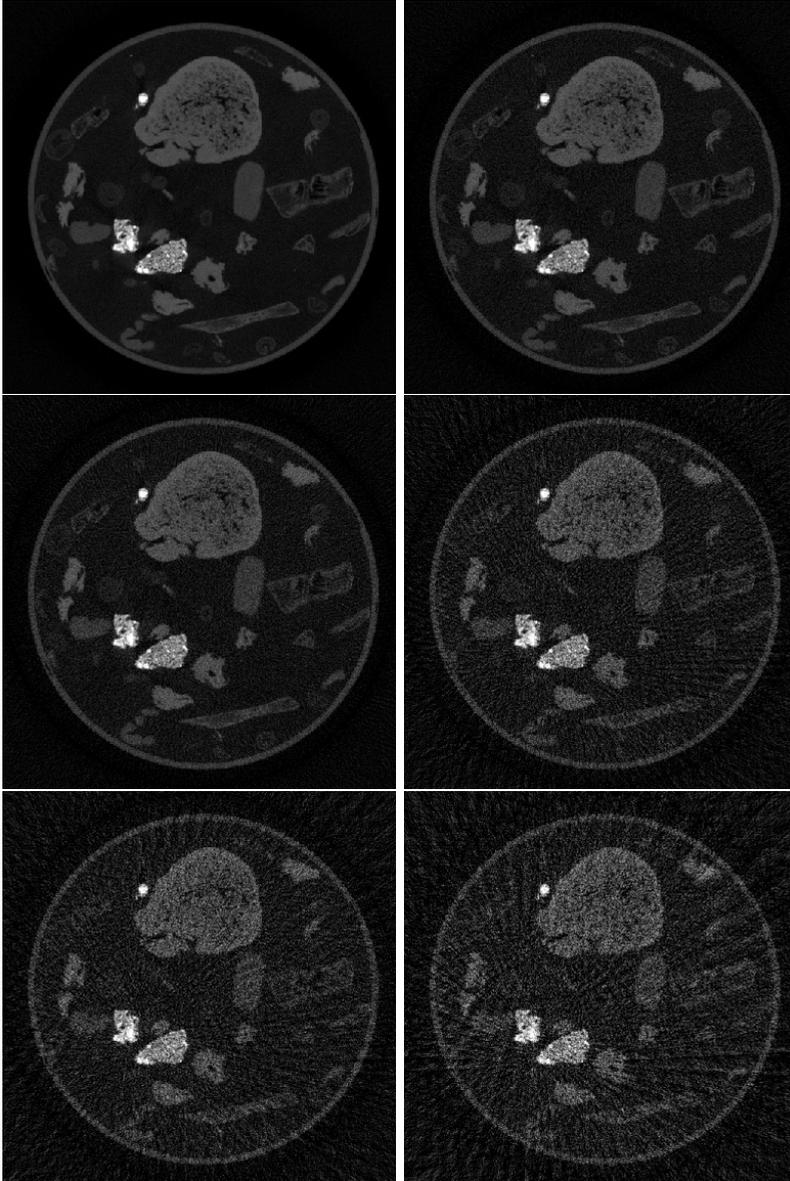


Figure 1.6: From top to bottom, left to right: FBP reconstructions of a sparse-angle CT acquisition based on 720, 360, 180, 120, 90, and 60 projections and a fully sampled CT acquisition.

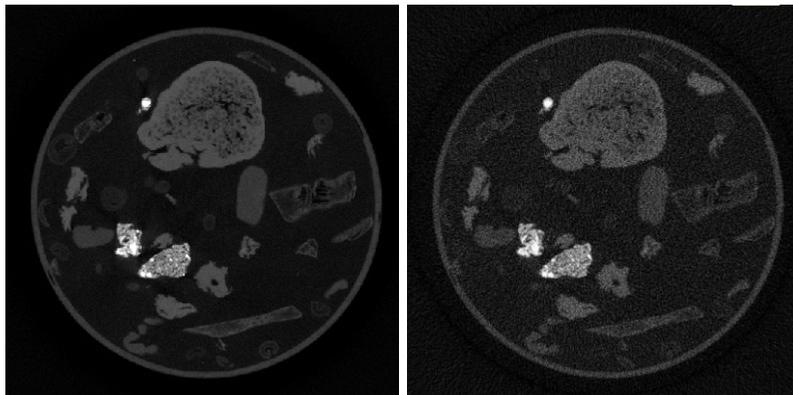


Figure 1.7: From left to right: FBP reconstruction of a high-dose and low-dose CT acquisition.

for the best image quality. The figure shows what artifacts are introduced by such a CT acquisition.

Beam-hardening and photon starvation

The standard computational models used in CT imaging rely on the assumption that the X-ray beam is monochromatic, assigning a single attenuation coefficient to each material. But this is not the case for common micro-CT scanners. In fact, the beam emitted from their X-ray sources usually provides a spectrum of photon energies. This spectrum shows a continuous distribution of photon energies composed of bremsstrahlung photons and discrete lines of characteristic radiation, which are dependent on the target material used within the X-ray tube. The low-energy photons within the beam are preferentially absorbed by the scanning object, which causes the beam spectrum to shift to higher energies as it passes through more and more material. The beam becomes progressively "harder" (increased average photon energy) and this changes the effective absorption coefficient μ_{eff} [158]. Therefore, the polychromatic nature of the X-ray beam leads to an error in the attenuation coefficients and the linear relationship between them and the material thickness. This is called beam-hardening and causes a variety of artifacts in images [35]. Two easily visible signs of these beam-hardening artifacts are cupping and streaking. When objects in the scanning sample appear brighter at the edges than at the center this is called a cupping artifact. Furthermore, beam-hardening can also create streaking artifacts which are displayed as dark and light streaks around very attenuating structures. A more severe effect related to the high attenuation of particular objects in the CT scanner is the so-called photon starvation which refers to the fact that the density of a material is so high that for an X-ray beam in a given direction the number of photons that penetrate the material is so low that it falls within the range of dark current fluctuations. With such a low measurement signal the reconstruction algorithm fails and creates star-like streaks originating in the highly dense structure. These image artifacts due

to beam-hardening can be minimized by so-called beam filtration, the placement of materials into the X-ray beam. With this the X-ray spectrum can be modified and its low-energy portion reduced, which is commonly referred to as "pre-hardening the beam". It shifts the mean photon energy towards higher energies and narrows the standard beam spectrum for CT imaging. In Figure 1.8, we show the beam spectra and reconstructions of two corresponding CT acquisitions: one uses an unfiltered beam spectrum emitted from an X-ray source operated with 60kV and 60W whereas the other uses a filtered beam spectrum (Thoraeus filter: Sn 0.1mm, Cu 0.2mm, Al 0.5mm) from an X-ray source operated with 90kV and 90W.

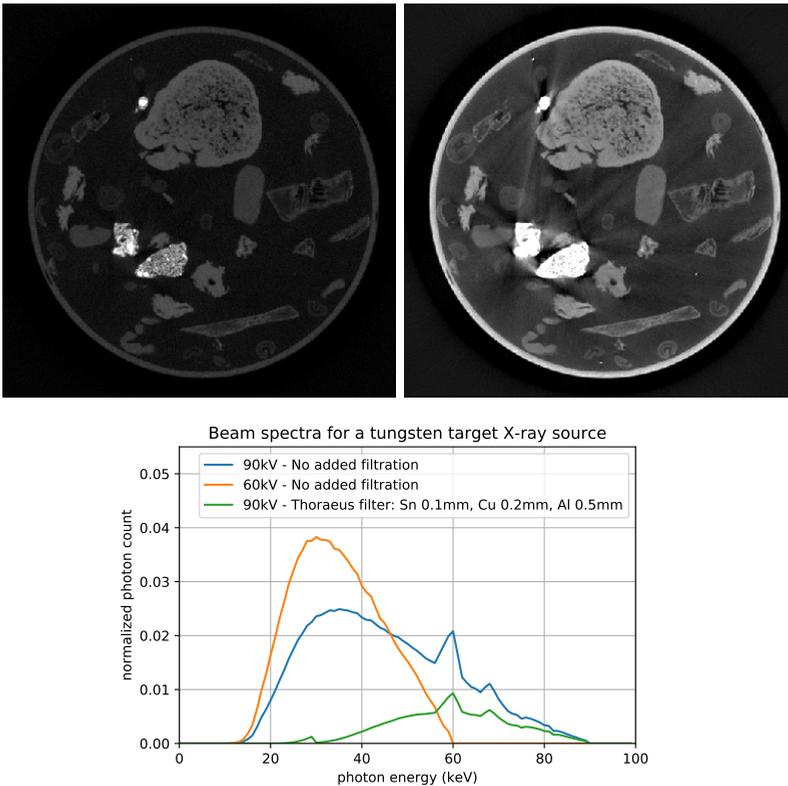


Figure 1.8: From left to right: FBP reconstruction of a filtered and an unfiltered CT acquisition exhibiting beam-hardening artifacts and the corresponding beam spectra (green and orange) simulated with the TASMIP software [22].

1.3 Machine learning

Artificial intelligence (AI) is the branch of computer science that focuses on creating systems capable of performing tasks that typically require human intelligence, such as

understanding language, recognizing patterns, and making decisions. Machine learning (ML), a sub-field of AI, focuses on the development of algorithms that allow computers to learn from and make predictions based on data [56]. In addition to computer science, which develops these new algorithms, a wide range of natural sciences has also exhibited a strong research interest in utilizing ML to enhance their respective fields.

Although ML has been already defined in 1959 by Arthur Samuel [175], a more modern but still general definition was given by Tom Mitchell: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks T , as measured by P , improves with experience E " [6]. In other words, the computer program increasingly improves its performance on carrying out a task that it has not been explicitly programmed for. Overall, machine learning is viewed as a disruptive technology that with the support of big data and accelerated computation is able to bring forth tremendous algorithmic innovations [215]. Within this broader field of ML lies deep learning (DL), a specialized area that focuses on techniques based on neural networks (NNs) and representation learning. The adjective deep emphasizes the multiple levels of representation involved to transform input data into more complex and abstract representations [122].

Following the initial definition for ML, methods can be grouped depending on what experience E , performance P , or task T they use. Examples of tasks carried out by a ML algorithm include classification, regression, transcription, machine translation, anomaly detection, denoising, and many more. The performance measures P often relate to the accuracy of a model or the error rate the ML model produces. The experience E can be understood to relate to the way a ML algorithm is trained. The two common general categories for this are unsupervised and supervised learning which refer to how the ML algorithm experiences a dataset [72]. For unsupervised learning (UL), it experiences a dataset with many features as a whole and tries to derive structure, clustering and relationships among possible variables in the data without feedback on the prediction result. In supervised learning (SL) with a given dataset encompassing input data X and correct output/target data Y the algorithm focuses on learning a mapping between those two. A comprehensive overview of different types of machine learning is given in Table 1.1.

In this thesis, we mainly focus on supervised learning methods. For this kind of algorithm datasets need to be composed of so-called training examples, with indices $i = 1, \dots, m$ (total number of training examples) which are given by a pair $(x^{(i)}, y^{(i)})$ of input and target variable. Formally, we want to learn a function $h : \mathcal{X} \mapsto \mathcal{Y}$, called hypothesis, mapping from input manifold \mathcal{X} to output manifold \mathcal{Y} . As a performance measure for the accuracy of our hypothesis function a so-called cost or loss function is introduced, which measures the "difference" between the predicted value \hat{y} and the actual target value y with some chosen metric.

Table 1.1: Characteristics of different types of machine learning (adopted from [36]).

Type of ML	Type of Data Provided	Mechanism
Supervised learning (SL)	Labelled data	The algorithm uses the pairs of input and target data to infer the relation that maps between the data.
Unsupervised learning (UL)	Unlabelled data	The algorithm searches for rules - if-then-associations - to discover patterns within the data.
Semi-supervised learning (SSL)	Small amount labelled, but mostly unlabelled data	The algorithm uses the small amount of labelled data to develop an initial model that is iteratively applied to the greater amount of unlabelled data.
Reinforcement learning (RL)	No labelled data necessary, but numerical performance scores	The algorithm tries to maximize a reward function or reinforcement signal to achieve a predefined goal under a given set of rules.

This metric varies depending on the application but a common choice is the mean squared error:

$$L(\hat{y}, y) = \frac{1}{m} \sum_{i=1}^m \left(\hat{y}^{(i)} - y^{(i)} \right)^2 = \frac{1}{m} \sum_{i=1}^m \left(h(x^{(i)}) - y^{(i)} \right)^2 \quad (1.3)$$

Those datasets are then commonly split into three separate subsets of different sizes: A training set used to optimize the ML algorithm on with the aim of the smallest possible loss. A validation set used to fine-tune and compare different ML algorithms. A testing set used to see how well the chosen algorithm generalizes to unseen examples. Depending on the available size of the dataset common splits are 60/20/20%, 80/10/10%, or 98/1/1%. It is of vital importance that all these subsets come from the same distribution to avoid systemic flaws. Particularly, the validation and testing set should be close to the training set distribution but also to future unseen examples. This also touches on the widely observed sim-to-real gap [152, 205]. Especially, early computer vision algorithms often used small-scale datasets under laboratory conditions for their development and lacked the ability to generalize to the real world. Having good data became therefore of increasing importance and led to the emergence of the field of data science, a new research area at the intersection of statistics, computer science, and machine learning devoted to maximizing the value from vast collections of information [184]. A popular adage of this field is the concept "Garbage In, Garbage Out", which underlines the idea that the quality of the output is inevitably connected to the quality of the input.

The acquisition and release of datasets developed into a more and more important research area and both publishers and conferences devote entire journals and submission tracks to this. This pays tribute to the fact that the releases of large-scale, open-source datasets such as MNIST [123], CIFAR [119] and ImageNet [46] have been key enablers in the field of computer vision that helped the research community to develop standardized benchmarks and continuously advance the state-of-the-art.

1.4 Machine learning for computed tomography

Over the past decades methods used for solving ill-posed inverse problems such as reconstructing a CT image from measurement data focused on incorporating regularization, priors and/or learned knowledge to improve the reconstruction quality. In the beginning, mathematical models for medical image reconstructions were mostly handcrafted and designed by expert knowledge or hypotheses on the reconstructed images. After that, hybrid approaches combining handcrafted and data-driven modeling emerged where parts of the models are learned from observed data while they still mostly rely on human design. More recently, advances in machine learning and developments regarding available data and computational resources enabled models which are mostly learned and rely only minimally on human design [225].

These models, however, require large-scale, versatile, experimental datasets to be trained on, and the field of X-ray CT lacked such datasets for developing machine learning methods. Especially two-dimensional, reconstructed CT slices would play an important role in advancing method development since the corresponding learning and reconstruction tasks require less computational resources compared to their three-dimensional counterparts. Generally, a large-scale benchmarking dataset for 2D computed tomography may be a first step to enabling similar breakthroughs in machine learning based CT image reconstruction.

Deep learning methods for CT image reconstruction have been categorized in different ways. A book about deep learning for biomedical image reconstruction by Ye et al. [219] divides models, for example, into three groups: Pre-processing, post-processing, and raw-to-image, where the CT images are directly reconstructed from the raw measurement data. On the one hand, the pre-processing methods face challenges, as analytical reconstruction algorithms are sensitive to errors introduced during the pre-processing of measurement data, which can lead to the emergence of new artifacts in the reconstructed images. On the other hand, also post-processing methods encounter problems, e.g. when the output of classical reconstruction algorithms of suboptimal data has too severe artifacts to be mitigated, or when they create structures/features that are not consistent with the data, so-called “hallucinations”. Lastly, the raw-to-image group encompasses a wide variety of different methods that differ in their methodological approach and their inherent challenges.

To further clarify the landscape of deep learning approaches for image reconstruction, a domain-based categorization was proposed by Ravishankar et al. [166], which classify methods into image-domain learning, hybrid-domain learning, AUTOMAP [230] (which learns a direct mapping between measurement and image domains), and sensor-domain learning. While this classification offers insight into where learned algorithms fit within the process of transforming raw data into reconstructed images (cf. Figure 1.9a), it mixes different methodological models.

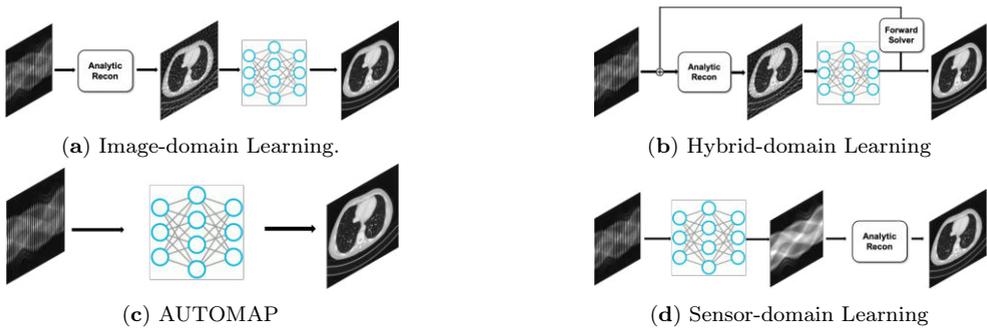


Figure 1.9: Domain-based categorization of deep learning methods for CT image reconstruction. Adopted from Ravishankar et al. [166]

Another recent survey [11] focused on this more methodological component in their general categorisation of supervised learning methods for solving inverse problems using data-driven models. Following mainly this last classifications, we consider the following four method categories in chapter 6: post-processing networks, learned / unrolled iterative methods, learned regularizer methods, plug-and-play methods.

While any strict categorization may overlook or misrepresent certain methodologies from the literature, such as self-supervised learning or unsupervised approaches, this framework effectively encompasses the majority of techniques found in the (weakly) supervised learning literature on data-driven CT reconstructions.

Due to the substantial mathematical and computational differences among the models outside of these four method categories, we have opted to exclude them from the comparison analysis presented in chapter 6. To establish a foundation for benchmarking, we prioritize established techniques from these four categories that can serve as reliable baselines, omitting some newer techniques based on, e.g. transformers and generative models, as explained in chapter 7.

1.5 Research questions

In this thesis, we present our contributions to advance learned algorithms for 2D X-ray computed tomography by answering the research questions outlined in this section. Below, we indicate how the individual chapters in this thesis deal with the research questions and how they influenced and built upon each other.

The starting point for our investigations was analyzing the problems in the development of ML methods for CT. As indicated in the general introduction to machine learning having good data is of vital importance for developing, training, and testing ML algorithms for CT reconstruction. However, in the imaging community algorithm development and data collection often do not go hand-in-hand, but are carried out by separate institutions with limited interaction. Algorithms are developed on simulated data and their generalization to real-world tomographic applications are rarely examined. Therefore, realistic experimental data is highly important but scarce and its acquisition immensely difficult.

Designing a data collection suitable for training ML methods for CT reconstruction requires a thorough understanding how objects with different characteristics have to be imaged differently to obtain high-quality CT scans. Especially, for a supervised learning setup it is necessary to acquire pairs of data which on the one hand exhibit image artifacts due to their challenging nature and on the other hand are acquired in an optimized way to be artifact-free.

Research question 1:

How can we improve the acquisition of CT scans through beam-filtration for challenging multi-material scanning samples such as cultural heritage objects?

In chapter 3 we show how untailed acquisitions of CT scans of multi-material objects can lead to reduced image quality and heavy visual errors called image artifacts, which can influence the perception or representation of information. We demonstrate how a tailored acquisition can reduce these artifacts and lead to a higher information gain (see Figure 1.10). We discuss how the X-ray beam properties and the beam-object interaction influence CT image formation and how to use filters to manipulate the emitted X-ray beam to improve image quality for multi-material objects. We showcase that this can be achieved with limited resources in a low-cost DIY fashion with thin sheets of metal as filters, 3D-printed filter frames and a filter holder. Secondly, we give a qualitative analysis of the influence of the CT acquisition parameters illustrated with two case study objects from the textile collection of the Rijksmuseum, Amsterdam, The Netherlands. With this we provide insights and intuitions on tailoring CT scans to cultural heritage objects. Thirdly, we extract a general concept of steps for museum professionals to design an object-tailored CT scan for individual cases.

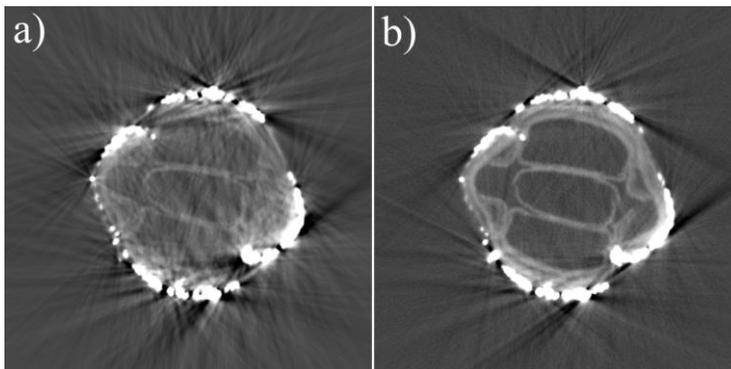


Figure 1.10: Reconstruction slice of a standard acquisition (50 kV, 700 μ A, no filter) a) and an object-tailored acquisition (70 kV, 1000 μ A, Thoraeus filter (Sn 0.25mm, Cu 0.5mm, Al 0.5mm) b) for the case study 2 "Purple velvet knife holder" from chapter 3

To advance learned algorithms for 2D X-ray computed tomography and to overcome their lack of generalization to real-world tomographic applications, a suitable CT dataset has to be acquired in an experimental manner instead of being simulated. Furthermore, this dataset needs to be designed in a versatile way such that its samples resemble the natural variations of images from, e.g. clinical applications, and offer the imaging community suitable data for different image reconstruction tasks.

Research question 2:

How can we acquire a versatile experimental CT dataset for the development of machine learning methods for CT image reconstruction?

In chapter 4 we design a versatile, open 2D fan-beam CT dataset suitable for developing machine learning techniques for a range of image reconstruction tasks such as supervised or unsupervised denoising, sparse-angle scanning, beam-hardening reduction, super-resolution, region-of-interest tomography, or segmentation. We describe in detail the steps involved in acquiring an unprecedented X-ray data collection by making extensive use of a highly flexible, programmable and custom-built X-ray CT scanner. Based on the insights from chapter 3 we choose suitable scan parameters for the acquisition such as beam filtration, X-ray tube voltage and current, detector exposure time, binning and averaging, the number of projection angles as well as source, object and detector positions. Furthermore, we develop a sophisticated, semi-automatic scan procedure that allows to automatize the collection of 50 slices during an 8.5h scan. In 111 scanning sessions (each with a different sample mix) and a total scanning time of more than 850 hours over a duration of almost five months a diverse mix of samples with high natural variability in shape and density was scanned slice-by-slice (5000 slices in total) with high angular and spatial resolution and three different beam characteristics: A high-fidelity, a low-dose, and a beam-hardening-inflicted mode. In addition, 750 out-of-distribution slices were scanned with sample and beam variations to accommodate robustness and segmentation tasks. We provide raw projection data, reference reconstructions and segmentations based on an open-source data processing pipeline (see Figure 1.11).

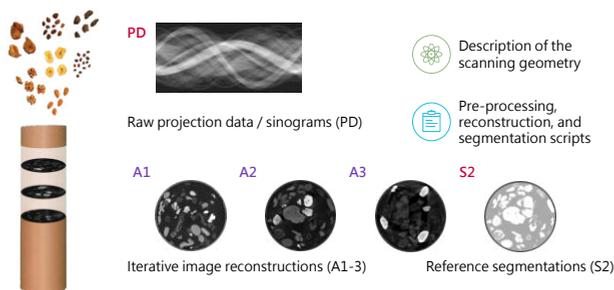


Figure 1.11: Overview of the scope of the 2DeteCT dataset.

Especially in the field of low-dose CT acquisitions and the denoising of CT data many of the existing research studies are based on simulated CT data. With the rise of ML algorithms used for denoising also their training is based on simulated low-dose CT data and their generalization to real-world experimental low-dose CT data is to be discussed.

Research question 3:

How does the performance of machine learning methods for denoising differ when trained on clean and simulated noisy CT data compared to clean and experimental noisy CT data?

In chapter 5 we utilize the large 2D computed tomography dataset for machine learning presented in chapter 4. We carry out for the first time a comprehensive study on the differences between the observed performances of algorithms trained on simulated noisy data and on real-world experimental noisy data. The study compares the performance of two common CNN architectures that are trained and evaluated on both simulated and experimental noisy data (see Figure 1.12). The results show that sinogram denoising performs better with training on simulated noisy data when evaluated in the sinogram domain on both experimental and simulated noisy test data. However, this performance does not carry over to the reconstruction domain where training on experimental noisy data shows a higher performance in denoising experimental noisy data. Training the algorithms in an end-to-end fashion from sinogram to reconstruction significantly improves model performance, emphasizing the importance of matching raw measurement data to high-quality CT reconstructions. The study furthermore suggests the need for more sophisticated noise simulation approaches to bridge the gap between simulated and real-world data in CT image denoising applications. It gives insights into the challenges and opportunities in leveraging simulated data for machine learning in computational imaging.

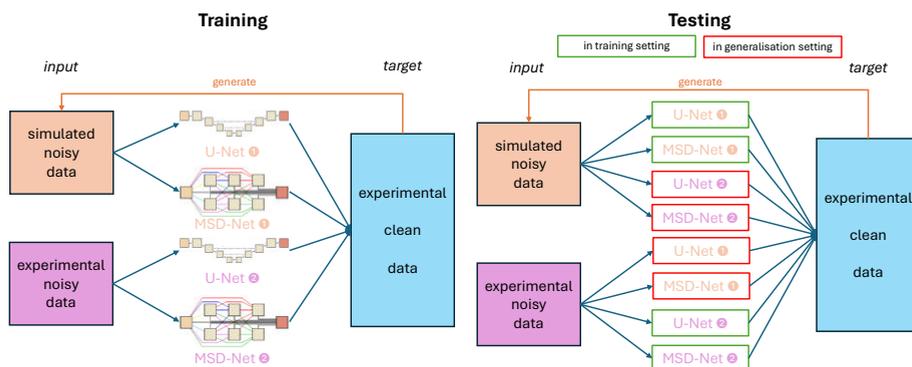


Figure 1.12: Training and testing scenarios for learned denoising networks (U-/MSD-Net illustrations adopted from [160]).

Throughout the last decades many different ML methods from various method categories have been developed to perform the most common CT image reconstruction tasks. Oftentimes, these algorithms were developed using datasets that are not openly available, utilize a lot of simulation, or are of various sizes. Furthermore, the corresponding research studies use different pre-processing pipelines and model implementations, which causes a general lack of comparability between the different ML methods. With the 2DeteCT dataset from chapter 4, this issue and the following question can be addressed.

Research question 4:

How do machine learning methods from different method categories perform on common CT image reconstruction tasks in comparison?

In chapter 6 we address how the lack of large-scale, open-access datasets has hindered the comparison of data-driven state-of-the-art methods in CT image reconstruction. We use the 2DeteCT dataset from chapter 4 for benchmarking machine learning based CT image reconstruction algorithms. We categorize these methods into post-processing networks, learned/unrolled iterative methods, learned regularizer methods, and plug-and-play methods, and provide a pipeline for easy implementation and evaluation. Using key performance metrics, including SSIM and PSNR, our benchmarking results showcase the effectiveness of various algorithms on tasks such as full data reconstruction, limited-angle reconstruction, sparse-angle reconstruction, low-dose reconstruction, and beam-hardening corrected reconstruction (see Figure 1.13).

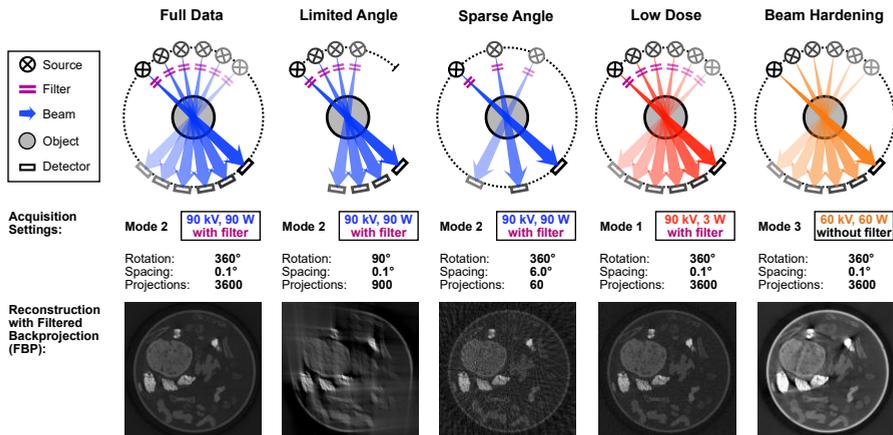
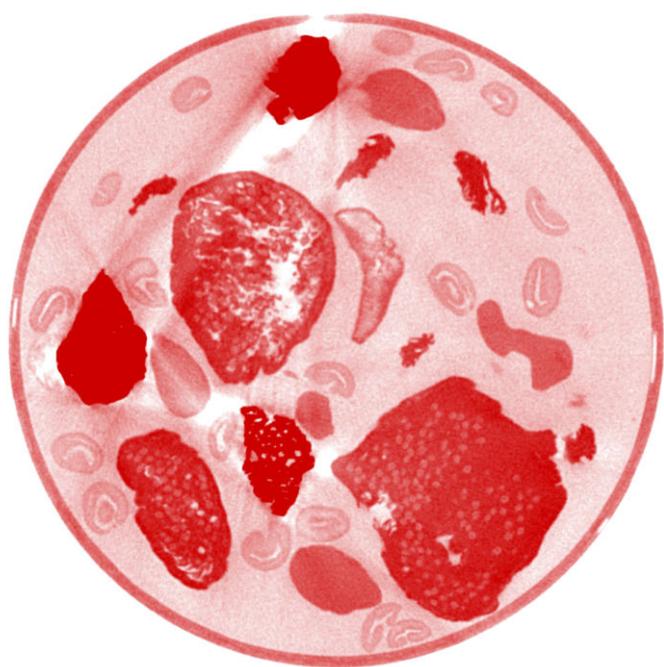


Figure 1.13: CT Image Reconstruction Tasks.



2

Tailoring CT Acquisitions in the FleX-ray Lab

One important concept of this thesis is tailoring CT acquisitions to the scanning samples. The relevance and impact of this concept is illustrated in more detail in chapter 3 for two cultural heritage case study objects. In this section, we present the specific laboratory CT scanner used for the experiments conducted in this thesis. The FleX-ray scanner is a custom-built, highly flexible X-ray CT scanner, developed by TESCAN XRE NV [216] located in the FleX-ray Laboratory at the Centrum Wiskunde & Informatica (CWI) in Amsterdam, The Netherlands. Generally, this laboratory is used to conduct proof-of-concept studies in the field of mathematics and computer science [43].

2.1 Functionalities of the FleX-ray scanner

The scanner has three main components: i) a cone-beam microfocus X-ray point source emitting polychromatic X-rays between 20 keV and 90 keV with a tube current between 10 μA and 1000 μA ; ii) a rotation stage, on which samples of different sizes (up to 40 cm \times 40 cm \times 40 cm) can be mounted; and iii) a CMOS (complementary metal-oxide semiconductor) flat panel detector with a CsI(Tl) scintillator (Dexella 1512NDT, [131]) with 1536 \times 1944 pixels, 74.8 μm^2 each, onto which the X-rays are projected. The offset counts (“dark currents”) and the maximal readout of this detector are given by $\sim 1,000$ and 65,535 counts respectively. Translation stages enable all three components to move independently from one another. An image of the scanning set-up is shown in Figure 2.1.

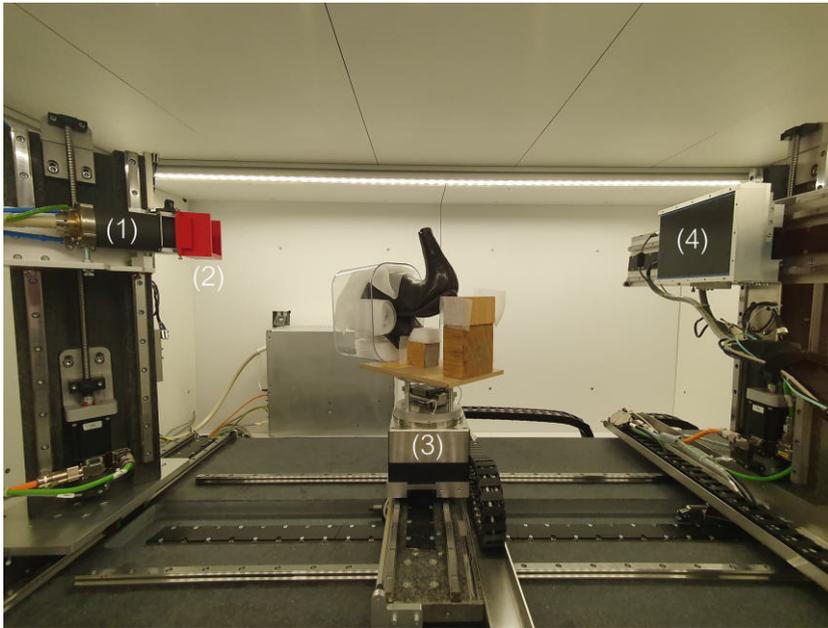


Figure 2.1: FleX-ray scanner setup with X-ray source (1), attached beam-filter holder (2), rotation stage with mounted object (3), and flat panel detector (4).

The main features of the FleX-ray scanner: X-ray voltage and current, motor movements for source, detector, and sample positioning, as well as detector features can be programmed for the full duration of a scan. These technical specifications, also listed in Table 2.1, enable better suited data acquisition and were instrumental to realize the CT scans in this thesis.

2.2 Extensions to the FleX-ray scanner

Although the FleX-ray scanner by itself already was a highly flexible micro-CT scanner, its operation required at times less reproducible and provisional research practices. To standardize and unify, for example, the positioning on the sample stage two extensions to the FleX-ray laboratory were created: sample stage disks and sample stage cylinders. Furthermore, the functionalities of the scanner were extended to use beam filtration for the CT acquisition by creating a beam filter holder and a set of exchangeable and combinable beam filter frames.

2.2.1 Sample stage disks

The rotation stage within the FleX-ray scanner is made of aluminium and has a sample stage disk made of aluminum as well placed on top of a carved out ring-like area. The

Table 2.1: Technical specifications of the FleX-ray scanner, HW: Hardware, SW: Software

Parameter	Value
Tube voltage	20 - 90 kV
Tube power	2 - 50/90 W ("microfocus" / "high power")
Tube current	10 - 1500 μ A (depending on the focus mode, standard 100)
Exposure time	11.7/14.3/38.5 - 5000 ms (dependent on hardware binning)
Number of frames to average	free choice in full integers
Hardware binning	1×1 (HW1SW1); 2×2 (HW2SW1); 4×4 (HW4SW1) HWX $X \times X$ pixels are grouped to SWY $Y \times Y$ pixels on the software side
Detector rows	1536 ($ROI = 1520; 8 - 1527$)
Detector columns	1944 ($ROI = 1912; 32 - 1943$)
Physical detector size	145.9 mm \times 114.9 mm
Source Object Distance (SOD)	ca. 15 mm - ca. 1009 mm
Source Detector Distance (SDD)	ca. 139 mm - ca. 1098 mm
Magnification Factor $\alpha = \frac{SDD}{SOD}$	1.087 - 73.133
Number of projections	free choice in full integers, usually multiples of 360 +1 are used: e.g. 1, 361, 721, 1081, 1441, 1801, 2161, 2521, 2881, 3241, 3601

diameter of this ring on which the aluminium sample disk is placed was measured to be 110mm and we produced several additional aluminum disks. We produced 10 disks with a plain top surface for basic scans that require no exact positioning or mounting within the rotation stage diameter. They can also be taped to wooden boards, sponge materials or anything else that might be necessary for mounting the respective sample for the CT scan. Furthermore, 10 disks with a special design were created. They feature a sketch of angle markers and circles of different radii for exact positioning. At designated locations there are also holes to fit in LEGO™ studs or poles for exact positioning (cf. Figure 2.2).

2.2.2 Sample stage cylinders

When acquiring a CT scan, we also want to limit the number of scattered photons in our projections. One source of this unwanted scattering is the rotation stage within the FleX-ray scanner which is made out of aluminium. This scattering can be circumvented, though, by elevating the sample a few centimeters above the rotation stage. For this we used the diameter of the ring on which usually the aluminium sample disk is placed and 3D-printed cylinders with this diameter with varying heights and designs on the top surface (cf. Figure 2.3).

The motivation for varying heights comes from the fact that, depending on the region of interest that we want to use for both the object and the detector, we need a higher sample stage cylinder to prevent the metal rotation stage to be within the field-of-view or too closely adjacent. We created three different designs which are available in heights of 20, 50, and 100mm. The first design features a plain top surface for basic

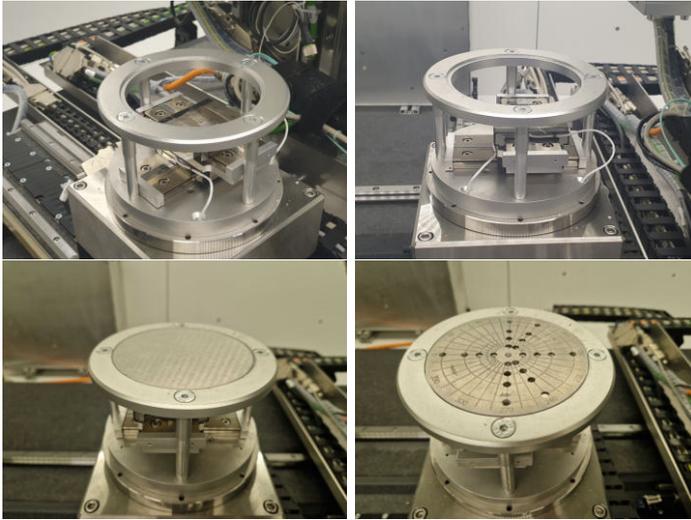


Figure 2.2: The rotation stage within the FleX-ray scanner with the two new sample stage disk designs (left: plain, right: with markers and holes).

scans that require no exact positioning or mounting within the rotation stage diameter. The second design encompasses a carved out volume of area $6.38\text{cm} \times 6.38\text{cm} \times 0.32\text{cm}$ (LxWxH) to fit in a LEGO™ base-plate of 8×8 studs. The third design re-models the design of the second sample stage disk and contains the same sketch of angle markers and circles of different radii for exact positioning including the holes at designated locations. These holes always reach half-way through the different heights of the sample stage cylinders.

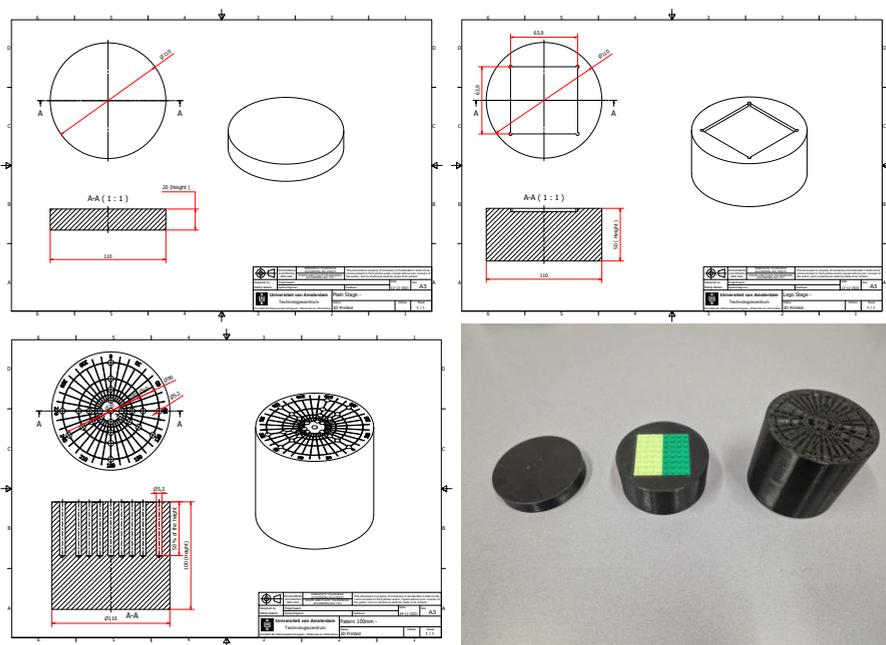


Figure 2.3: Technical sketches and 3D-printed realization of the three different heights and design choices for the sample stage cylinders.



Figure 2.4: The set of 3D-printed filter frames with thin sheets of metal of different thicknesses and the 3D-printed beam filter holder.

2.2.3 Beam filtration

As described in section 1.2, the interplay between a broad beam spectrum and the energy-dependent absorption of the sample's materials can create a range of image artifacts. With so-called beam filtration, the placement of materials into the X-ray beam, we can modify the X-ray spectrum and reduce its low-energy portion which is commonly referred to as "pre-hardening the beam". This shifts the mean photon energy towards higher energies and narrows the standard beam spectrum for CT imaging. It can minimize image artifacts due to beam-hardening and avoid saturation of the detector.

This filtration of the beam typically happens at or near the X-ray tube window in the direct path of the X-ray beam. The filters applied to the X-ray beam are commonly thin sheets of metal, such as aluminum (Al), copper (Cu), iron (Fe), tin (Sn), and tungsten (W), but studies have shown that combination filters of different materials and thicknesses are even more effective in reducing image artifacts [77, 87, 96]. These compound filters combine the different attenuation characteristics of various materials and result in more effective filtering by reducing intensity of low-energy photons over a wider band.

To enable a stable, exact, and reproducible beam filtration for the Flex-ray scanner, we designed a tool that can be mounted on the X-ray tube and allows for the placement of various filters into the beamline. This tool should hold the filters in a fixed position and allow also the use of multiple filters of different thicknesses and materials.

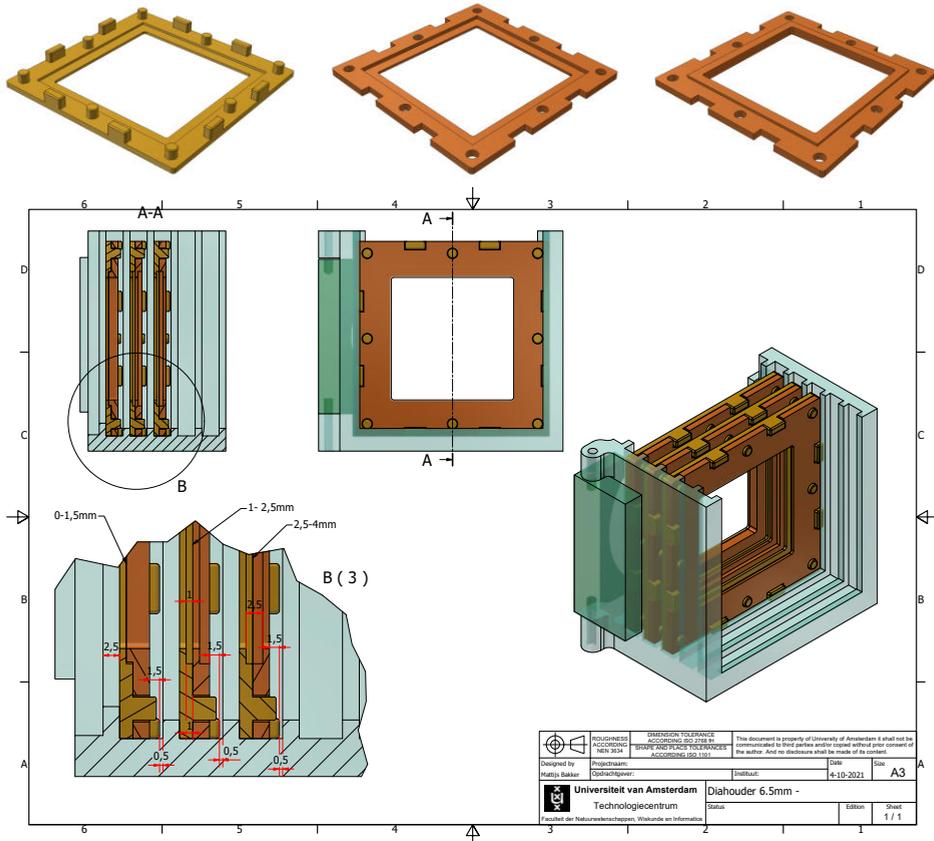


Figure 2.5: Technical sketches of filter holder and filter frames.

We designed a 3D-printed beam filter holder similarly to a slide projector with five slits. The corresponding filter frames were designed so that they are always tightly positioned within these slits and can host filter materials with a thickness ranging between 0.1 and 4.0 mm. The technical sketches and 3D-printed realizations can be found in Figures 2.5, 2.4, and 2.6.

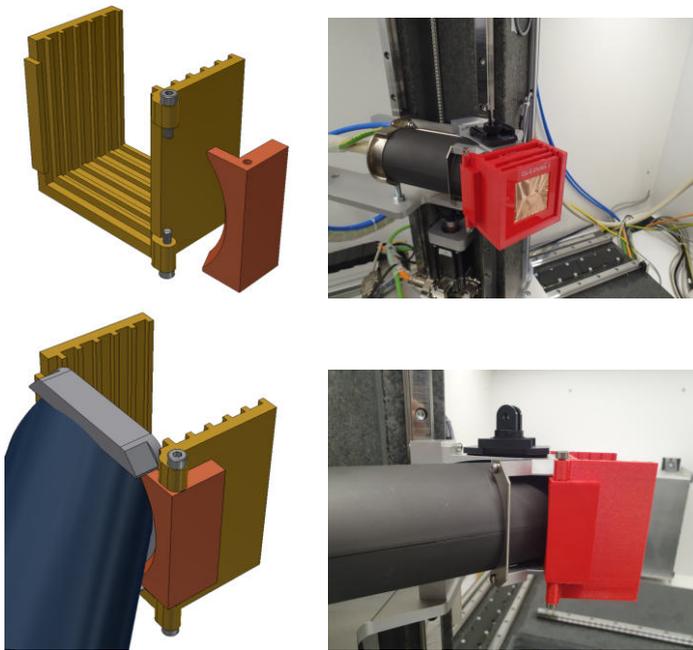


Figure 2.6: Setup of beam filter holder on X-ray tube.

2.3 Guideline for choosing suitable acquisition parameters

Setting up a CT acquisition for a chosen scanning sample can be an extensive and laborious task. Over the course of this PhD thesis, we developed a guideline for the FleX-ray scanner to help researchers with a step-by-step process to determine suitable acquisition parameters. Of course, some of the steps are more specific to our CT system and need to be adapted for a different system, but they might give the reader an initial idea and basic insight in how different choices influence the outcome of the CT acquisition.

We found four common basic objectives when scanning a sample: low noise, high angular and spatial resolution, high contrast, and fast scans. For a good signal-to-noise ratio it is important to consider how different acquisition parameters influence the noise level within the CT images such as source current, source voltage, exposure time, number of projections, and number of averaged images [172]. The number of projections necessary to have sufficient sampling and high angular resolution can be calculated considering the Nyquist-Shannon theorem with $P \geq \frac{\pi}{2} S$ [94]. Depending on the experiment, there might be limitations in the choice of one or multiple of the acquisition parameters, but it might be possible to compensate for these choices with adjusting other values.

As a prerequisite for applying the subsequent guideline, the minimal and maximal available values of the following three CT acquisition parameters should be known: tube voltage, tube current, exposure time. Furthermore, the dynamic range of the detector in use should be determined, namely the lower bound, given by the offset counts (“dark currents”), and the maximal readout of the detector (at FleX-ray Laboratory: $\sim 1,000$ and $65,535$ counts). This is done to identify the necessary photon flux for the CT acquisition. The aim should be to have counts of at least a two- to three-fold of the dark current at any location in all radiographs and never to saturate the detector.

1. Determine a suitable SOD and SDD to achieve the desired resolution in the CT reconstruction of the object under investigation. Use the intercept theorem: $\frac{\text{voxel size}}{SOD} = \frac{\text{detector pixel size}}{SDD}$ to calculate the corresponding scaling factor $\frac{SOD}{SDD}$ and to determine whether the object under investigation can fit onto the detector with this geometry: $\text{geometry diameter} = \text{detector width} \times \frac{SOD}{SDD}$ as well as $\text{geometry height} = \text{detector width} \times \frac{SOD}{SDD}$. If the object does not fit, it might be necessary to acquire a tiled scan where multiple reconstructions are stitched together vertically or the detector needs to be placed at multiple positions and a bigger virtual detector is stacked together.
2. Empirically determine the rotational position with the lowest detector signal which corresponds to the largest attenuation in the line of sight. This usually corresponds to the thickest part of the object and/or where the object is composed of denser materials.

3. Investigate the minimal photon energy needed to penetrate the object in this rotational position by slowly increasing the tube voltage in small steps of e.g. 5kV while observing the minimal detector signal. The measured signal should be a two- to three-fold of the dark current (at FleX-ray Laboratory: 2000-3000 counts).
4. Choose adequate filtration that can filter out the low energy part of the spectrum, i.e. photons that have a lower energy than the energy found in the step above. The mean photon energy within the beam spectrum should be above the minimal penetration photon energy and also the minimal photon energy of the spectrum should be close to that. This may require to use a peak tube voltage (kVp) that is considerably higher than the tube voltage corresponding to the minimal photon energy. To find a suitable beam filtration for the object under investigation the corresponding beam spectra can be simulated with the TASMIP software [22]. For lower minimal photon energies thin sheets of Copper (e.g. 0.01 - 0.50mm) can be used whereas higher energies may require thin sheets of Tin (e.g. 0.10 - 0.50mm) or compound filters such as the Thoraeus filter of (e.g. Sn 0.25mm, Cu 0.5mm, Al 0.5mm) that can effectively filter out photons carrying an energy of 1.5 keV to 70 keV [93].
5. Exclude the K-edges of any metals within the object by adjusting the tube voltage accordingly. A list of absorption edges can be found, for example, at the following website [57].
6. Calculate the maximal feasible acquisition time per projection image of the detector from the total available scanning time to achieve the highest signal-to-nois-ratio (SNR) that satisfies the given time and saturation constraints. Also, consider taking multiple frames for averaging each projection, if the total available scanning time allows for that. The three important values for this are: First, the minimal feasible exposure time available determined by the electronics of the CT scanner / detector $t_{exp,min}$; second, the maximal exposure time restricted by the saturation of the detector $t_{exp,max-det}$; third, the maximal exposure time restricted by the total available scanning time $t_{exp,max-time}$. The second upper bound can be simply measured when moving the scanning object out of the field of view and radiating the detector with the previously found tube voltage, chosen beam filtration and the maximal tube current and increasing the exposure time from the minimal value until the detector saturates. The other upper bound can be calculated by the following formula:

$$t_{exp,max-time} = \frac{t_{total-image-series}}{(\#PPI + \#proj) * \#avg} \quad (2.1)$$

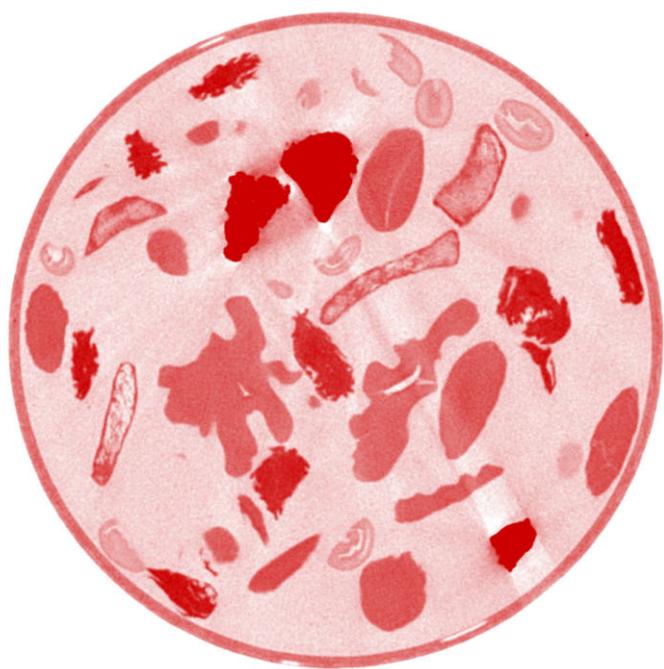
where the variables are given as:

- $t_{exp,max-time}$ - maximal exposure time restricted by the scanning time
- $t_{total-image-series}$ - available scanning time for the total image series

2.3. GUIDELINE FOR CHOOSING SUITABLE ACQUISITION PARAMETERS

- $\#_{PPI}$ - number of pre-processing images which is the sum of all dark and flat field images (before+after)
 - $\#_{proj}$ - number of projections
 - $\#_{avg}$ - number of averages
7. Find a balance for the trade-off between the reduction of beam-hardening image artifacts, improved contrast, and reduced noise by adjusting the tube voltage and tube current within the given boundaries.

In this section, we have outlined the essential steps for setting up a CT scan and choosing suitable acquisition parameters such as SOD, SDD, tube voltage and current, beam filtration, and exposure time tailored to specific samples. By following these guidelines, researchers can effectively optimize their CT scanning outcomes while balancing factors like noise, resolution, contrast, and scan speed. Although some of the contents of this chapter are specific to the FleX-ray Lab, the extensions can be implemented in other laboratories as well and the structured approach above can foster a deeper comprehension of how varying parameters influence the CT scanning process.



3

Beam filtration for object-tailored X-ray CT of multi-material cultural heritage objects

Computed Tomography (CT) is a non-invasive X-ray imaging technique. The scanning setup consists of an X-ray source, a detector and a rotation stage on which the object of interest is mounted. A single X-ray image, also called radiograph or projection, is created by shooting X-rays at the detector, with the object in between. Part of the radiation is attenuated by the object, based on material properties such as density, thickness, and atomic composition. The resulting image on the detector is an overlay of the materials in the direction of the X-rays. For a CT scan, the object is rotated, while 2D X-ray images are taken at small angular intervals. Afterwards, computer algorithms are used to create a tomographic reconstruction of the data. This is a 3D representation of the interior of the object, a grey scale map where the pixel/CT values represent the relative densities of materials in the object. This 3D image can for example be viewed as a stack of 2D images, called slices, which show a virtual cut through the object [79, 83]. Since the development of X-ray CT, its applications broadened from the medical sector to various other fields, e.g. manufacturing industry, food industry, and material sciences [196].

This chapter is based on:

M. B. Kiss, F. G. Bossema, P. J. van Laar, S. Meijer, F. Lucka, T. van Leeuwen, and K. J. Batenburg. “Beam filtration for object-tailored X-ray CT of multi-material cultural heritage objects”. *Heritage Science* 11.1 (2023), p. 130.

Computed Tomography in Cultural Heritage Computed tomography has also proved to be a powerful non-invasive tool to analyse cultural heritage objects [34]. It has been successfully applied to a variety of such objects over the past decades: for example, wooden statues for dendrochronology research [49] and large wooden objects [24, 167], Egyptian funerary masks [192], antique glass [61], musical instruments [50, 185], ancient basketry [9], anthropological studies [135, 143] and historical paintings [146, 193]. In cases where it is not possible to see the interior of an object using visual inspection, CT enables museum professionals to obtain 3D information about the interior of the object. In particular, CT scanning gives a much more detailed overview than radiography because it is three-dimensional, allowing to isolate specific features and parts by virtually cutting through the object in any location and direction. These insights can help with the conservation or restoration of the objects, as well as provide contextual information on the object's history or making process [209].

Challenges specific to CT in Cultural Heritage Cultural heritage objects exist in a wide variety and have characteristics which present challenges for CT scanning: multi-scale internal features, a diversity of sizes and shapes, and multi-material objects [23]. The last challenge is even greater when the multiple different materials of the object have varying densities, especially when one of the materials is a metal [86]. A multi-material object might be for example made out of fabric, leather, metal thread and different kinds of beads as well as designed in a multi-layered way. Hidden intermediate layers that could give insight into the object's history and making process might be difficult to analyse. Additionally, the actual dimensions and characteristics of these layers are usually not visible from the outside. Investigating these characteristics using X-ray radiographs (2D X-ray projections) will only give a limited amount of information, whereas CT-scanning will be more informative. Acquiring CT scans of these objects with standard lab/museum CT scanners in an untailed way yields images with heavy visual errors (see Figure 3.1) called image artifacts¹.

Difference to medical CT challenges Although both the medical and the cultural heritage sector need an acceptable image quality for interpretation and face similar challenges in the CT acquisition, they are fundamentally different [34, 70]. In medical CT scans the scanned subjects (humans) are similar with respect to their anatomy and the material composition. The human body is generally more homogeneous with respect to its density distribution compared to cultural heritage objects. Although metals may be present in the form of prostheses, these are confined to specific regions whereas metals in cultural heritage can also occur scattered throughout the object. Furthermore, the energy regime of the photons is different and improvements to medical CT have been extensively researched, e.g. the tuning of tube voltage and current, using filters and software-based artifact correction algorithms [156]. Lastly, the medical sector lacks an extensive adaptability of acquisition parameters since commercial CT scanners are used that have a limited number of adjustable settings.

¹The word artifact is used in this work only in reference to visual errors in CT images, not in reference to an object shaped by human workmanship (sometimes referred to as 'artefact').

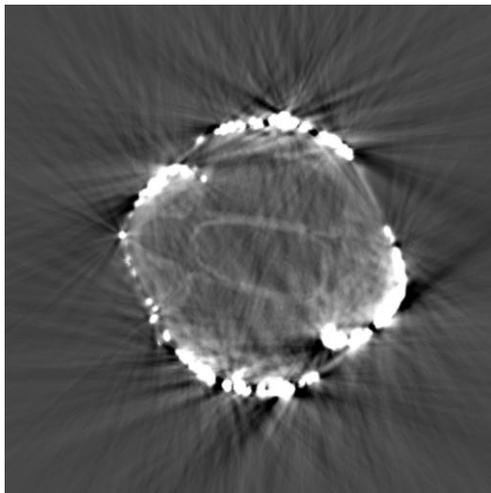


Figure 3.1: Heavy visual errors (image artifacts) exhibited by an untailored CT scan of a multi-material cultural heritage object.

In contrast, objects scanned in the cultural heritage sector are all different in size and material composition and there is limited knowledge on what is in the object and what it is made of. As CT scans are typically carried out with laboratory-setups, the acquisition parameters can be adjusted and researchers can interactively investigate the effect of these parameter choices [231]. We will show that the aforementioned challenges can be overcome by tailoring the acquisition parameters to the object under study. This enables not only cleaner and more expressive 2D X-ray projections (radiographs) but also more informative reconstructions. Therefore, it is advisable to analyse the objects to be scanned for taking possible causes of image artifacts into account (cf. "Background" section).

Contributions The contributions of this work are as follows: Firstly, we briefly introduce several of the key factors that influence computed tomography image formation in the context of cultural heritage objects to make this work accessible to a broad audience including museum professionals as well as X-ray imaging specialists. We address the underlying physics and discuss the use of filters to manipulate the emitted X-ray beam to improve image quality for these multi-material objects. This can be achieved with limited resources in a low-cost DIY fashion as described in the "Hard- and software" section. Secondly, we give a qualitative analysis of the influence of the CT acquisition parameters illustrated with two case study objects from the textile collection of the Rijksmuseum, Amsterdam, The Netherlands (cf. "The case studies" section), which were scanned at the FleX-ray Lab of the Centrum Wiskunde & Informatica, Amsterdam, The Netherlands. With this we provide insights and intuitions on how to design a CT scan and on choosing suitable acquisition parameters, illustrated by these case studies. Thirdly, we extract a general concept of steps for

museum professionals to design an object-tailored CT scan for individual cases.

As this work is positioned at the interface between scientific imaging and cultural heritage, it will be necessary to sometimes make certain obvious statements, to ensure a common ground of understanding between professionals with different backgrounds and technical knowledge.

3.1 Background

In this section we will address how the multi-material nature of cultural heritage objects presents a challenge for CT imaging that can lead to reduced image quality and errors in the perception or representation of information called image artifacts. These image artifacts can hinder effective reconstruction and visualization of the acquired data but also limit the information or complicate its interpretation. The three main factors that contribute to lower image quality are (i) the polychromatic X-ray spectrum and metal artifacts; (ii) the contrast between materials; (iii) noise and detector saturation.

3.1.1 Polychromatic X-ray spectrum and metal artifacts

The X-ray beam used for CT imaging commonly has a polychromatic nature, meaning that the beam emitted from the X-ray source emits a spectrum of photon energies. Exceptions for these polychromatic beams can be found in specialized synchrotron facilities which generate a monochromatic beam, with photons of one specific energy. Furthermore, the attenuation of X-ray photons traversing a material is energy dependent. The attenuation coefficient of materials commonly present in cultural heritage objects decreases as the photon energy increases. [5]

The standard computational models used in CT imaging rely on the assumption that the X-ray beam is monochromatic, assigning a single attenuation coefficient to each material. However, the polychromatic nature of the X-ray beam leads to an error in the attenuation coefficients and the linear relationship between them and the material thickness due to the energy-dependence of the attenuation. This is called beam hardening and causes a variety of image artifacts. [35, 158]. Two signs of these beam hardening image artifacts are "cupping" and "streaking". When regions in the scanned sample appear brighter (higher density) at the edges than at the center (decreased brightness) this is called a cupping artifact. Streaking artifacts show as dark and light streaks around structures with a high density. Visual examples of both image artifacts can be found in Figure 3.2.

An extreme example of such dense materials are metals. The attenuation coefficients of metals are much higher than those for lighter materials such as ceramics, wood, or textiles and can lead to incomplete attenuation profiles. In particular, around the K-edge of metal elements - a sudden increase in X-ray absorption when the energy of the X-rays is just above the binding energy of the innermost electron shell - the absorption properties change substantially and cause errors in the reconstructed CT

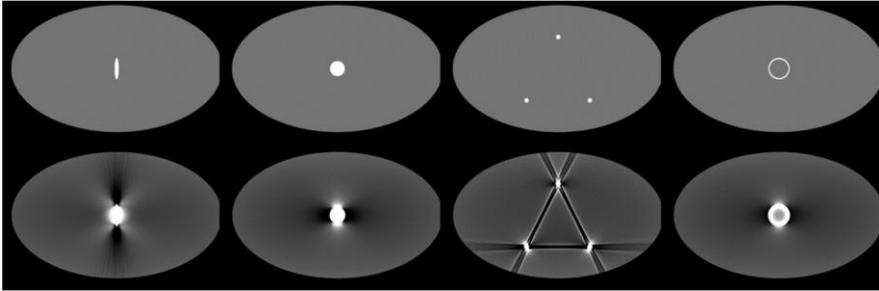


Figure 3.2: Simulated scans without (top row) and with (bottom row) beam hardening, showing that dark streaks occur along the lines of greatest attenuation, and bright streaks occur in other directions. Also note the subtle "shadows" beneath the surface, which is caused by beam hardening and is called cupping artifact. Reprinted with permission from [21].

image. Metal objects in the field of view can therefore also lead to severe streaking image artifacts. This is due to a combination of beam hardening and photon starvation but can also originate in extreme contrasts within the radiographs. Photon starvation is an effect where the attenuation of a material is so high that almost no signal for a given X-ray beam is detected behind the object in some of the orientations. Without a measured signal, star-like streaks originating in the metal structure appear in the CT reconstruction (cf. Figure 3.2).

3.1.2 Contrast between materials

The concept of contrast is generally hard to measure and quantify. A common aim when conducting CT scans is to have a high difference in absorption between the features of interest inside the object and their surroundings, including the scanning medium (air, sand, water, etc.) [174]. A high contrast-to-noise ratio (CNR) enables visual interpretability and might be very case-dependent. For many materials common in cultural heritage the absorption curve has a steeper slope for lower photon energies. Therefore, lower photon energies in the X-ray beam enable more contrast between the different materials [142]. This property can be utilized to achieve higher contrast between the different materials by creating a beam spectrum with photons whose energy is low but sufficient to penetrate the object.

3.1.3 Noise and detector saturation

In a CT scan, all angular projections are acquired using a fixed exposure time and source current, which jointly determine the beam intensity that reaches the detector after interacting with the object. When acquiring the projection images, it must be ensured that for each individual projection the detector measures a sufficient signal across the X-ray image, while at the same time not saturating the detector at any particular location.

For a given exposure time, the maximum source intensity is limited by the values measured at the detector without an object present between the X-ray source and the detector (the so-called "flat-field"), which should not trigger detector saturation. The minimum suitable values for intensity and exposure time can be found by rotating the object to find the line of sight that has the largest attenuation, resulting in the lowest signal at the detector across the full CT scan. As a rule of thumb, this value should be at least a factor 2-3 greater than the signal measured with the source turned off (the so-called "dark-field") or around 5% of the total count range [188].

It is noteworthy that the photons in the low-energy range of the spectrum contribute to detector saturation, while at the same time may hardly contribute to the actual signal measured while scanning the object. Filtering out these photons at the source can therefore be useful in improving penetration of the object while avoiding detector saturation. Within these limits, increasing the beam intensity will lead to a higher signal-to-noise ratio (SNR) of the projections, resulting in reduced noise in the reconstructed CT image.

3.2 Methods

In this section we describe the concepts and resources used for the investigation of the two case study objects.

3.2.1 Concept of beam filtration

As described in the background section, the interplay between a broad beam spectrum and the energy-dependent absorption of the sample's materials can create a range of image artifacts. In this section, we discuss the concept of beam filtration which is the placement of materials into the X-ray beam that modify the spectrum of the beam. We describe how beam filtration can reduce the low-energy portion of the X-ray spectrum, commonly referred to as "pre-hardening the beam", and allows mainly higher energy X-rays to pass through the scanned object. This narrows the standard beam spectrum for CT imaging as well as shifts its mean photon energy towards higher energies, minimizing image artifacts due to beam hardening and avoiding saturation of the detector.

The filtration of the beam typically happens at or near the X-ray tube window in the direct path of the X-ray beam. The filters applied to the X-ray beam are commonly thin sheets of metal, such as aluminum (Al), copper (Cu), iron (Fe), tin (Sn), and tungsten (W), but studies have shown that combination filters of different materials and thicknesses are even more effective in reducing the image artifacts [77, 87, 96]. These compound filters combine the different attenuation characteristics of various materials and result in more effective filtering by reducing intensity of low-energy photons over a wider band. One well-established compound filter is the so-called Thoraeus filter, which consists of tin, copper and aluminum at varying thicknesses. This compound filter can effectively filter out photons carrying an energy of 1.5 keV

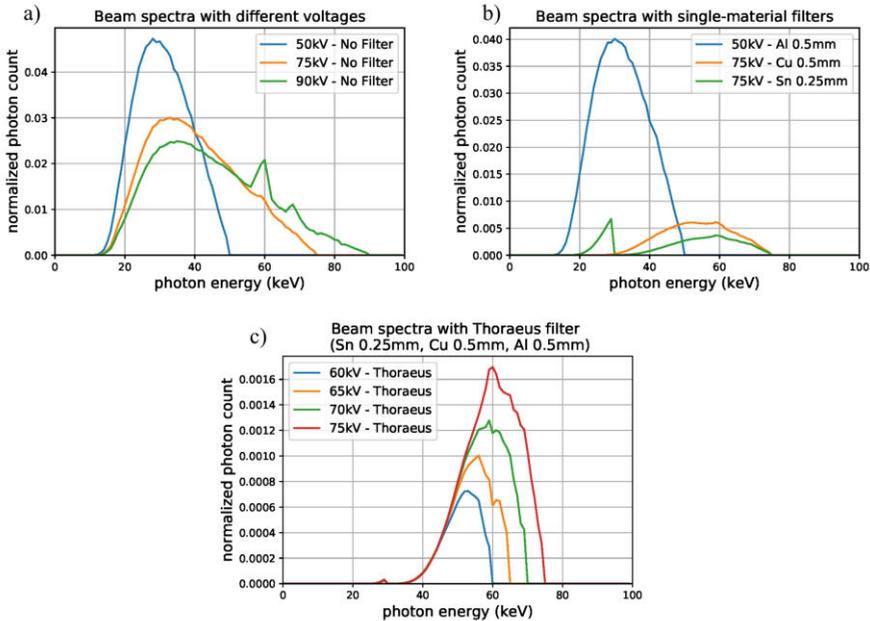


Figure 3.3: Beam spectra of a tungsten target X-ray source with an X-ray exit window made of $300\ \mu\text{m}$ Beryllium simulated by TASMIP software [22]. a) at different tube voltages 50kV, 75kV, 90kV (A continuous distribution of energies composed of Bremsstrahlung photons and discrete lines of characteristic radiation [for tungsten 58 and 69keV] can be seen. The maximal photon energy of the beam spectrum is determined by the tube voltage. This maximal energy though is hardly represented by photons, while most of the photons are at roughly one third of this peak energy); b) at tube voltages of 50/75kV filtered with either aluminum, copper or tin of varying thicknesses (see legend); c) at varying tube voltages 60, 65, 70, 75kV filtered with a Thoraeus compound filter of Sn 0.25mm, Cu 0.5mm, Al 0.5mm

to 70 keV [93]. In practice, very low energy photons are filtered out already by the X-ray tube exit window itself. In Figure 3.3 the beam spectra of a tungsten target X-ray source operated at different tube voltages and with different filters are shown.

3.2.2 Hard- and software

This explorative study is a collaborative work of conservators, curators and researchers from the Rijksmuseum, Amsterdam, and CT imaging scientists from the Centrum Wiskunde & Informatica (CWI) Amsterdam. All scans were conducted in the custom-built FleX-ray Lab at CWI, which contains a highly flexible X-ray CT scanner, developed by TESCAN XRE NV [216]. Generally, this laboratory is used to conduct proof-of-concept studies in the field of mathematics and computer science [43]. The scanner has three main components: i) a cone-beam microfocus X-ray point source emitting polychromatic X-rays between 20 keV and 90 keV with a tube current between $10\ \mu\text{A}$ and $1000\ \mu\text{A}$; ii) a rotation stage, on which samples of different sizes (up to

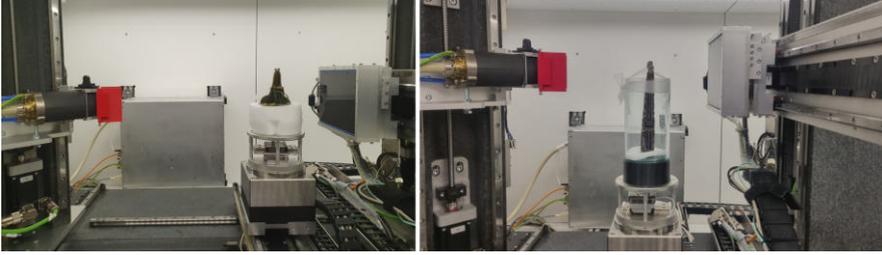


Figure 3.4: Scanning setup for the case study objects in the FleX-ray scanner

40 cm \times 40 cm \times 40 cm) can be mounted; and iii) a CMOS (complementary metal-oxide semiconductor) flat panel detector with a CsI(Tl) scintillator (Dexella 1512NDT, [131]) with 1536 \times 1944 pixels, 74.8 μm^2 each, onto which the X-rays are projected. The off-set counts (“dark currents”) and the maximal readout of this detector are given by $\sim 1,000$ and 65,000 counts respectively. Translation stages enable all three components to move independently from one another. An image of the scanning set-up is shown in Figure 3.4.

We point out that the maximum tube voltage available for the FleX-ray scanner is lower than other commercial systems that range between 40 to 225 kV. The purpose of this work however is to investigate the two case study objects and to give general guidance on designing object-tailored CT scans. Particular results will be dependent on the individual scanners used. Specifically, the available maximum tube voltage will have a strong impact on the ability to penetrate the objects. For some of the experiments in this work the maximum tube voltage of the FleX-ray scanner limits the shaping of the beam spectrum tailored to the given materials in the objects but because of their small thickness, it was possible to reduce the metal artefacts sufficiently to reveal the layers and other materials in the object.

For the purpose of beam filtration we designed an extension to the current FleX-ray Lab setup to have a stable, exact and reproducible way of filtering the X-ray beam. The key idea is to attach a tool on the X-ray tube that enables to place various filters into the beam line. These filters should be held in a fixed position and also the use of multiple filters of different materials and thicknesses should be possible. Therefore, we modelled a beam filter holder for 3D printing, which can be mounted on the tube, and corresponding filter frames in which the filter materials can be clamped in. Similarly to a slide projector these filter frames can then be placed into one of five slits of the beam filter holder. We designed the filter frames so that they are always tightly positioned within the slits and can host filter materials with a thickness ranging between 0.1 and 4.0 mm. The technical drawings can be found in Figure 3.5 and the construction files can be made available upon request.

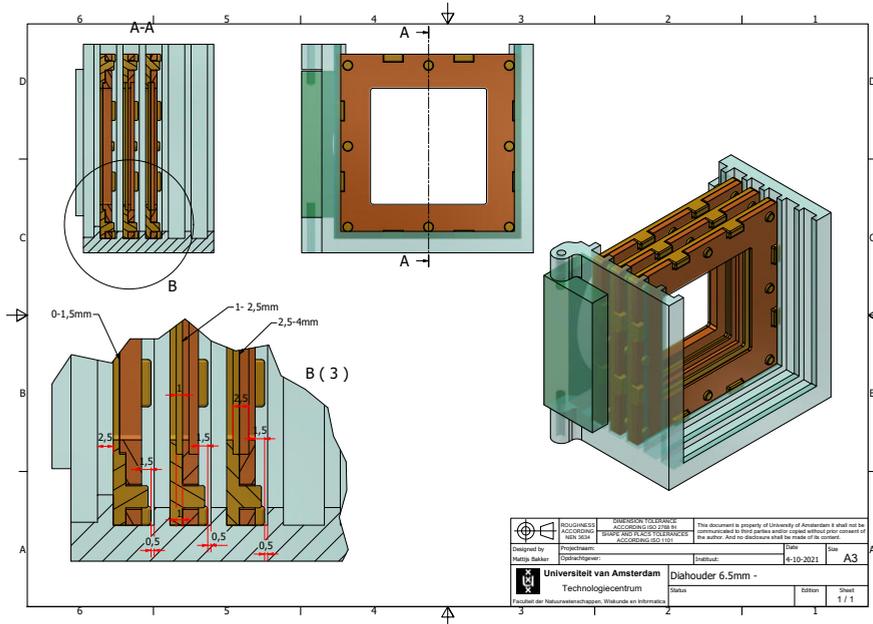


Figure 3.5: Technical drawings of the 3D-printed DIY filter holder.

For the processing of the CT data and 3D reconstruction in this article, the FleXbox software [118] was used. First the data was flat- and dark-field corrected using the corresponding recorded dark-fields (D) and flat-fields (F). Each projection image consists of raw photon counts per detector pixel. With the dark-fields we correct for the off-set counts of the detector system and the flat-field images are used to normalize pixel-dependent sensitivities of the detector. This way the projection images (P) can be corrected and converted into a beam intensity loss image I following the Beer-Lambert law with the following formula:

$$I = \frac{P - D}{F - D} \quad (3.1)$$

After applying the negative logarithm a reconstruction was obtained using the FDK algorithm [59]. Apart from the flat- and dark-field corrections the data was not post-processed with any computational processing filters to correct/reduce image artifacts. Our purpose in this study is to single out the impact of the different acquisition settings on the image formation.

3.3 Experiments and results

In this section we show the influence of different filters and acquisition parameters on the CT image formation and give general guidance on choosing suitable acquisition parameters.

3.3.1 The case studies

To demonstrate the need for an object-tailored acquisition process for cultural heritage objects, we have chosen two case study objects from the textile collection of the Rijksmuseum, Amsterdam, The Netherlands. The objects under investigation were a late 16th-century green velvet purse [169] (case study object 1) and an early 17th-century purple velvet knife holder [170] (case study object 2) which are shown in Figure 3.6. In the following section we will investigate the influence of CT acquisition parameters on the image formation and give general guidance on how to scan these objects. Restricted by the limited time that the cultural heritage objects were available for scanning at our institution we calculated the maximal exposure time available to each scan and kept it at a static value across all scans to be able to conduct all the scans in the different acquisition modes within the available time. Keeping the exposure time constant across different measurements furthermore allowed for a better comparison between the scans and with this we show the necessity for an object-tailored CT acquisition.

To our knowledge there is currently no suitable framework that would give a good indication of improved image quality for visual interpretation and analysis for cultural heritage objects. As there is no ground truth to compare reconstructions to, and as the interplay between artifacts, noise, and contrast will lead to complications for any quantitative metric such as PSNR, SSIM or MSE, we chose to focus here on qualitative evaluation in the context of the ability to answer specific questions about the object based on the CT-images.

The aim of the CT scans was to uncover information about very thin layers within the objects. Therefore, we chose the full detector resolution with 1911×1520 pixels with no hardware binning and used 2881 projections (motivated by the Nyquist-Shannon theorem [94]) for both case studies. All projections have been taken with one frame without averaging. Furthermore, 50 dark- and flat-fields both before and after the scans have been acquired for all case study scans to allow for an adequate pre-processing of the projection data according to formula 3.1.

To be able to detect very thin layers with our scans we aimed to have a voxel-size within our object of $54.9 \mu\text{m}$ (green velvet purse) and $45.0 \mu\text{m}$ (purple velvet knife holder) and positioned the object and detector accordingly. Due to this choice of resolution the objects did not fully fit on the detector in the vertical direction. Therefore, we conducted scans with the detector and source at different heights, to obtain tiled scans of the full object. These were first reconstructed independently and then the resulting volumes were stitched together using the merging capabilities of the Flexbox software [118]. For the velvet purse we used two and for the knife holder four vertical tiles. The total scan time and data size amounted to 163 min and 201.3 GB for the six scans of case study 1 and 138 min and 184.4 GB for the five scans of case study 2.



Figure 3.6: Case study objects "Green velvet purse" (BK-KOG-29 [169]) and "Purple velvet knife holder" (BK-NM-3086 [170]) from the textile collection of the Rijksmuseum, Amsterdam, The Netherlands.

Case study object 1: "Green velvet purse"

At first sight, this peculiarly shaped object looks like a case for a pair of scissors with a small, leather lined purse on its front with a drawstring. However, the fact that it is not possible to open the case itself complicates probing this hypothesis. Of the three examples surviving in the Netherlands (Rijksmuseum BK-KOG-29, Amsterdam Museum KA-18660, and Oranje-Nassau Museum) two are covered with plain velvet and the third is embroidered with seed pearls and gold thread. Other examples, in France (Musée de Cluny, Paris) and Italy (Bargello, Florence) are similar to the Dutch examples. In the Amsterdam Museum example, three coins were found, one of which dates from 1591, indicating that it might have served as a purse for coins. Besides the function as a scissor-case or purse, the strong resemblance of the object's shape to a phallus makes an erotic connotation possible as well [8].

Through the use of CT scans, we hoped to determine how the purse was made, which materials were used and what gives it its shape. One of our main research questions was whether the object box is empty, indicating it might have been a case, or if there is something inside and if so, what. The answers to these questions can be found in the section *Significance for the image interpretation and analysis of the object*.

The purse has dimensions of roughly 11.0 cm in length, 7.5 cm in width and 3.0 cm in thickness. The materials used are velvet on the exterior with metal eyes on the top, as well as leather lining in the pouch on the front. The velvet is adorned with a

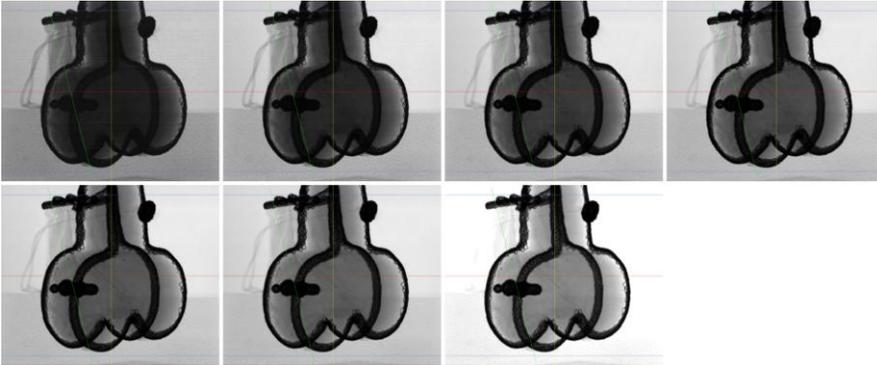


Figure 3.7: From left to right, top to bottom: Radiographs of the "Green velvet purse" with different tube voltages between 20kV and 50kV in 5kV steps with the same tube current (1000 μ A), same exposure time (200 ms) and no filtration.

drawstring of braided gold thread with a silver core. These metal parts in particular make it difficult to scan and have to be investigated with radiographs to make sure that there is limited photon starvation and that sufficient radiation reaches the detector.

Insights and intuitions on the scan design and acquisition parameter choices. To determine the minimal photon energy necessary to penetrate the object, we empirically determined the rotation of the scanned object under which we observe the lowest detector counts. This corresponds to the largest attenuation in the line of sight and was found to be in the orientation visible in Figure 3.7 where the braided gold/silver wire balls overlay the drawstring outline. In this orientation we increased the tube voltage with steps of 5kV while observing the minimal detector count.

We observed that for tube voltages between 20 and 40 kV the object is not well penetrated resulting in dark/occluded areas where almost no signal was detected behind the object in this line of sight. When using tube voltages above 40 kV we see a sudden increase in the minimal count which implies that a minimal photon energy of 40/45 keV is needed to penetrate all regions of the object in this orientation and in the current configuration of Source-Object-Distance (SOD) and Source-Detector-Distance (SDD). Simultaneously, we observed that without filtration the detector is saturated for a tube voltage of 50 kV and a tube current of 1000 μ A.

Furthermore, we reviewed the material composition of the scanned object, i.e. silver and gold, and their main K-edges lie at 25.51 keV and 80.72 keV. To exclude these photon energies, the tube voltage should be set below 80 kV and the filtration should be chosen such that it filters out all photons below an energy of e.g. 30 keV to exclude the silver K-edge. We point out that a tube voltage of 80kV is low for these materials and realize that such low tube voltages are a limiting factor for penetrating thick layers of metal but for our particular CT-system the tube voltages available are bound by regulatory requirements. However, thin layers of metal as presented by the case

study objects can also be penetrated with the available tube voltages and enabled still acceptable results. With higher tube voltages it might then be advisable to filter out photons with an energy lower than e.g. 85 keV.

To give the readers of this work insights and intuitions on how different choices of acquisition parameters would influence the resulting radiographs and reconstructions we designed in total six CT scans for the "Green velvet purse". The six conducted scans include two untailed and unfiltered acquisitions, three single-material filtered CT scans as well as one with a Thoraeus compound filter.

For the untailed scans a standard choice simply based on the radiographs and the measured detector signal would result in an unfiltered scan with a tube voltage of 50 kV and a tube current of 700 μA . A second choice aiming for better object penetration would be choosing a higher tube voltage of 75 kV as well as a reduced tube current of 200 μA to avoid saturation of the detector.

For the filtered scans common filter choices would for example be thin sheets of aluminum filtration to filter out very low-energy photons [61] or to use small amounts of copper or tin filtration at higher tube voltages. Lastly, we conducted a scan with a Thoraeus compound filter. To have enough signal we set the maximal tube current of 1000 μA for all of them.

The case study object has not been moved between the different acquisitions to avoid the necessity to register the reconstructed image volumes. If the object would have been moved between the scans a registration of the image volumes would be necessary to be able to compare the scans directly with each other since the orientation of the 2D slices might be slightly different. Therefore, the acquisition parameters were changed digitally and the replacement of the chosen filtration was carried out very carefully avoiding to touch the scanned object. All scans have been conducted in the "High Power"-mode of the FleX-ray scanner which uses a focal spot size below 45 micrometer for operations up to 90W [43]. The exposure time selected for all the scans was limited by the total available scanning time and had to be restricted to a static value of 200 ms. All other details of the scan settings for this case study that we selected for illustrative purposes can be found in Table 3.1.

Furthermore, all six corresponding beam spectra used for scanning the case study 1: "Green velvet purse" can be found in Figure 3.3. The main insight from this is that the mean photon energy in the spectra shifts continuously to higher energies from ~ 25 keV to ~ 60 keV from scan 1 to scan 6. This means that a higher fraction of the X-ray photons within the beam spectra have sufficient energy to be able to penetrate the object and that dependent on the chosen filtration the severity of photon starvation artifacts will be reduced.

Radiographs and reconstructions of case study 1: "Green velvet purse". In the following section we will show radiographs and the reconstructions of these six CT scans and analyse them. In Figure 3.8, two horizontal cross-section slices from the reconstruction at different heights in the object and the radiographs of all the different scans (1-6)

are presented. We have selected these two cross-sections because they show two levels of severity of the beam hardening and metal artifacts. Once caused by four and once by eight metal trimmings visible in the slices.

In the radiographs for the different acquisition parameters it is visible that changing the tube voltage from 50 kV to 75 kV led to an increased object penetration. Adding aluminum filtration decreased the overall signal but there is no other change in the radiograph observable. Adding stronger filtration leads to less occlusion in the radiograph, which remains only noticeable where multiple metal parts of the velvet purse overlap in the line of sight. However, it also significantly decreases the number of counts and the radiographs appear darker and noisier. The shown projection images have not been adjusted in brightness/contrast level to visually illustrate the decreasing detected signal. Therefore, it is hard to see that there is an increased penetration of the metallic parts of the velvet purse. In the reconstruction the increased penetration of the object can be observed more clearly and how this helps the analysis of the object.

Analysing the reconstruction slices shows an overall image quality improvement from scan 2 towards scan 6. Adding aluminum filtration shows almost no visible difference. The reconstructions of the scans with 75 kV tube voltage show a better penetration of the object but with no filtering there are severe beam hardening and metal image artifacts visible in the cross-sections. With increasing filtering these image artifacts decrease while the noise level rises. Finally, scan 6 has almost no image artifacts within the purse and only small and reduced image artifacts on the corner trimmings. This can be explained by the fact that the minimal photon energy necessary to penetrate the object was determined to be 40/45 keV and the chosen Thoraeus filtration filters out photons with energies between $1.5 \text{ keV} \leq E_{\text{phot}} \leq 40 \text{ keV}$. For the second slice with eight metal trimmings the image artifacts within the left and right part of the purse are almost fully removed but in the middle still heavy image artifacts can be observed. This is due to the overlap of multiple metal trimmings in the line of sight for most of the projection images. Although scan 6 exhibits a higher noise level and even shows beginning ring artifacts due to very low detector signal, the reconstruction cross-section slices provide more interpretable information for the cultural heritage

Table 3.1: Acquisition parameters of the scan of case study 1: "Green velvet purse"

	Scan 1	Scan 2	Scan 3	Scan 4	Scan 5	Scan 6
Beam parameters						
Tube voltage	50 kV	50 kV	75 kV	75 kV	75 kV	75 kV
Tube current	700 μ A	700 μ A	200 μ A	1000 μ A	1000 μ A	1000 μ A
Filters	None	Al 0.5 mm	None	Cu 0.5 mm	Sn 0.25 mm	Sn 0.25 mm Cu 0.5 mm Al 0.5 mm
Camera settings						
Exposure time	200 ms					
Nr. of projections	2881					
Nr. of averages	1					
Geometry						
Source-Object-Distance (SOD)	447 mm					
Source-Detector-Distance (SDD)	609 mm					
Object voxel size	54.9 μ m					

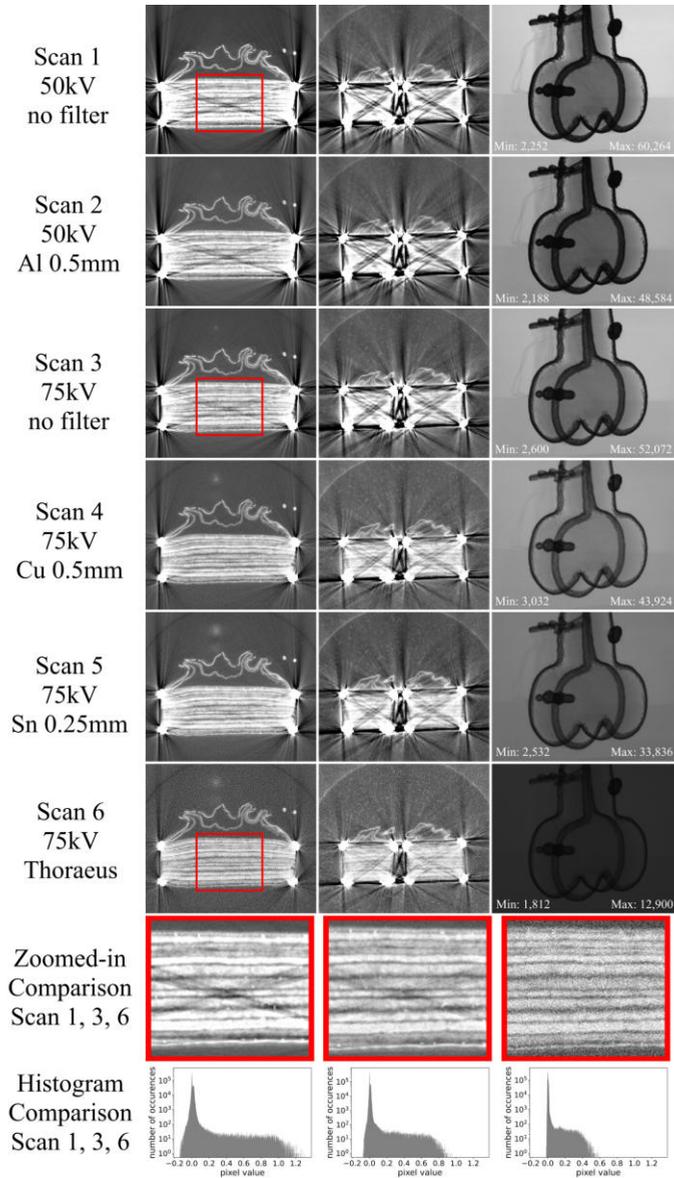


Figure 3.8: From left to right, top to bottom: Two reconstruction cross-sections and one radiograph for all six conducted scans of the "Green velvet purse" with the specific acquisition parameters listed in Table 3.1. The cross-section images shown are taken from slice 1698 and slice 2108. Additionally, comparisons of zoomed-in regions and histograms of scan 1, 3, and 6 of slice 1698 illustrate the improved quality of the object-tailored scan 6.

object's analysis because of the limited beam hardening and photon starvation artifacts. This is also visible in the histograms of the reconstructed slices (cf. Figure 3.8 which show that less pixels have negative or very high values in comparison to the majority of the pixels in the object-tailored scan 6). With more available scanning time the noise level can be compensated by increasing the exposure time but could not be done for our experiments because of time restrictions regarding the scanning of the cultural heritage objects.

Significance for the image interpretation and analysis of the object. Through the object-tailored CT scan (scan 6), in which the shadows caused by the metal thread trimmings were filtered out, it was possible to closely examine the interior of the purse. In both the unfiltered and filtered reconstructions, it became clear that the purse itself is not hollow, but instead stuffed with several layers of a certain material. Counting the layers and identifying the material of this stuffing was complicated in the unfiltered scan, due to the image artifacts introduced by beam hardening. In the filtered reconstruction, it was possible to identify 11 layers of a non-woven material. The fact that this material showed irregular thicknesses, no evidence of weaving, and no different properties on its different sides, as would be the case for leather, strongly suggests that felt was used.

It was possible to identify a leather lining on the pouch on the front of the object. The CT scan, however, clearly revealed that this lining was only attached at the top and bottom of the pouch, and detached throughout the rest of the pouch (cf. Figure 3.8). The velvet is furthermore damaged in certain parts, and repaired with a paper-like material with a matching green colour.

The object-tailored CT scan allowed through its artifact reduction for an easier analysis of the cultural heritage object, for the identification of the filler material based on its structural characteristics as well as provided insights into its build-up. While it was possible to disprove the hypothesis of the purse being hollow and made for holding scissors or other small objects, its precise use remains unknown for now.

Case study object 2: "Purple velvet knife holder"

This early-17th century purple velvet knife holder is elaborately decorated with gold and silver embroidery, and freshwater seed pearls in floral motifs. The removable cap reveals golden decorations of two hands holding a flaming heart and three internal compartments, the contents of which are unfortunately lost. The question arises what it could have contained, besides presumably a knife and a fork, the case being too narrow to contain a spoon unless quite small and shallow. The materials used, and type of decorations present, are reminiscent of a contemporaneous pouch at the Rijksmuseum (BK-NM-11110).

The CT scan was conducted to investigate the construction and material composition of the velvet knife holder. The shape and depth of the third cavity might provide insight into what kind of third object the case once held. Additionally, we wanted

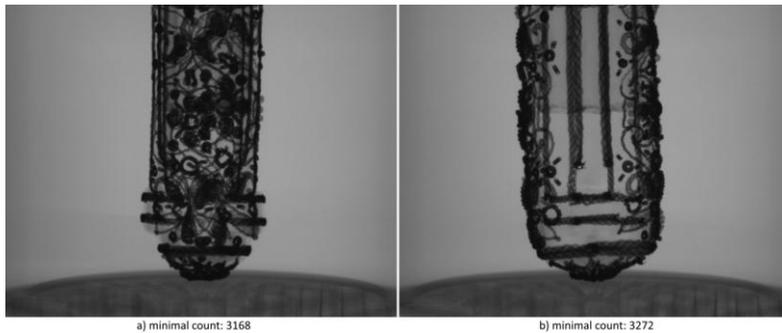


Figure 3.9: Radiographs of the 0 degree and 90 degree position of the knife holder with same tube voltage (35 kV), the same tube current (1000 μ A), same exposure time (100 ms) and no filtration.

to discover if the textile was pre-embroidered and how it was attached to the case. The answers to these questions can be found in the section *Significance for the image interpretation and analysis of the object*.

The Rijksmuseum knife holder has dimensions of roughly 22.0 cm in length, a maximum of 3.4 cm in width and a maximum of 2.5 cm in thickness. The relevant materials used are paper or wood in its interior, with a velvet exterior decorated with gold and silver thread as well as small pearls.

Insights and intuitions on the scan design and acquisition parameter choices. For this second case study object, the availability of prior information of scanning a similar object (case study object 1) in terms of materials (velvet, leather, gold/silver threads) simplified the scan design process. In contrast, the dimensions of this object were a bit different. A reconstruction with a standard, untailed parameter selection, which exhibits severe image artefacts, was already shown in Figure 3.1. Here, we will focus on the importance of choosing the correct rotation and the trade-off regarding the fine tuning of the tube voltage. The impact of the different parameter choices will be shown with the corresponding beam spectra, radiographs and reconstructions.

The purpose of the radiographs shown in Figure 3.9 is to illustrate the importance of quantitatively determining the respective minimal count that corresponds to the largest attenuation in the line of sight. Although it might be non-intuitive the 0 degree position is causing a higher attenuation than the one in the 90 degree position. Without determining this rotational position there might be lines of sight where the object is not penetrated sufficiently. This can lead to image artifacts due to photon starvation.

After finding this rotational position we determined the minimal photon energy needed to penetrate the object again. We observed that the needed minimal photon energy is 35/40 keV and that with no filtration for a tube voltage of 60kV our detector was already saturated. For a tube voltage of 50 kV we saw sufficient minimal detector signal when not taking the object's characteristics into account. But with prior

knowledge from the previous case study we decided again to use a "Thoraesus filter" and to stay below a tube voltage of 80 kV because of the K-edges of gold.

To illustrate the trade-off between the reduction of beam hardening image artifacts, improved contrast, and reduced noise we designed four CT scans with identical acquisition parameters which only differ in their respective tube voltage (see Table 3.2). To have enough signal we had to choose the maximal tube current of 1000 μA for all of them. The exposure time selected for all the scans was limited by the total available scanning time and had to be restricted to a static value of 200 ms

Table 3.2: Acquisition parameters of the scans of case study 2: "Purple velvet knife holder"

	Scan 1	Scans 2-5
Beam parameters		
Tube voltage	50 kV	60/65/70/75 kV
Tube current	500 μA	1000 μA
Filters	None	Sn 0.25 mm Cu 0.5 mm Al 0.5 mm
Camera settings		
Exposure time		200 ms
Nr. of projections		2881
Nr. of averages		1
Geometry		
Source-Object-Distance (SOD)		234 mm
Source-Detector-Distance (SDD)		389 mm
Object voxel size		45.0 μm

The impact of the different parameter choices can be already seen in the corresponding beam spectra in Figure 3.3 and will be described and illustrated in more detail in the radiographs and reconstructions in Figure 3.10. When looking at the four different beam spectra it can be seen that three important changes are introduced: (i) The change of the maximal photon energy in the spectrum; (ii) the change of the mean photon energy in the spectrum; (iii) the change in overall intensity of the beam. This will influence the penetration of the X-ray photons in the object, the contrast of the images, and the overall SNR.

Radiographs and reconstructions of case study 2: "Purple velvet knife holder". In Figure 3.10 we show the results of our investigation trying out different tube voltages between 60 and 75 kV while keeping all other acquisition parameters the same. This was done since for many materials common in cultural heritage the absorption curve has a steeper slope for lower photon energies. This leads to higher contrast between the different materials. The radiographs in Figure 3.10 show, similarly to the beam spectra in Figure 3.3, that with lower tube voltage also the overall signal decreases which makes the radiographs a bit more noisy. For tube voltages of 60 or 65 kV the signal measured by the detector is very low and is even close to the dark currents of our detector.

At the same time the change of tube voltage influences the extent of the beam hardening image artifacts due to the shift of the mean and maximal photon energy in

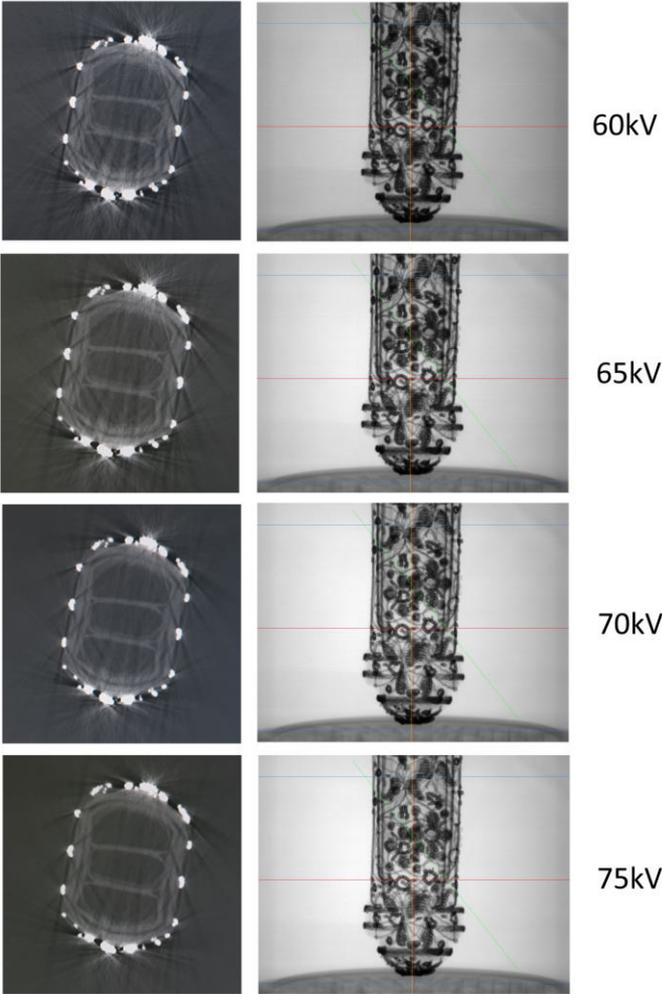


Figure 3.10: From top to bottom: Reconstruction slices and radiographs of the "Purple velvet knife holder" of scans 2 - 5 with the specific acquisition parameters listed in Table 3.2. The cross-section images shown are taken from slice 1084.

the beam spectrum. Whereas for scans 4 and 5 there are almost no image artifacts present in the inside of the object, scans 2+3 show more dominant image artifacts. Also, the beam hardening image artifacts in the cavities close to the outer casing are more prominently visible for these two. Nevertheless, there is an increased contrast noticeable between the different materials. This shows that there is a trade-off between having less image artifacts, improved contrast and reduced noise. Scan 4 exhibits a reduction of beam hardening effects with overall acceptable noise as well as contrast levels and was found qualitatively the best trade-off between these objectives.

Significance for the image interpretation and analysis of the object. A clear improvement is noticeable when comparing the unfiltered reconstruction slices of scan 1 with the filtered reconstruction slice of scan 4. In the scan 1 reconstruction (as presented in Figure 3.1), features of the inner structure are hard to interpret as a result of the image artifacts introduced by beam hardening. In the scan 4 reconstruction, however, these features are more well-defined and easier to distinguish.

This difference in scan quality, supplemented with visual observation of the object, allowed us to gain a better understanding of the materials used and the construction of the object. The three compartments are separate cylindrical shapes, made from (possibly stiffened) paper, that are inserted into a leather outer shell that holds them together. The velvet exterior is attached to this leather outer shell, presumably using glue. The middle cylinder, or cavity, gets pressed together by the adjoining cavities, as the case gets narrower towards its bottom. The shape of the cavities did not give any indications on what the third object could have been that the knife holder once held. The crisp features of the filtered reconstruction allowed us to investigate whether the leather outer shell had any holes, indicating that the exterior was embroidered after attaching it to the object. As no holes were found, however, we concluded that the pearls and floral motifs were applied to the velvet prior to it being glued to the object. The golden plated strips running along the sides of the case covering the seams could perhaps be added after application, as these are most likely glued onto the velvet as opposed to being stitched.

3.3.2 General concept of steps for an object-tailored CT scan

With the insights and intuitions we gained while investigating the two case study objects, testing various acquisition parameters and observing their influence on the CT image formation, we reflected on a general concept of steps for designing an object-tailored CT scan. Although all investigations were carried out for a specific CT system as well as specific objects and need to be optimized for individual cases, the following six steps can guide cultural heritage professionals through the process of finding suitable CT acquisition parameters:

1. Determine a suitable SOD and SDD to achieve the desired resolution in the CT reconstruction of the object under investigation. Please be aware that a later change in the tube current or exposure parameters might affect the focal spot size and reduce your resolution.

2. Empirically determine the rotational position with the lowest detector signal which corresponds to the largest attenuation in the line of sight.
3. Investigate the minimal photon energy needed to penetrate the object in this rotational position by slowly increasing the tube voltage while observing the minimal detector signal.
4. Choose adequate filtration that can filter out the low energy part of the spectrum.
5. Exclude the K-edges of any metals within the object by adjusting the tube voltage accordingly.
6. Calculate the maximal feasible acquisition time per projection image of the detector from the total available scanning time to achieve the highest SNR that satisfies the given time and saturation constraints. Also, consider taking multiple frames for averaging each projection, if the total available scanning time allows for that.
7. Find a balance for the trade-off between the reduction of beam hardening image artifacts, improved contrast, and reduced noise by adjusting the tube voltage and tube current within the given boundaries.

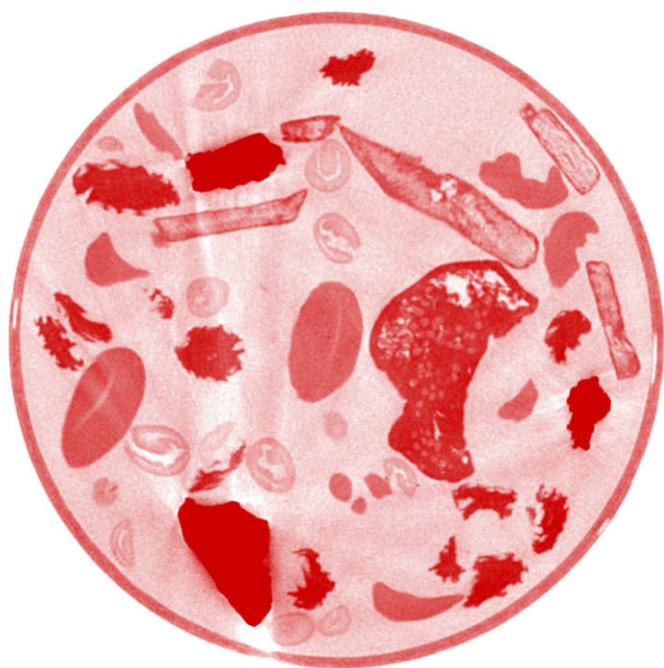
3.4 Discussion and conclusion

In this study we have introduced how the multi-material nature of cultural heritage objects exhibit challenges for CT scanning and presented several key factors that influence CT image formation. We addressed their underlying physics and discussed how the concept of beam filtration can improve the image quality of CT scans. Illustrated on two case study objects the influence of CT acquisition parameters was investigated and we gave general guidance on how to design a CT scan and how to choose suitable acquisition parameters. We showed that an object-tailored CT acquisition of two case study objects yielded a clear image quality improvement and helped the cultural heritage experts in their analysis. Lastly, we extracted a general concept of steps for museum professionals to design an object-tailored CT scan for individual cases.

Despite the clear image quality improvement of this object-tailored method compared to an untailored CT acquisition, the main disadvantages are firstly that through the filtering the beam intensity is reduced, which results in a decreased SNR, and secondly that the beam hardening problem cannot be completely eliminated [87]. To counter the degraded X-ray signal and greater image noise, longer exposure times and averaging can be used [92]. Depending on the available scanning time and capabilities of the scanning facility these disadvantages can therefore be mitigated. Overall, the object-tailored choice of acquisition parameters can help though to produce a cleaner image by absorbing the lower energy photons that do not reach the detector and that tend to scatter more. The beam filtration also enables the use of a higher voltage,

current and longer exposure times since it prevents the X-ray detector to become saturated through the low-energy photons in the beam spectrum.

The results presented in this manuscript are of course for a specific CT system as well as specific objects and need to be tailored to new individual cases. With the concept of steps for an object-tailored CT scan design we provided a short guidance for this, while the two case studies provide insights and intuitions on choosing suitable acquisition parameters that take the objects' characteristics into consideration and improve image quality in CT reconstructions.



4

2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning

X-ray computed tomography (CT) is a non-invasive X-ray absorption-based imaging technique used in a range of fields, including medicine, manufacturing industry, food industry, and materials science. For a CT scan, X-ray projection images of an object are taken from multiple angular positions. To obtain a reconstruction of this acquired data, an inverse problem has to be solved through analytical methods such as filtered back-projection or iterative reconstruction algorithms. The overarching methodology to reconstruct images from these measurements is called computational imaging.

In recent years, the field of computational imaging focused on developing data-driven methods for image reconstruction [166]. With that focus also the need for large datasets increased. In particular, image reconstruction based on deep learning (DL) methods, such as deep neural networks (DNNs), requires a large amount of realistic data for both evaluating the developed methods on real world applications as well as for constructing the method itself. For example, supervised learning approaches optimize the network parameters based on training data composed of a large number

This chapter is based on:

M. Kiss, S. Coban, K. Batenburg, T. van Leeuwen, and F. Lucka. “2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning”. *Scientific Data* 10.1 (2023), p. 576.

of representative pairs of input and desired ideal output data of the network (i.e., the ground truth).

While in various application fields of DL there already exist large and open data collections such as MNIST [123] for handwritten digit recognition or IMAGENET for image classification/processing [46], suitable experimental data collections for computational imaging with high-quality ground truth reconstructions and/or segmentations are scarce. For magnetic resonance imaging (MRI), the fastMRI dataset [116] is a larger dataset containing raw (unprocessed) k-space data of knees and brains acquired across multiple institutions and scanners. There are some datasets available for X-ray CT but unfortunately they lack certain desirable characteristics: The Mayo clinic low-dose CT challenge of 2016 [141] with 30 patient scans consisting of roughly 70 slices each has a fairly small number of scan subjects. Although their new release of 2021 [144] has 300 patients another important downside of both datasets is that noisy reconstruction and projection data is simulated from clean reconstructed volumes. The LoDoPaB-CT dataset [124] contains over 40,000 scan slices from around 800 patients selected from the LIDC/IDRI database. But despite the large size of the data collection it still uses simulated low photon count measurements and not experimental data. The walnut dataset [47] provides 42 three-dimensional cone-beam CT (CBCT) scans of walnuts. Although it provides raw experimental data, the applicability of the dataset is limited through the small number of samples of the same object type and its design for a specific task in 3D CBCT. This makes it less useful for more general methods development. Overall, the few available datasets are limited in their applicability to one computational imaging task.

A key disadvantage of available datasets for X-ray CT is that they commonly use commercial CT solutions with licensed software that have no (or limited) access to raw projection data or the specifics of the experimental acquisition. Therefore, mathematical and computational studies typically rely on artificial data simulated with varying degrees of realism. To develop and train algorithms for computational imaging tasks such as low-dose reconstruction, limited or sparse angular sampling, beam-hardening artifact reduction, super-resolution, region-of-interest tomography or segmentation it is necessary to have corresponding experimental training data. The field of X-ray CT still lacks such a large-scale, versatile, experimental dataset for machine learning. Especially two-dimensional, reconstructed CT slices would be useful for method development since the corresponding learning and reconstruction tasks require less computational resources compared to their three-dimensional counterparts.

Acquiring such a large 2D CT dataset encompasses various requirements: First, research groups need to have a scanning facility readily available and be able to also use it for a large-scale, time-extensive data collection process. Second, the geometry and other acquisition parameters of this scanner must be highly adjustable to collect a dataset that can be used for a wide range of machine learning applications. Third, similar image characteristics as encountered in medical CT would be preferable because of the great importance of medical imaging as an application area of X-ray CT means. Last, it is necessary to limit manual intervention during the acquisition process to be

able to reach a high number of acquired CT reconstruction slices. This requires the ability to automate the acquisition process as much as possible.

In this work, we describe in detail the steps involved in acquiring an unprecedented X-ray data collection by making extensive use of a highly flexible, programmable and custom-built X-ray CT scanner: The first step was to choose suitable scan parameters for the acquisition such as beam filtration, X-ray tube voltage and current, detector exposure time, binning and averaging, the number of projection angles and source, object and detector positions. The aim behind these choices was to acquire a rich projection dataset for each image slice that can be used for a wide range of imaging tasks such as supervised or unsupervised denoising, sparse-angle scanning, beam-hardening reduction, super-resolution, region-of-interest tomography or segmentation. It therefore provides a starting point for algorithm development with realistic experimental data. As a second step, a scanning object had to be designed in such a way that 2D slice scans resemble image features found in medical abdominal CT scans. For this a cylindrical tube was filled with a mix of samples of similar density but different shapes immersed in a powder. In the third step, an experimental set-up and a script generator program was developed that allowed to automate the collection of 50 slices during an 8.5h scan. In 111 scanning sessions (each with a different sample mix) and a total scanning time of more than 850 hours we acquired 5,000 slices over a duration of almost five months. Each of these slices was acquired in three different acquisition modes resulting in "clean", "noisy", and "artifact-afflicted" reconstructions. Furthermore, additional 750 out-of-distribution (OOD) slices were acquired, for which each ingredient was scanned separately, and scans with different parameters and/or samples were designed. This was done to accommodate robustness tests or help with segmentation / foreign object detection tests.

To make this dataset accessible to a broad range of researchers including those who do not have high-performance computing facilities readily available, we provide also reference reconstructions and segmentations, additionally to the raw projection data in sinograms and an implementation of the complete computational pipeline based on open-source software (cf. Figure 4.1).

The structure of this work is as follows: The section "Experimental design" describes the overall study design, the experimental design of the acquisition and data processing protocols. In the section "Data records" the file structure of the data collection is described, while the "Usage notes" section contains all details on how to access and use it.

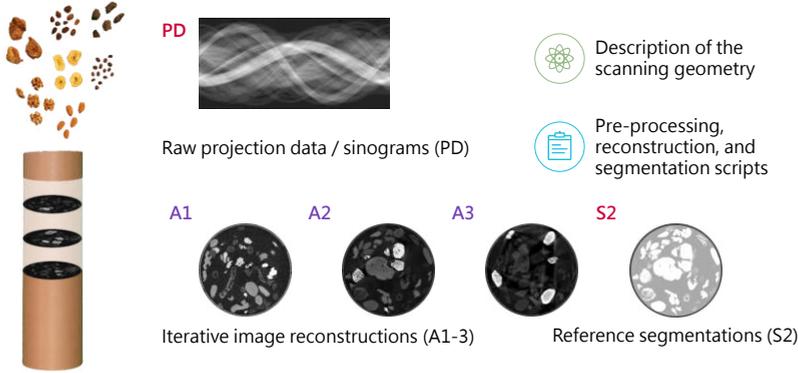


Figure 4.1: Overview of the scope of the 2DeteCT dataset.

4.1 Methods

4.1.1 X-ray computed tomography scanner

The data collection has been acquired using a highly flexible, programmable and custom-built X-ray CT scanner, the FleX-ray scanner[43], developed by TESCAN-XRE NV (<https://info.tescan.com/micro-ct>), located in the FleX-ray Lab at the Centrum Wiskunde & Informatica (CWI) in Amsterdam, Netherlands. It consists of a cone-beam microfocus X-ray point source (limited to 90 kV and 90 W) that projects polychromatic X-rays onto a 14-bit CMOS (complementary metal-oxide semiconductor) flat panel detector with CsI(Tl) scintillator (Dexella 1512NDT, [131]) and 1536×1944 pixels, $74.8 \mu\text{m}^2$ each. To create a 2D dataset, a fan-beam geometry was mimicked by only reading out the central row of the detector. Between source and detector there is a rotation stage, upon which samples can be mounted, cf. Figure 4.2. The machine components (i.e., the source, the detector panel, and the rotation stage) are mounted on translation belts that allow the moving of the components independently from one another. Furthermore, we developed an in-house software toolbox for designing executable scan scripts for the scanner. With these, the scanner performs sophisticated acquisition protocols automatically without human intervention. The acquisition procedure that we programmed with this toolbox will be described in more detail in the "Data acquisition" section.

4.1.2 Experimental design

The three general aims of this data collection are as follows. First, to provide the computational imaging community with a dataset that encompasses raw experimental measurement data that can be used to develop and test techniques in CT imaging for real world applications. These also include medical CT by producing similar image

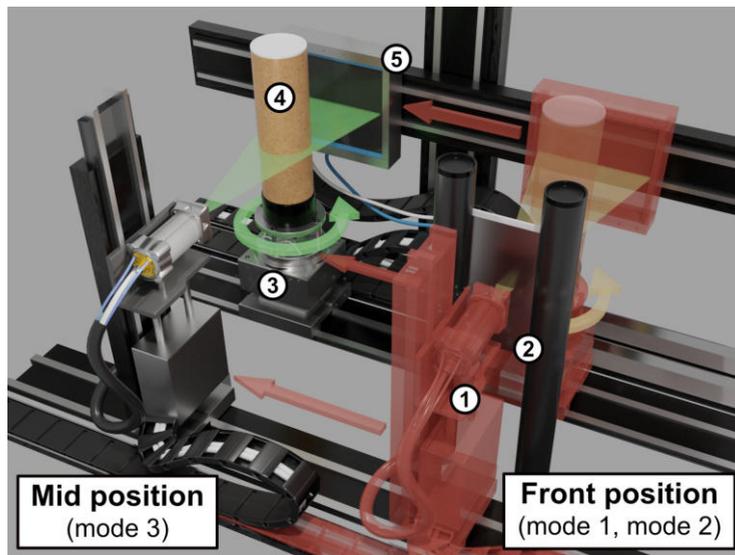


Figure 4.2: FleX-ray Lab: the computed tomography set-up used for the data acquisition. 1) cone-beam X-ray source; 2) Thoraeus filter sail; 3) Rotation stage; 4) Sample tube; 5) Flat panel detector. The objects 1, 3, 4, and 5 move from their red transparent front position to the mid position for the acquisitions of mode 3. In both positions 3,601 projection images per slice are taken while the object rotates 360 degrees.

features and contrast in the dataset slices as exhibited in medical abdominal CT scans. Second, the possibility to expand the current scope of the dataset by adding more detailed multi-class segmentations or by adding more slices with the same or a different sample mix because of a reproducible setup. Third, to have a large enough dataset such that it can be used for developing deep learning algorithms for different computational imaging applications, including low-dose acquisition, limited or sparse-angle scanning, beam-hardening artifact reduction, super-resolution, region-of-interest tomography or segmentation. The three key features to achieve this were firstly, the design of a semi-automatic data acquisition; secondly, finding a scanning object and sample mix that is both, diverse enough as well as stable over the long scanning time; and thirdly, creating a scanning setup and experimental design that enables the aforementioned applications. A total of 9 months went into the experimental design, developing and scripting the semi-automatic data acquisition, selecting and testing the different samples, designing the scanning setup and determining suitable acquisition parameters.

Semi-automatic data acquisition

To maximize the number of scanned CT slices it was strictly necessary to limit the amount of human intervention during the data acquisition. The idea was to use the flexibility of the scanner and its ability to be scripted to automate the acquisition

process as much as possible. The only necessary human interaction would be to prepare the next sample mix and to start a scanning protocol, which then would acquire a certain number of slices automatically. In our optimized setting, acquiring a batch of 10 slices takes a fixed time of 1h 42mins, while 50 slices can be acquired in 8h 34mins. This way, the idle time of the scanner, e.g. at night, could be used to reach a large amount of slices.

Sample preparation

One of the aims of this dataset was to produce images with similar image features and contrast as abdominal medical CT scans. To achieve this, a container representing the body was used as a scanning object and was filled with a mix of cm-scale objects mimicking the organs/bones submersed in a background medium resembling connective tissue. Furthermore, this mix of sample objects should have a high natural variability in both inter and intra-sample shape and density (see Table 4.1). In particular, one of the samples should be dense enough to correspond to bones/teeth and introduce beam-hardening effects. Lastly, this sample mix should stay stable during long lasting high-intensity X-ray exposure and therefore have especially a certain temperature stability.

To fulfill these requirements a variety of dried fruits and nuts were scanned and their appearance as well as their relative intensities in the reconstructed image slices were evaluated. Although the actual attenuation coefficients of the samples are not the same as for organs, bones and connective tissue in medical CT scans, the ratio between dense and soft regions are similar. Furthermore, some of the dried fruits and nuts have shapes that resemble organs, e.g. walnuts resembling brain tissue. The tested samples included: dried apricots, bananas, dates, figs, mangoes, raisins, and coffee beans as well as almonds, cashews, hazelnuts, para nuts, peanuts, pecans, pistachios, and walnuts. Furthermore, different stone types were tested for the sample selection as objects that introduce beam hardening effects such as different types of basalt, granite, lime stone, lava stone, marble, quartz, and slate. To avoid air volumes between the samples, various filler materials were investigated: cereal-based coffee powder, sand, saw dust, (powdered) sugar, salt, and sweetener. The requirements were similar or lower density than water, contrast to other samples, no/limited air bubbles, not too coarse in its fine structure, rather homogeneous, and temperature stable regarding aggregate state, physical extension, as well as density/humidity.

Object	Density (g/cm^3)
Cardboard tube	0.689
Cereal coffee powder	0.260
Banana	0.422
Coffee beans	0.432
Walnut	0.494
Almond	0.507
Raisin	0.612
Fig	0.629
Lava stone	1.5 - 1.9*

*depending on its porousness of 20 - 50%

Table 4.1: Reference densities of the final sample mix as published by the Agricultural Research Service of the U.S. Department of Agriculture (<https://www.aqua-calc.com>)

After evaluating all sample mix tests, the following sample mix was selected: almonds, dried banana chips, coffee beans, dried figs, lava stones, raisins, and walnuts immersed in cereal-based coffee powder as a filler material. These samples were chosen to have three categories of densities, where in each category the objects differ mainly in size, shape, and fine structure but have a similar density. This is comparable to the medical imaging domain in which most organs have a similar density but differ mainly in size, shape, and fine structure. For our dataset, dried banana chips and coffee beans were chosen in the low-density category. In the medium-density category two samples had to be chosen from a diverse mix of nuts. Due to the bigger size of their fine structure walnuts were chosen over pecans. For the second nut, almonds were selected over cashews, hazelnuts, para nuts, peanuts, and pistachios because of their medium size and their simpler shape. In the high-density category raisins were selected as a small-size sample and figs were found to be the most interesting samples in terms of fine structure in comparison to apricots, dates, and mangoes. The respective densities of these samples can be found in the appendix in Table 4.1. Since X-ray absorption is related to the density and thickness of the scanned material, these values are a first indicator for the measured intensities and their contrast to each other.

To avoid the samples to dry up too much over time through the long exposure to high energy X-ray radiation, the sample mix was replaced three times in total. The amount of each sample within the final mixes can be found in Table 4.2 and was based on yielding a roughly equal share of volume within the sample mix.

Sample	Mix 1 (slices 1 - 1,800)	Mix 2 (slices 1,801 - 3,720)	Mix 3 (slices 3,721 - 5,000)	Mix OOD (slices 5,521 - 6,370)
Cereal coffee powder	400 g	400 g	400 g	400 g
Banana	75 g	79 g	80 g	80 g
Coffee beans	60 g	58 g	54 g	54 g
Walnut	84 g	73 g	79 g	79 g
Almond	117 g	112 g	111 g	111 g
Raisin	111 g	100 g	110 g	110 g
Fig	282 g	285 g	290 g	290 g
Lava stone	171 g	154 g	177 g	177 g
Fresh fig	0 g	0 g	0 g	5 pieces*
Grape	0 g	0 g	0 g	121 g*
Hazelnut	0 g	0 g	0 g	105 g*
Pistachio	0 g	0 g	0 g	87 g*
Peanut	0 g	0 g	0 g	55 g*
Titanium prostheses screws	0 g	0 g	0 g	4 screws*

Table 4.2: Sample distribution for the three different sample mixes. * only one of them included in the mix during each OOD scan

The aforementioned sample mix was put into a sample container as a scanning object. The requirements for this were apart from a stable positioning and temperature/radiation exposure stability that the container did not absorb too much of the X-ray radiation. After testing different paper and plastic (PE, PVC, etc.) based containers a cylindrical cardboard tube was selected.

Because of the restrictions imposed by the inner dimensions of the scanner and the maximal diameter usable on the scanner sample stage (109.4 mm) a cardboard tube with 10 cm inside diameter and 34 cm of height was used. To prevent unwanted

scattering from the aluminum sample stage the sample tube was elevated by a few centimeters with a 3D-printed PETG sample stage cylinder of 50.0 mm in height. To ensure that the setup remains the same this cylinder can be steadily and reproducibly positioned into a carved out circle on the sample stage and the cardboard tube was positioned centrally onto this sample stage cylinder using superglue.

For the additional 750 out-of-distribution (OOD) slices, each ingredient in the mix was scanned separately and new samples were included to accommodate robustness tests or help with segmentation or foreign object detection tests. The OOD sample objects were fresh figs, grapes, hazelnuts, pistachios, peanuts and titanium prostheses screws. The latter were chosen to have objects in the sample mix that create even more severe artefacts than the lava stones and furthermore are used in clinical practice.

Scanning setup and parameter choices

The scanning setup had to be designed in such a way that the different application areas for the data collection could be served. Namely, denoising, sparse-angle scanning, beam-hardening reduction, super-resolution, region-of-interest tomography or segmentation. Since data for sparse-angle, super-resolution and region-of-interest tomography can be generated from scans with a large amount of angle projections and high resolution, the objectives were to acquire at least a noisy, a beam-hardening artifact-inflicted and a "clean" scan as a ground truth and starting point for constructing a high-confidence segmentation.

Therefore, data has to be acquired in three different acquisition modes: Mode 1 being the noisy, low-dose acquisition; mode 2 being the clean, high-dose acquisition; and mode 3 being the beam-hardening artifact-inflicted acquisition. To achieve both beam-hardening artifact-inflicted and "clean" images from the same sample it is necessary to scan with and without beam filters. Since the FleX-ray scanner does not encompass an automatic filter wheel we developed a new kind of setup within our scanner cabinet. To be able to scan multiple vertical slices with a filter we built a "filter sail" which was placed between the X-ray tube and the sample tube. Since the source, rotation stage and detector can be moved at the same time it was possible to get this filter sail into the beam axis without human interaction during the process (cf. red arrows in Figure 4.2).

For this a second scanning position in the scanning cabinet was identified which is located in the front instead of the middle of the scanner (on the transversal axis). After testing that this scanning position yields indistinguishable results with otherwise identical acquisition parameters to the "mid position" a scanning script was developed that will move between these two position depending on which acquisition mode shall be used at that moment.

The "front position" was used for mode 1 and 2 whereas mode 3 was acquired in the "mid position". This distinction was necessary since mode 1 and 2 require a filter setup while measurements in mode 3 are acquired without a filter. This means in the "front position" the X-rays go through the filter sail whereas in the "mid position"

they are not filtered. To limit the time spent on motor movements the scans were carried out in batches of 10 slices in each mode.

The next choice for the scanning setup regarded the positioning within the scanner cabinet and encompassed three objectives: Sufficient photon flux, high resolution, and good detector coverage. These are mainly influenced by two parameters, the so-called Source-to-Object Distance (SOD) and Source-to-Detector Distance (SDD). The bigger the SDD is, the more parallel the beam geometry is. At the same time though the photon flux decreases. While it is desirable to have a parallel beam geometry, a decrease in photon flux necessitates longer scanning times or the noise increases which prolongs the data acquisition significantly. The SOD and its ratio with the SDD, called magnification factor $mag = \frac{SDD}{SOD}$, determine the resolution of the scan. This means what length inside the object is covered by one detector pixel det and accordingly, the resolution can be calculated as follows: $res = \frac{det}{mag}$. We strived to maximize spatial resolution subject to the constraint that the size of the scanned object does not exceed the size of the detector.

The positioning of both sample tube and detector was limited by the dimensions of the inside of the scanner, the sample tube, the size of the detector, and the possible motor movements. With a minimal motor position distance between detector and rotation stage of 63 mm on the magnification axis, 529 mm was found to be a suitable SDD and increasing the distance to the sample tube to an SOD of 431 mm increased the footprint of the scanned object with respect to the width of the detector. This resulted in a sufficient photon flux and a resolution of $60.95 \mu\text{m}^3$. Although the size of our scanning object (10 cm diameter) is much smaller than a patient's body in medical CT scans (up to 50 cm diameter) the ratio between the typical resolution and size are comparable (60 and $300 \mu\text{m}$ respectively).

After the selection of the sample mix and sample tube as well as the positioning suitable beam parameters for the desired application areas had to be found. Two main problems arise though for the chosen scanning setup and experimental design: Firstly, lab X-ray sources emit a spectrum of X-ray energies. They have a so-called polychromatic beam in contrast to synchrotron facilities which have a monochromatic beam consisting only of X-rays of one distinct energy [5]. The broader the beam spectrum is, the larger the occurring beam hardening effects are [35, 158]. Secondly, low energy photons are absorbed more strongly by larger objects limiting the amount of detected photons. Therefore, a low average beam energy leads to high noise. Hence, the beam spectrum of the X-rays needed to be optimized to produce "clean" reconstructions with limited noise and beam-hardening artifacts.

For this different combinations of tube voltage and filters to shape the beam spectrum were tested. First, the lower bound of the X-ray energy required to penetrate the object was determined. A tube voltage of 40.0 kV produced virtually no signal on the detector and the noise was too high. For a tube voltage of 60.0 kV and maximum current of $1000 \mu\text{A}$ a sufficient amount of photons was measured, but without using filters noticeable beam hardening artifacts were observed in the reconstructions. Next, the tube voltage was set to the maximum of 90.0 kV and beam filtration was used

to improve the image quality. To reduce the beam hardening effects the low-energy photons were filtered out by placing thin sheets of metal between the X-ray tube and the sample tube. A variety of filters of different materials (Al, Cu, Sn, W) and thicknesses (0.01 mm - 0.50 mm) as well as combinations of them were tested. A popular compound filter in CT imaging, the so-called "Thoraeus filter" [93], showed the best performance. Its compound consists of a tin filter, followed by a copper and after that a aluminum filter of varying thicknesses and effectively reduces the amount of photons carrying an energy of 1.5 keV to 70.0 keV. After testing a variety of thicknesses for the different compounds the final filter setup was composed of Sn = 0.1 mm, Cu = 0.2 mm, and Al = 0.5 mm. To have sufficient signal the tube current had to be set again to the maximum of 1000 μ A. Figure 4.3 illustrates that with the chosen compound filter almost all photons with energies below 40.0 keV are filtered out of the beam spectrum which will reduce the beam hardening artifacts.

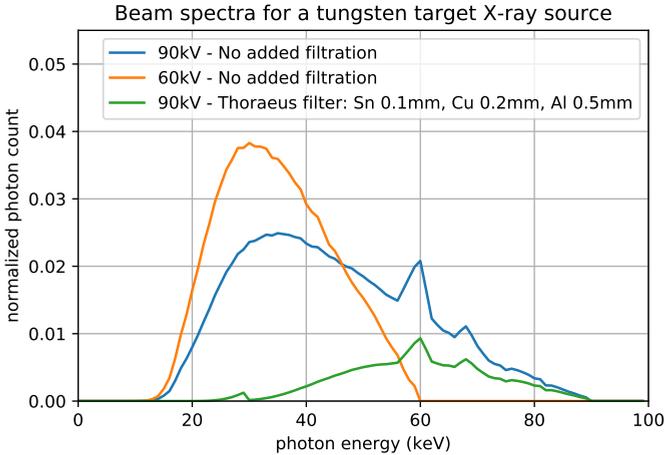


Figure 4.3: Beam spectra of a tungsten target X-ray source with an X-ray exit window made of 300 μ m Beryllium operated at 60kV with no added filtration and 90kV tube voltage with no added filtration and filtered with a Thoraeus filter of Sn = 0.1mm, Cu = 0.2mm, Al = 0.5mm simulated by TASMIP software [22].

Tube voltages employed by medical CT scanner commonly range between 80 and 140 kV and sometimes also use tin filtration between 0.4 and 0.8 mm [161]. The maximum tube voltage available for our particular CT-system is bound by regulatory requirements to 90 kV but nevertheless can be considered comparable to medical CT scanners. For the low-dose acquisition setting we chose a current of 33.3 μ A resulting in a tube power of 3.0 W. According to Lee W. Goldman the relationship between dose and tube current is linear [70]. Therefore, mode 1 is a 1/30 dose acquisition in comparison to mode 2. Note however, that we maximized the dose in mode 2 to obtain the best possible image quality while ensuring sufficient slice throughput. In medical imaging, the "full" dose is typically chosen as low as reasonably achievable (ALARA)

to still obtain images with sufficient quality for the clinical task at hand. In clinical practice, “tube current”-“exposure time” products range between 50 and 400 mA s [168]. The clean, high-dose acquisition of mode 2 has a “tube current”-“exposure time” product of ~ 18.0 mA s, whereas the one of the noisy, low-dose acquisition of mode 1 is given by 0.6 mA s. However, this comparison has to be taken with caution since the scanning object is much smaller compared to a medical CT scan and the geometry of the scanning setup largely differs from a clinical setup. The circumference of our scanning object is roughly 35cm whereas 50 to 100 cm are standard abdominal circumferences for children and adolescents [168].

Lastly, the exposure time and the number of projections for the acquisition of the scans was chosen. For the latter an application of the Nyquist-Shannon sampling theorem to CT yields that the achievable image resolution is not limited by the angular sampling rate if the number of projections is chosen greater than the number of detector pixels times $\pi/2$ [94], which amounts to $1912 \times \frac{\pi}{2} = 3003$ in our setting. Therefore, 3,601 projections were chosen where the first and last projection coincide to have a standard angular increment of 0.1 deg. For the exposure time 50 ms (20 Hz) ensured that all projections acquired for one slice are obtained within 3min scanning time without saturating the detector in any of the acquisition modes. All projection images were taken without any hardware binning or averaging. The summary of the acquisition parameters used can be found in Table 4.3.

Acquisition parameter	Mode 1	Mode 2	Mode 3	Noise-OOD
Tube voltage	90.0 kV	90.0 kV	60.0 kV	90.0 kV
Tube power	3.0 W	90.0 W	60.0 W	1.5 W
Filters used	Thoraesus* ¹	Thoraesus	No Filter	Thoraesus
Exposure time	50.0 ms			
Effective detector pixel size	74.8 μ m			
Source to object distance * ²	431.020 mm			
Source to detector distance * ²	529.000 mm			
Number of projections	3601			
Angular increment	0.1 deg			

Table 4.3: Summary of the acquisition parameters used. *¹ (Thoraesus = Sn 0.1mm, Cu 0.2mm, Al 0.5mm), *² these quantities are based on the motor readings of the Flex-ray scanner which get translated into physical quantities and are subject to alignment errors.

4.1.3 Data acquisition

As described in the section "Experimental design" the data acquisition was done in a semi-automatic fashion. Using our in-house script generator (cf. Section "X-ray computed tomography scanner") we developed a scan protocol that can acquire 50 slices in all three acquisition modes in one continuous session lasting 8h 34mins. The most time consuming processes in this acquisition protocol after acquiring the 3601 projections are the motor movements to change between the acquisition modes since mode 1 and 2 are acquired in the "front position" while mode 3 is acquired in the "mid position". This means that scanning them directly after each other would prolong

the acquisition duration. To ensure that the sample mix does not move noticeably between the different acquisition modes, 10 slices have been scanned consecutively before switching the acquisition modes.

For each acquisition mode and each 10 slice batch a dark-field and flat-field consisting of 100 averaged projections each were acquired for slice 1. Afterwards the 3,601 projections are acquired while the sample stage is rotating continuously. Subsequently, the source and detector move down by 1mm and the next 3,601 projections are acquired. This process is repeated until the 10th slice of the batch, after which also a post-batch flat-field is acquired. Then the acquisition parameters are changed for mode 2 and the above process starts again before they are changed once again for mode 3 and the process repeats one more time. After that the next 10 slice batch is scanned starting again in mode 1. A visualization of the scanning procedure can be found in Figure 4.4. Depending on the available time and/or scanner errors occurring between 10 and 50 slices were acquired per scanning session. In total 5,000 slices were acquired in 111 sessions which lasted between 1h 42mins (10 slices) and 8h 34mins (50 slices).

Additionally to these slices acquired with the standard sample mix and the above mentioned acquisition parameters, the following out-of-distribution (OOD) scans were acquired:

- "pure-sample-ODD": Only one type of sample is mixed with the filler material, scanned in the same way as the standard sample mix.
- "foreign-objects-ODD": A new type of sample not contained in the standard sample mix is added and the resulting new mix is scanned with the standard settings. The foreign objects used are fresh figs, grapes, hazelnuts, peanuts, pistachios, and titanium prostheses screws, cf. Table 4.2.
- "Noise-ODD": the standard sample mix with a tube voltage of 90.0kV, the Thoraeus filter but an even lower tube power of 1.5W compared to the 3.0W used in mode 1. `mode1` contains the 1.5W measurements, `mode2` the usual noisy 3.0W and `mode3` the usual "clean" 90.0W measurements.

Remark. During the acquisition of the dataset the detector broke down. It was exchanged by TESCAN-XRE NV and the FleX-ray scanner has been re-calibrated before resuming the dataset acquisition. Since for every 10 slice batch both dark- and flat-fields are acquired the pixel sensitivities and dark currents of the individual detector should not play a role. Table 4.4 lists which slices have been acquired on which detector and with which sample mix.

4.1.4 Computational processing

The above described data acquisition process yielded a total of 540,195,000 files for the 5,000 standard slices in three modes each. The scanning of every slice produced per mode 3,601 projection data files (images of size 1 x 1912 in TIFF format) and for

every 1st and 10th slice of a 10-slice-batch per mode there are either an additional pre-batch dark- and flat-field or an additional post-batch flat-field, respectively.

Sinogram production

To facilitate using the data collection, the 3,601 projection data files for one slice and mode were combined into one sinogram (image of size 1912 x 3601 in TIFF format) and stored in a folder together with copies of the dark- and flat-fields which belong to the respective 10-slice-batch. The script used for this (`sinogram_production.py`) can be found on GitHub: <https://github.com/mbkiss/2DeteCTcodes>.

Reconstruction production

The sinograms contain the raw photon counts per detector pixel that have to be corrected by offset counts (“dark currents”) via the dark-field (D) and pixel-dependent

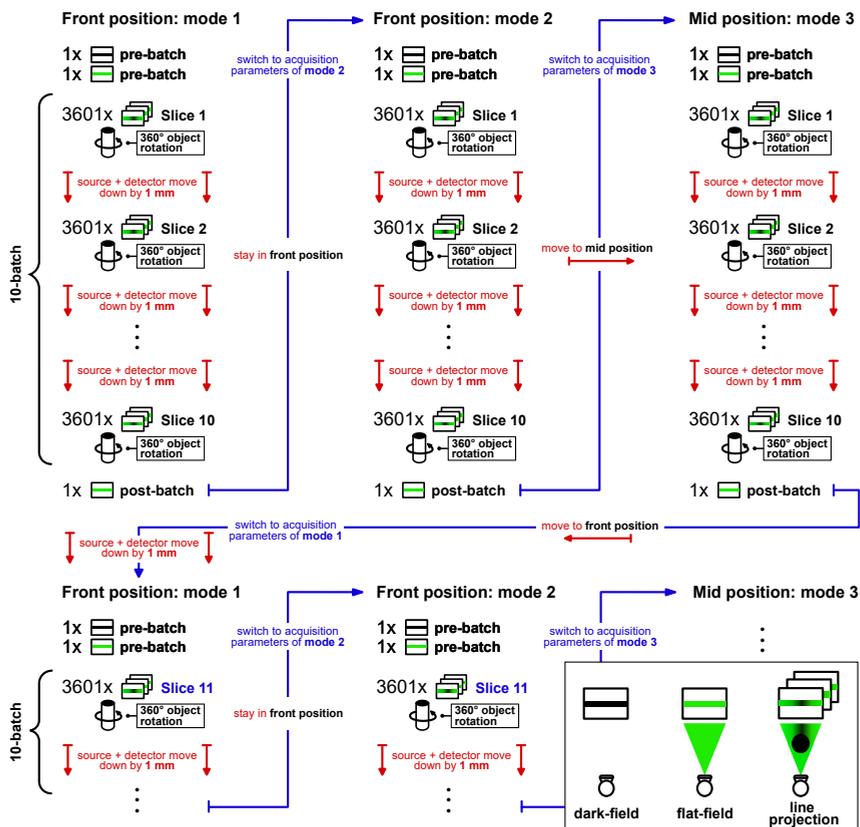


Figure 4.4: Visualization of the scanning procedure.

Sample	Detector 1	Detector 2
Mix 1 (slices 1 - 1800)	slices 1 - 1800	-
Mix 2 (slices 1801 - 3720)	slices 1801 - 2830	slices 2831 - 3720
Mix 3 (slices 3721 - 5000)	-	slices 3721 - 5000
Fig (OOD Pure)	slices 5521 - 5570	-
Almond (OOD Pure)	slices 5571 - 5620	-
Banana (OOD Pure)	slices 5621 - 5670	-
Raisin (OOD Pure)	slices 5671 - 5720	-
Walnut (OOD Pure)	slices 5721 - 5770	-
Coffee beans (OOD Pure)	slices 5771 - 5820	-
Lava stone (OOD Pure)	slices 5821 - 5870	-
Mix 3 (OOD Noise)	-	slices 5871 - 5920
Titanium prostheses screws (OOD Mix 3)	-	slices 5971 - 6070
Peanut (OOD Mix 3)	-	slices 6121 - 6170
Pistachio (OOD Mix 3)	-	slices 6171 - 6220
Hazelnut (OOD Mix 3)	-	slices 6221 - 6270
Grape (OOD Mix 3)	-	slices 6271 - 6320
Fresh fig (OOD Mix 3)	-	slices 6321 - 6370

Table 4.4: Overview of all slices in the dataset assigned to their respective sample mix and used detector.

sensitivities via the flat-fields (F). According to the following formula the combined sinograms (S) can be corrected and converted into a beam intensity loss image (I) following the Beer-Lambert law after applying the negative logarithm to it:

$$y = -\log(I) = -\log\left(\frac{S - D}{F - D}\right) \quad (4.1)$$

The conversion of the sinograms into beam intensity loss images can in some cases yield negative or zero pixel values which were then replaced by the value 1×10^{-6} to ensure strictly positive values as a pre-requisite for the subsequent negative logarithm transform. Although the filtered back-projection (FBP) is a widely used analytical technique to solve the inverse problem of CT reconstruction, noisy and beam-hardening artefact-inflicted measurements yield reconstructions with streaking artifacts. [33]

Therefore, the reconstructions for each slice were obtained by using an iterative reconstruction technique to solve a non-negative least squares (NNLS) problem using 100 iterations of Nesterov accelerated gradient descent [151] with a step size of $\tau = 1/L$, where L is the Lipschitz constant of the forward operator. The forward and backward projection operators were implemented using the CUDA kernels in the ASTRA toolbox. All reference reconstructions were computed from downscaled sinograms ($956 \times 3, 601$), yielding reconstructions of $1, 024 \times 1, 024$ to limit the memory size of the dataset. To reduce the extent of artifacts - especially due to beam hardening in `mode3` - a reconstruction plane of 233.0 mm^2 was used that was resolved by $2, 048^2$ pixels of physical size $113.8 \mu\text{m}^2$. After the reconstruction the images were cropped to the central square of $1, 024 \times 1, 024$. The computation for one reconstruction took about 63s on a NVIDIA GeForce RTX 3070 with 8GB of GDDR6 memory and an Intel Core

i7-10700KF 8-core processor. The total dataset was reconstructed on a gpu-server with 4 NVIDIA GTX 1080Ti (11GB) and 2x Intel Xeon 8-core processor in 83.25 hours. Examples of reconstructed slices are shown in Figure 4.5. The reconstructions for `mode1` display a strong noise due to the limited amount of photons while image slices for `mode2` appear clean, noise- and artifact-free and those for `mode3` exhibit strong beam-hardening artifacts.

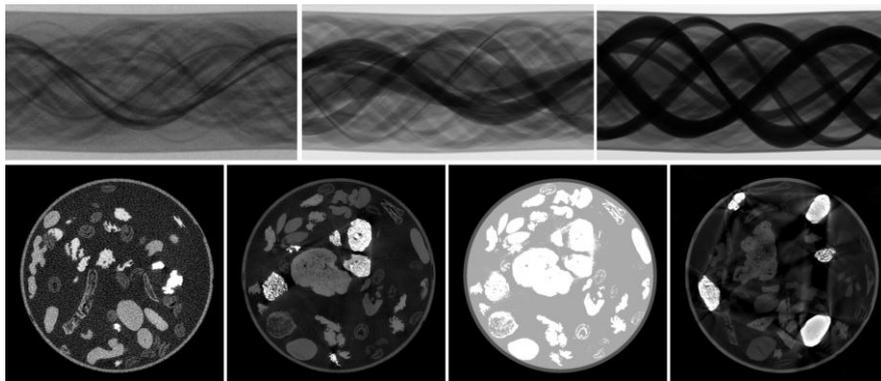


Figure 4.5: From left to right: Sinograms and reconstructions from slice 1661 (`mode1` -low-dose), slice 4300 (`mode2` - artifact-free), slice 560 (`mode3` - artifact-inflicted) and segmentation of slice 4300 based on the `mode2` reconstruction. To illustrate the variation in the sample mix we selected different slices for each mode.

Segmentation production

The reference 4-class segmentation is based on the noise- and artifact-free reference reconstructions of `mode2`. The segmentation distinguishes between background, tube wall, filler material and sample mix objects and is created in six steps. First, a fixed mask for the tube wall is matched to the correct location within the respective slice by maximizing the total sum of the overlap between the mask and the reconstruction via the Nelder-Mead simplex algorithm [65]. Second, everything outside of the tube wall is classified as background. Third, the sample mix objects and tube wall are identified by applying a thresholding algorithm based on average thresholds $[0.00110607, 0.00407358]$ found through a three-class multi-otsu thresholding [129] for some sample slices. Fourth, the sample mix objects are distinguished by removing the tube wall and background from the thresholding segmentation. Fifth, the filler material is identified by removing all other classes from the whole image. Sixth, after checking that all pixels are classified once but not more than that the four-class segmentation is put together and integer values are assigned to the different classes.

The values are: background - 1, tube wall - 2, filler material - 3, sample mix objects - 4. The computation for the segmentation of the total dataset took about 4.25 hours on a NVIDIA GeForce RTX 3070 with 8GB of GDDR6 memory and an Intel Core

i7-10700KF 8-core processor. An example of such a four-class segmentation is shown in Figure 4.5.

4.2 Data records

The 2DeteCT dataset is published as open access on zenodo (<https://zenodo.org>) in 12 repositories. The complete data collection is organized as follows: There are a total of 5,000 slices with standard sample mix and standard acquisition parameters which are split into 10 ZIP archives containing 1,000 slices each. Additionally there are 750 OOD slices which are split into two ZIP archives. To simplify the usage of the data collection the raw projection data are split from the reference reconstructions and segmentations. Each of the raw data archives is ca. 40 GB in memory size and the corresponding reference reconstructions and segmentations amount to ca. 12 GB archives. The separate DOIs for the raw data are as follows: Slices 1 - 1,000 [102], slices 1,001 - 2,000 [104], slices 2,001 - 3,000 [106], slices 3,001 - 4,000 [108], slices 4,001 - 5,000 [110], OOD slices 5521 - 6370 [112]. The reference reconstructions and segmentations can be found under these DOIs: Slices 1 - 1,000 [103], slices 1,001 - 2,000 [105], slices 2,001 - 3,000 [107], slices 3,001 - 4,000 [109], slices 4,001 - 5,000 [111], OOD slices 5521 - 6370 [113]. Each slice folder `slice00001 - slice05000` and `slice05521 - slice06370` contains three folders for each mode: `mode1`, `mode2`, `mode3`. In each of these folders there are the sinogram, the dark-field, and the two flat-fields for the raw data archives, or just the reconstructions and for `mode2` the additional reference segmentation.

- `sinogram.tif` is a 16-bit unsigned integer TIFF file of shape $1,912 \times 3,601$ containing the measured projection data of 3,601 projections acquired in a full rotation of 360° combined into one sinogram (cf. Figure 4.5); size: 13,460KB.
- `dark.tif` is a 16-bit unsigned integer TIFF file of shape $1,912 \times 1$ containing the dark-field measurement; size: 3.82KB.
- `flat1.tif` and `flat2.tif` are 16-bit unsigned integer TIFF files of shape $1,912 \times 1$ containing the flat-field measurements before and after every 10-slice-batch scanned; size: 3.82KB.
- `reconstruction.tif` is a 32-bit floating point TIFF file of shape $1,024 \times 1,024$ containing the NNLS reconstruction computed from the pre-processed sinogram according to 4.1 by 100 iterations of Nesterov accelerated gradient descent (cf. Figure 4.5); size: 4,097KB.
- `segmentation.tif` is 8-bit unsigned integer TIFF file of shape $1,024 \times 1,024$ containing the reference segmentation based on the `mode2` reconstructions (cf. Figure 4.5); size: 1,025KB.

4.3 Technical validation

The FleX-ray scanner is subject to regular maintenance and calibration. Log-files of the scans have been recorded with which we can trace what happened during the scans. Furthermore, the sanity of all collected data was checked via `sinogram_production.py` including that the number of files, their names and their dimensions are correct. Finally, a histogram analysis of all sinograms was performed to ensure that there is no over-saturation present.

4.4 Usage notes

4.4.1 Raw projection data

The projection data for each slice are shared as combined sinograms of 16-bit unsigned integer TIFF files containing the raw photon counts per detector pixel. TIFF files can be interpreted and manipulated by common image visualization software such as ImageJ [177] or scientific computing languages such as MATLAB [139] or Python [191], e.g., through the `imageio` package [115]. In order to be used by tomographic reconstruction algorithms, they typically need to be pre-processed as described above and as shown in the provided scripts.

4.4.2 Reconstructions and segmentations

In general, all reconstructions described in the previous sections can be computed from the projection data with the scripts provided. Depending on the computational resources available this could, however, require a large amount of computing time. Therefore, reference reconstructions are included in the data collection as well. They can be also used as comparison images to test novel reconstruction algorithms, or as ground truths for supervised learning algorithms. Furthermore, they can be used for CT image analysis tasks. Each reconstruction is given by a 32-bit floating point TIFF file.

The reference segmentations of the `mode2` reconstructions can be computed by the users with the provided script `segmentation_production.py` as well but have been included in the uploads as well for the same reasons. Each segmentation is given by an 8-bit unsigned integer TIFF file.

4.4.3 Expansion possibilities

Although the current scope of the 2DeteCT dataset is already offering a lot of versatile applications, the reproducible setup and experimental design as well as the availability of our highly-flexible laboratory X-ray CT scanner enable us to expand the dataset upon reasonable request. Expansion possibilities include among others: First, adding more slices with the same sample mix to increase the size of the data collection to host possible coding challenges. Second, including various new samples

in the sample mix or using an entirely different sample mix for smaller expansions of the data collection. Third, adding more detailed multi-class segmentations to the dataset to train more powerful segmentation algorithms. We encourage the computational imaging community to approach us for suggestions or collaborations on such expansions.

4.4.4 Further usage

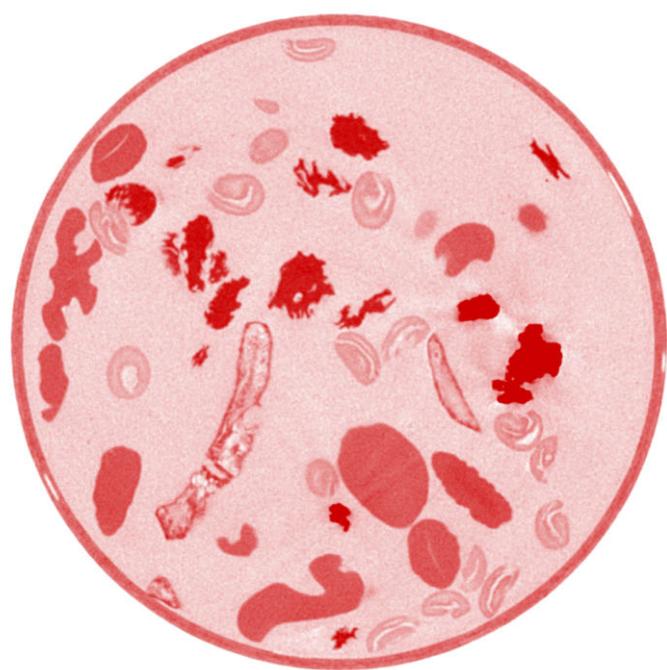
This dataset can be used for developing both classical and ML-based algorithms for a variety of computational imaging applications, including low-dose acquisition, limited or sparse-angle scanning, beam-hardening artifact reduction, super-resolution, region-of-interest tomography or segmentation. With the provided raw projection data and the reference reconstructions in the three different modes, high-fidelity, low-dose, and beam-hardening-inflicted, it is possible to create training data pairs for supervised learning in both the sinogram and the reconstructed image domain. For example, either the A1 (low-dose) or A3 (beam hardening-inflicted) iterative reference reconstructions can be paired with the A2 high-fidelity reference reconstruction (cf. Figure 4.1) to train a denoiser or an artifact-reduction algorithm. For image segmentation the high-fidelity measurements or reference reconstructions of `mode2` and the provided 4-class reference segmentation can be used. Alternatively, the user could use their own multi-class segmentation approach. The provided reconstruction scripts can also easily be modified by the user to represent different scan scenarios: Limited or sparse angle tomography can be experimentally simulated by simply loading only subsets of the raw projection data; Super-resolution experiments can be conducted by either artificially binning the raw projection data into larger pixels, or by binning the reconstructed volumes into larger voxels; Region-of-interest tomography can be achieved by a combination of suitable sinogram sub-sampling and binning. In each of these cases, the provided iterative reconstruction using the full data can be set as a ground truth.

With this dataset we provide a starting point for general algorithm development with experimental X-ray CT data. Due to the design of the dataset, trained machine learning algorithms may generalize to the medical imaging domain, but further research with raw clinical projection data would be needed to demonstrate this. For ethical and regulatory reasons such data is commonly not publicly available for medical CT scanners [207].

4.5 Code availability

Python scripts for loading, pre-processing and reconstructing the projection data in the way described above are published at <https://github.com/mbkiss/2DeteCTcodes>. They make use of the ASTRA toolbox, which is openly available on (www.astra-toolbox.com) or accessible as a conda package (`conda install -c astra-toolbox astra-toolbox`). ASTRA is currently only fully supported for Windows and Linux. Installing it on Mac OS is possible but in the current state very involved and version-dependent.

All reference reconstructions provided have been computed with the Python scripts. Furthermore, while the scripts allow for angular sub-sampling the projections and the reference reconstructions were computed with all projections as mentioned in the subsection "Reconstruction production" above.



Learned denoising with simulated and experimental low-dose CT data

Computed tomography has proven itself as a powerful non-invasive imaging technique in many fields such as materials science, industrial testing, and medicine. It uses X-ray technology to create detailed cross-sectional images of the scanned object using computational methods. Since it uses harmful radiation the imposed dose on objects and patients raises concerns and safety guidelines have been established to minimize radiation exposure [32, 52]. The ALARA principle [140], which stands for "As Low As Reasonably Achievable" advises healthcare providers to use the lowest possible radiation dose necessary to produce high-quality images. However, the minimization of radiation dose through lowering the tube current or exposure time seriously degrades the resulting CT images if no corresponding noise compensation is applied before or during image reconstruction [82, 127]. Noisy images can also occur when there are, for example, constraints on the available time or the number of projection angles. In either setting, it is desirable to reduce the amount of noise through computational methods.

Like in many other research fields, recent developments in computational imaging have focused on developing machine learning (ML) approaches to tackle its main challenges. To improve the performance of algorithms, ML methods are used for different image

This chapter is based on:

M. B. Kiss, A. Biguri, C.-B. Schönlieb, K. J. Batenburg, and F. Lucka. "Learned denoising with simulated and experimental low-dose CT data". *arXiv preprint arXiv:2408.08115* (2024).

processing tasks. These tasks are for example segmentation, artifact removal, or noise reduction.

Generally, these ML methods heavily rely on the availability of high-quality data on which they are trained. When there is a lack of such data, usually existing data is augmented, or new data is generated artificially through simulations. These simulations mimic the problem the ML algorithms shall solve and try to resemble real-world data as good as possible.

The fundamental question arising from this approach is to which extent algorithms trained on simulated data are applicable to real-world experimental data. This work is investigating the performance of noise reduction for two common convolutional neural networks (CNNs). These networks are trained on either simulated or experimental noisy data and are applied to both experimental and simulated noisy data.

Typically, researchers would not have access to raw measurement data because CT manufacturers consider them proprietary. This severely limited both, the analysis of noise simulations but also the performance comparison of algorithms trained on simulated data [207]. The data used in this work are 2D slices of X-ray computed tomography images published in the carefully designed study “2DeteCT – A large 2D experimental, trainable and expandable CT data collection for machine learning” [98]. This experimental data was acquired by the group for Computational Imaging at the Centrum Wiskunde & Informatica and is openly available on zenodo [102–113]. The data collection consists of 5,000 distinct image slices acquired in three different modes. The resulting images are either clean, noisy, or artifact-inflicted.

Using the paired data of clean and noisy images, we create a setting for supervised learning that the CNNs can be trained on for noise reduction. In this work, the clean data is used as a measurement basis to add computationally fast, yet accurate simulated noise. With this data collection and the newly simulated data we have three types of sinograms available: an experimental noisy, an experimental clean and a simulated noisy sinogram. We can pair those sinograms with the clean reconstructed target images to show the difference between training on simulated noisy data and using experimental noisy data for the task of learned denoising.

In this paper, we utilize a large 2D computed tomography dataset for machine learning to carry out for the first time a comprehensive study on the differences between the observed performances of algorithms trained on simulated noisy data and on experimental noisy data. For this we train two common neural networks such as the generic U-Net [173] and the more tailored MSD-Net [160] on both types of noisy data, experimental and simulated. These networks are applied to the data they have been trained on but also to their respective counterparts. The evaluation follows via quantitative metrics in the sinogram and reconstructed image domain as well as qualitative visual inspection in the reconstructed image domain only.

The structure of this study is as follows: After a brief overview of related work in noise modelling and mitigation in the field of computed tomography, we focus on the pre-processing of computed tomography data, previous noise simulation approaches

and how they influenced our choices for simulating noisy training data. In the following subsections we describe the preparation of our training data, the method development, the employed comparison metrics, and how we set up the computational experiments. In the results and discussion section, we present the empirical selection of the noise level for our simulated noisy data and analyse the performance of the differently trained networks applied to the two types of noisy data. We focus on three aspects in our analysis: The choice for the evaluation domain, the influence of the image content, the choice of the training setting.

5.1 Methods

5.1.1 Related work

There is a vast amount of literature investigating the theoretical derivation of accurate noise models for computed tomography images. Generally, they agree that the image noise is directly related to the imaging process and its design criteria such as exposure time, pixels size, slice width, and reconstruction algorithm [58]. Faulkner et al. [58] therefore distinguish between algorithmic and non-algorithmic contributions to noise, and between spatial as well as statistical errors in a CT scan. They note that the statistical noise in the reconstructed images is independent of the number of projections and that the uncertainty is only dependent on the total number of detected photons. Hsieh [83] distinguishes between two principal sources of noise in CT measurement data: quantum noise and electronic noise. Yu et al. [221] showed that the latter usually can be neglected except when the number of detected photons is low and approaches the electronic noise floor. Furthermore, they emphasize that the major difficulty in simulating very-low dose CT measurement data is photon starvation artifacts. These become apparent in reconstructed image slices as ripples or rings in the central region or streaking artifacts between high-density regions. Yu et al. [221] furthermore concluded that their proposed method is not able to simulate images with very-low dose because the photon starvation artifacts are quite complicated. Additionally, they reiterated the call of numerous researchers to get access to raw CT data to allow for testing algorithms for iterative reconstruction and noise reduction [155]. Manufacturers of clinical CT scanners usually introduce nonlinear filters [83] on the measured data to counter beam-hardening and photon starvation artifacts. Therefore, real-world experimental raw data prior to this nonlinear filtering would enable more accurate noise simulations but are usually unavailable.

In practice, it is very challenging to bound the concept of noise in CT image reconstruction from artifacts originating from sources such as sample movement, geometric misalignment, or under-sampling. In this work, we choose to confine our investigations purely to the noise in the sinograms induced by the photon detection in the detector. We note, however, that in reality it is hard to completely disentangle these artifacts and their origins.

5.1.2 Noise simulation

Zainulina et al. [223] concluded in their work that adding noise to the images artificially could bias the predictions of a convolutional neural network (CNN) depending on the accuracy of the noise simulation. This noise simulation requires an in-depth understanding of the actual CT system and might not be feasible at times. The noise in low-dose CT measurement data is influenced by many factors such as the quantum noise, the logarithmic transformation of the measurements, or pre-reconstruction corrections for system calibration which makes modeling the noise in the reconstructed images particularly challenging [130].

Pre-processing A common practice to pre-process the projection data, consisting of raw photon counts per detector pixel, is the so-called dark- and flat-field correction. The dark-fields (D in Equation 5.1) represent the offset counts of the detector system and the flat-fields (F in Equation 5.1) are the values measured when irradiating the detector without an object present between the X-ray source and the detector. These two additional measurements are usually acquired before and/or after the acquisition of the 360° projections and used to remove the dark currents of the detector and to normalize its pixel-dependent sensitivities. With this, the sinograms (S) can then be converted into a beam intensity loss image (ILI) following the Beer-Lambert law after applying the negative logarithm to it according to the formula:

$$y = -\log(ILI) = -\log\left(\frac{S - D}{F - D}\right) \quad (5.1)$$

Such calibrated projection data no longer follows a compound Poisson distribution but is close to a Gaussian distribution with signal-dependent variance [127]. Furthermore, it has been shown that particularly the logarithm operation significantly amplifies the noise when the signal is low [82]. If we want to denoise the low-dose CT measurements before reconstruction, this is best done in the stage of the beam intensity loss image (ILI), so before taking the negative logarithm. If we would denoise the X-ray absorption sinogram (y) instead of the beam intensity loss image (ILI), the application of the negative logarithm would have amplified the noise and changed its distribution.

Based on these findings and considerations we pre-process the raw sinograms of the experimental noisy measurements to beam intensity loss images (ILI) as shown in Equation 5.1 and apply the denoising before taking the negative logarithm. For the preparation of the simulated noisy sinograms we pre-process the sinograms of the high-dose CT measurements of "mode 2" with dark- and flat-field corrections before applying simulated noise to them. With this we prevent our data to experience distributional shifts that might influence the performance of the denoising networks.

Noise simulation approaches To date, there have been proposed several different approaches to simulate the noise in CT measurement data [7, 17, 63, 121, 137, 138, 140, 206–208]. Under the condition that the raw data of a high-dose and low-noise scan is available many studies simulated low-dose and high-noise projection data by

applying synthetic Poisson noise or a combination of synthetic Poisson and Gaussian noise to a high-dose scan [53, 54]. The common models for noise simulation use a relatively simple model of CT acquisition considering a monochromatic X-ray source. This source generates photons that are attenuated by a scanned object and detectors counting surviving photons which are governed by Poisson statistics. More complicated methods range from a detailed characterization of signal statistics of X-ray CT [206–208] over noise equivalent quanta [137, 138] to accounting for energy-integrating detectors [121, 137]. The interested reader may be pointed to the study of Zabic et al. giving a broad overview on the state-of-the-art [222].

To motivate our noise simulation approach we highlight what approaches have been used in practice by previous publications in the field. In particular, there are three noise challenges that have been conducted in the past ten years that have attracted attention to deep learning based denoising. Firstly, the Mayo clinic low-dose CT challenges of 2016 [141] and of 2021 [144] which encompass 30 and 300 patient scans respectively of roughly 70 slices each with noisy reconstruction and projection data simulated from clean reconstructed volumes. Secondly, the LoDoPaB-CT dataset [124] which uses 800 patient scans selected from the LIDC/IDRI database and contains over 40,000 scan slices. Thirdly, the IEEE ICASSP Grand Challenge 8 [20] which also utilizes the LIDC/IDRI database and contains 1010 3D cone-beam CT (CBCT) images. All three noise challenges rely on vendor reconstructed images that subsequently are backprojected to create corresponding projection data / sinograms which then are supplemented with simulated noise. Whereas the first two publications simulate their noisy data only by applying Poisson noise to the projection data, the third generates CBCT projection data with a custom noise simulator that accounts for photon counts, flat-fields, electronic sources, and detector cross-talk as sources of noise. Similar approaches have been undertaken by Bruno de Man et al. from GE research [45, 212] and Jingyan Xu and Benjamin M. W. Tsui [217] and shall be the basis for this work’s noise simulator as well.

Chosen noise simulation approach In this work, we use a simplified version of the noise model used in XCIST [212]:

$$I_i = f_{CONV} \sum_k E_k \cdot \mathcal{P}(DQE_{ik} \cdot (A_{ik} + S_{ik})) + \mathcal{N}(\sigma_{electronic}) \quad (5.2)$$

$$I = \Gamma_{\sigma_{cross-talk}} [I_1, I_2, \dots, I_I]^T \quad (5.3)$$

where i is the pixel index of the detector I , E_k is the energy level with energy index k , A_{ik} are the incident photons in the pixel, S_{ik} the scattered photons in the detector. DQE_{ik} is the detector quantum efficiency and f_{CONV} the energy to electron conversion rate. The noise process is described by \mathcal{P} , a Poisson random generator, and $\mathcal{N}(\sigma_{electronic})$ is a zero mean Gaussian random generator with standard deviation $\sigma_{electronic}$. Finally, $\Gamma_{\sigma_{cross-talk}}$ is a $\mathbb{R}^{D \times D}$ matrix that models detector cross-talk, defined as a fraction of the signal $\sigma_{cross-talk}$ that is shared between adjacent pixels.

This describes a full model of the detector behaviour given incident photons. In XCIST, the incident photons can be simulated by a Monte Carlo particle simulation based on a known source energy spectra and material decomposition of a sample. If the precise behavior of the energy-integrating detector is well understood for each energy level, the parameters f_{CONV} , DQE_{ik} , and E_k can be incorporated. These parameters relate to the conversion of incident photons to measurements. However, for machine learning applications, the physics simulation would demand an unreasonably high computational time (several years for a sufficiently large dataset), necessitating simplifications of the model. In particular, the approximations done in this work assume that the measurement photons were produced by a monochromatic source ($k = 1$) and that there are no scattered photons measured ($S_k = 0$). Additionally, both the detector quantum efficiency of the pixels DQE_k and the photon-to-electron conversion rate f_{CONV} are assumed to be equal to one. As a detector specific calibration of these values is unknown and not easily obtainable without specialized lab equipment. This means that all photons reaching the detector are assumed to be measured and no loss of signal is present. These assumptions of course limit how close the simulation is to reality and following paragraph discusses their effects. The assumption of a monochromatic source is very common in the field of CT reconstruction and is the basis of the most commonly used version of the Radon transform. The effect of this assumption on the noise simulation is that there is no energy-dependent noise being added, but not that there is no noise. The assumption that no scattered photons are measured refers to omitting spatially dependent scatter which can be assumed to be low in comparison to the measured signal. However, the modeled Poisson noise itself still considers intensity-dependent scatter, as all noise in CT comes from photons that "attenuated", i.e. did not follow a straight path. Assumptions on the detector quantum efficiency of the pixels and the photon-to-electron conversion rate as well as the aforementioned two are simplifications that are made to limit the computational complexity of the noise simulation model. Adding noise that accounts for those effects requires is nowadays mainly available via Monte Carlo physics simulators like the mentioned XCIST software that we base our model on. This software has been used by GE Healthcare to validate their models for clinical implementation and can be assumed to be reasonably accurate. However, their computational footprint makes them infeasible to use for ML-sized datasets and requires us to limit the parameters of our noise simulation.

Thus, the chosen noise simulation approach to model the final measurement in the detector I_D is:

$$I_i = \mathcal{P}(A_i) + \mathcal{N}(\sigma_{electronic}) \quad (5.4)$$

$$I = \Gamma_{\sigma_{cross-talk}} [I_1, I_2, \dots, I_I]^T. \quad (5.5)$$

5.1.3 Training data

For the development of a ML-based denoising algorithm the most important element is adequate high-quality training data. In a supervised training framework that means

that there are pairs of input and target data. The algorithm is trained on these data pairs and learns a mapping from the input images to the target images. Zainulina et al. [223] concluded that such supervised deep learning methods show the best performance, but the requirement of paired images may not always be easy to accomplish. For the case of image denoising, this means noisy CT sinograms/reconstructions as an input and noise-free or "clean" CT sinograms/reconstructions as a target data.

Since the publication of the 2DeteCT dataset in 2023 [98], these paired images for supervised learned CT denoising are available. The 2DeteCT dataset comes with pairs of real measurements of the same object, one with near-zero noise, and one with high levels of noise. For this work, we deal with three types of sinograms: experimental clean sinograms that have been measured, experimental noisy sinograms that have been measured, and simulated noisy sinograms (based on the experimental clean data) that were simulated to contain the same level of noise as the experimental noisy sinograms. The corresponding acquisition parameters of this experimental data can be seen in Table 5.1).

In the remainder of this chapter, we will use the term "experimental noisy data" in

Table 5.1: Summary of the acquisition parameters of the 2DeteCT dataset, adapted from [98]. Mode 1 corresponds to the experimental low-dose, high-noise data; Mode 2 corresponds to the experimental high-dose, low-noise data. The Thoraeus filter is a compound filter made of Sn 0.1mm, Cu 0.2mm, Al 0.5mm. The SOD and SDD values are based on the motor readings of the FleX-ray scanner which get translated into physical quantities and are subject to alignment errors.

Acquisition parameter	Mode 1	Mode 2
Tube voltage	90.0 kV	90.0 kV
Tube power	3.0 W	90.0 W
Filters used	Thoraeus	Thoraeus
Exposure time	50.0 ms	
Binned detector pixel size	149.6 μm	
Number of binned detector pixels	956	
Source to object distance (SOD)	431.020 mm	
Source to detector distance (SDD)	529.000 mm	
Number of projections	3601	
Angular increment	0.1 deg	

reference to raw low-dose CT measurement data acquired by a real-world experimental CT system. The term "simulated noisy data" will be used for artificially generated data for which artificial noise was applied to "clean" raw measurement data.

For experimental noisy data, the creation of corresponding image pairs requires a careful acquisition design to avoid that the algorithms would also learn a transformation or change of image content. The exact same CT slice needs to be scanned twice which makes it necessary to change the acquisition settings without infringing with the scanned object. The five main influencing acquisition parameters for the noise level within the CT images have been identified as source current (I), source voltage

(V), exposure time (t), number of projections (n_{proj}), and number of averaged images (n_{avim}) [172]. Overall, the quantum noise in the reconstructed CT images is then inversely proportional to the square root of the number of detected photons. The aforementioned factors have their individual proportional influence on this number, which is given by:
$$\frac{V^{1.3}}{\sqrt{I \times t \times n_{proj} \times n_{avim}}}$$

Analysing this formula, we can determine the relationship between the acquisition parameters and the corresponding noise level in the reconstructed CT slices. Since the used tube voltage V not only influences the noise level but also changes the energy of the used X-ray photons, a change of this factor was omitted. The number of averaged images n_{avim} could not be decreased further than one and since the scanner was already operated close to the shortest possible exposure time t , changing that parameter was also not feasible. To avoid artifacts due to insufficient sampling of the object we did not decrease the number of projections n_{proj} . Therefore, the tube current I was the only feasible option to change and both the noisy and the clean CT scans were acquired with the exact same parameters except for the tube current. For the clean data this was 1000 μA whereas the noisy data had a 30 times smaller tube current of 33.3 μA .

For simulated noisy data creating corresponding image pairs is more straightforward. Given the "clean" data acquisition, a modification of the noise model in Equation 5.4 can be used to simulate artificial noise into the clean image. Given the noise-free incident photons A_i and that the outcome of the Poisson process can be described as an addition $\mathcal{P}(A_i) = A_i + P_i$, where P_i is just the noisy photons, we can rewrite Equation 5.4 for the acquisition of clean data as:

$$I_i^{clean} = A_i + P_i^{clean} + \mathcal{N}(\sigma_{electronic}), \quad (5.6)$$

where the assumption of $P_i^{clean} = 0$ can be made. This is not strictly true, but for a sufficiently large incident photon count A_i it is approximately true. For the noisy acquisition thus the following holds:

$$I_i^{noisy} = A_i + P_i^{noisy} + \mathcal{N}(\sigma_{electronic}). \quad (5.7)$$

$$I_i^{noisy} = I_i^{clean} + P_i^{noisy} \quad (5.8)$$

$$I_D^{noisy} = I_D^{clean} + \Gamma_{\sigma_{cross-talk}} [P_1^{noisy}, P_2^{noisy}, \dots, P_I^{noisy}]^T. \quad (5.9)$$

To appropriately simulate the low-dose measurements I_D^{noisy} , the noise distribution part of the total signal P_i^{noisy} has to be produced, i.e. the Poisson component of the noise. Technically, A_i would be a different number of photons for the clean and noisy images, as the noise mostly arises from the low photon count in our experiments and simulations. However, direct measurement of photon counts is not available and thus direct extraction of this noise from measured data is not possible. Therefore, the noise is parameterized by multiplying the flat-field corrected sinogram $ILLI \in [0, 1]$ (see paragraph "Pre-processing") by a parameters corresponding to the number of photons in vacuum, I_0 , and generating Poisson statistics from its result as

$$P_i^{noisy} = I_0 \cdot ILLI_i^{clean} - \mathcal{P}(I_0 \cdot ILLI_i^{clean}). \quad (5.10)$$

In this model, I_0 is the parameter to control the level of noise added to the clean data, a lower value representing noisier data. Based on the value in the XCIST software [212], a $\sigma_{cross-talk}$ of 5% of the signal is added.

5.1.4 Method development

The noise simulation and the algorithms for learned denoising in this work have been developed in LION (Learned Iterative Optimization Networks)[132], an open-source toolbox for learned tomographic reconstruction implemented in Python. With a designated data loader for the 2DeteCT dataset and with CT experiments set up in a reproducible way it serves as an environment for a standardized comparison of the methods described below.

For the learned denoising algorithms we selected two common convolutional neural networks (CNNs) for image processing tasks that have been used for both natural images but also for computed tomography images in particular: The generic U-Net [173] and the mixed-scale dense neural network (MSD-Net) [160]. The U-Net, originally developed for the segmentation of biomedical images, has been adopted in many fields as a baseline for image reconstruction based on neural networks. The MSD-Net has proven to be particularly effective for computed tomography [125, 159]. Its three main advantages are as follows: First, it has an advanced neural network architecture that uses dilated convolutions instead of traditional scaling operations to learn features at different scales. Second, it uses significantly fewer feature maps and trainable parameters which makes training it less computationally demanding and reduces the risk of over-fitting. Third, it has been applied to denoising large tomographic images and it has been proven that it can be easily applied to similar problems with minimal changes [160].

5.1.5 Comparison metrics

To evaluate the performance of the CNNs trained on either experimental or simulated noisy data, we consider two main comparison cases. In the first case, we test the performance of the algorithms in the setting that they have been trained on, i.e. settings in which they are supposed to work well. This means that if an algorithm is trained for denoising simulated noisy data this situation is used to score their overall performance. The same holds true for algorithms trained on experimental noisy data. For this we compare the output of our learned denoisers to the "clean" target data using the comparison metrics described below. In the second case, we want to compare the performance of the algorithms in settings for which they have not been trained for. This serves the purpose of checking their generalization to other tasks. It answers the question whether the algorithm generalizes to another noise model and its severity. In other words, whether the learned algorithms can also denoise input data without being trained on the specific noise of that data. This is particularly interesting for the case in which the learned denoisers are trained on simulated data and applied to experimental noisy data.

Using these comparison cases, we require comparison metrics with which we can evaluate the performance of the algorithms. Namely, how close the denoised images obtained from these algorithms are to the ground truth images. These metrics have to be able to measure two qualities: How well does the algorithm recover the structure of the imaged object from the noisy data? How well does it restore a good signal with respect to the overall noise in the reconstructed image?

Two commonly used metrics for these tasks are the structural similarity (SSIM) [202] and the peak signal-to-noise ratio (PSNR) [69]. The SSIM is a metric that indicates in a range from 0.0 to 1.0 how similar the compared image is to a ground truth, where 1.0 means they are identical. The PSNR is a metric that calculates the ratio between the highest attainable value of a signal and the strength of corrupting noise that impacts the fidelity of the image. Higher values in both metrics indicate a better algorithm performance. It is worth noting that these two commonly used quantitative metrics, may not be suitable for tomographic reconstruction or scalar fields [30, 31]. In reconstruction tasks such as CT imaging in medicine, PSNR and SSIM do not necessarily reflect a task-dependent better image [73, 136]. Therefore, it is suggested that evaluations consider such downstream tasks of the imaging rather than solely relying on traditional metrics. Additionally, the unbounded nature of CT images poses challenges for metrics like PSNR and SSIM, as the range of pixel values can vary. Different approaches to evaluating reconstruction performance, such as clipping or preserving the result range, can significantly impact reported performance. However, these metrics are still commonly used for a quantitative assessment of images. Since we are interested in measuring performance differences rather than rating the performance itself, they are also used in this work.

In our performance analysis we follow previous work by Zeng et al. [224] who argued that image artifacts due to beam hardening and photon-starvation are particularly difficult to evaluate meaningfully with quantitative metrics in the sinogram domain. They require a visual inspection in the reconstructed image domain. Therefore, we also include a qualitative, visualization-based evaluation between the results of denoised low-dose CT scans in the reconstructed image domain.

5.1.6 Computational experiments

For this work, we first applied the denoising in the projection domain, i.e. denoising beam intensity loss images (*ILI*), for three reasons: i) quality of denoising reconstructed images depends on the used reconstruction method; ii) artifacts caused by the noise in the projection domain are harder to remove after reconstruction; iii) noise in the projection domain is spatially uncorrelated. After evaluating the results of this approach we additionally trained denoising algorithms with an optimization in the reconstruction domain mapping directly from sinogram to reconstruction. For this, we included an FBP reconstruction in the pipeline of the models described below and visualized in Figure 5.1.

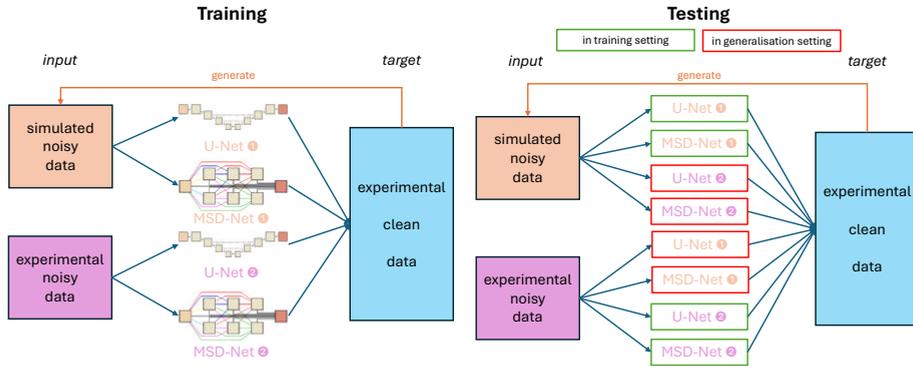


Figure 5.1: Training and testing scenarios for learned denoising networks (U-/MSD-Net illustrations adopted from [160]).

We prepared the training data for our learned denoisers by simulating noisy data from the "clean" experimental measurement data as described in Section "Training data" and using the unchanged clean data as ground truth target data. Consequently, there are two respective image pairs for supervised learning available: First, the simulated noisy data as an input and the experimental clean data as a target. Second, the experimental noisy data as an input and the experimental clean data as a target.

These image pairs were split into $\sim 80\%$ training data (3930 slices), $\sim 10\%$ validation (550 slices) and $\sim 10\%$ testing data (470 slices). Each algorithm was trained for 100 epochs using the Adam optimization algorithm [97]. The final model parameters were selected based on minimal validation loss. The computations were carried out on a GPU-server with 4x RTX 2080Ti (11GB), 384GB RAM, and 2x 16-core Xeon CPUs as well as a GPU-server with 2x RTX A6000 (48GB), 1TB RAM, and 2x 16-core Xeon CPUs.

After the training of the two neural network architectures on the two supervised learning settings, each of the four resulting trained networks was applied to their own test sets but also to the test sets of the data type they have not been trained on. A visual overview of this is given in Figure 5.1.

5.2 Results and discussion

5.2.1 Empirical selection of noise level

For our comprehensive study on the differences between the observed performances of algorithms trained on simulated noisy data and on experimental noisy data it was particularly important to have noise levels in our simulated noisy data that are representative of the noise levels present in our experimental noisy data. Therefore, we tried out various values of I_0 for our noise simulation approach and compared

both the resulting simulated noisy data and the experimental noisy data to the "clean" sinogram data with respect to PSNR and SSIM. Furthermore, the quantitative comparison was also carried out in the reconstruction domain, i.e. comparing the FBP-reconstructed images of the experimental and simulated noisy data to the "clean" reference reconstructions of the 2DeteCT dataset. Observing similar numerical values w.r.t. PSNR and SSIM for both experimental and simulated noisy data we can argue that our noise model generates simulated noisy data with a similar noise level. The results of this comparison can be found in Table 5.2. For all noise levels of the simulated noisy data the SSIM and PSNR values in the sinogram domain are significantly larger than the respective values for the experimental noisy data. A visual comparison of the different noise levels in the sinogram domain proofed uninformative as displayed in Figure 5.2.

Corresponding quantitative and qualitative analyses in the reconstruction domain showed similar image metrics for both the simulated and experimental noisy data. For a noise level of $I_0 = 200$ the PSNR value is closest to the same metric for the experimental noisy data, whereas the SSIM value shows its best agreement for a noise level of $I_0 = 300$. Since the task at hand is learned denoising, we chose to rely on the agreement with respect to the PSNR value and chose a noise level of $I_0 = 200$ for our computational experiments. A qualitative inspection of the images in Figure 5.3 agrees with this parameter choice.

A detailed inspection of Figure 5.3 furthermore showed a strong influence of the attenuation of the objects in each scan on the similarity between reconstructions based on simulated and experimental data. Simulated noisy image slices with no or only small objects with high attenuation (stones) appear to be visually close to the experimental noisy images. However, if those objects are bigger or grouped closely, the experimental noisy images show streaking artifacts caused by beam hardening, not visible in the reconstructions of the simulated noisy data. As previously mentioned, the noise model used in this work assumes mono-energetic photons and consequently cannot capture this behaviour. For the high-dose measurement data, which is the basis for the simulated noisy data, a high enough number of photons is detected and the reconstructed images do not present streaking artifacts due to beam hardening and some level of photon starvation.

5.2.2 Sinogram denoising

The quantitative analysis of the performance of the different CNNs trained on either experimental or simulated noisy data was carried out in both the sinogram and the reconstruction domain (FBP of model output) and is presented in Table 5.3. The evaluation in the sinogram domain shows that for both CNN architectures, U-Net and MSD-Net, the training on simulated noisy data performs better in both application cases, experimental and simulated noisy data. Applying the U-Net trained on experimental noisy data to simulated noisy data performs similarly well whereas the MSD-Net trained on experimental noisy data is not able to generalize well. Applying the U-Net trained on simulated noisy data to experimental noisy data yields a lower

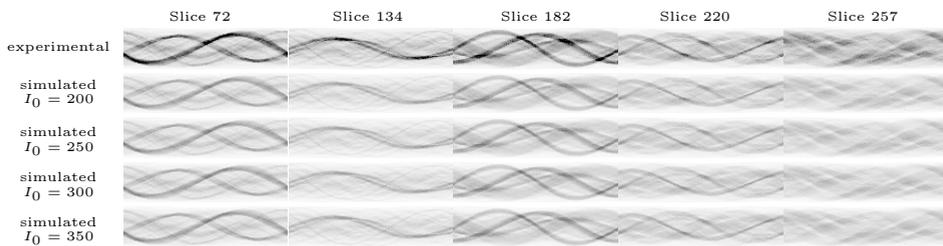


Figure 5.2: Visual comparison of the sinograms of experimental and simulated noisy data with different levels of I_0 (200, 250, 300, 350) from the 2DeteCT dataset for the slices with indices 72, 134, 182, 220, 257.

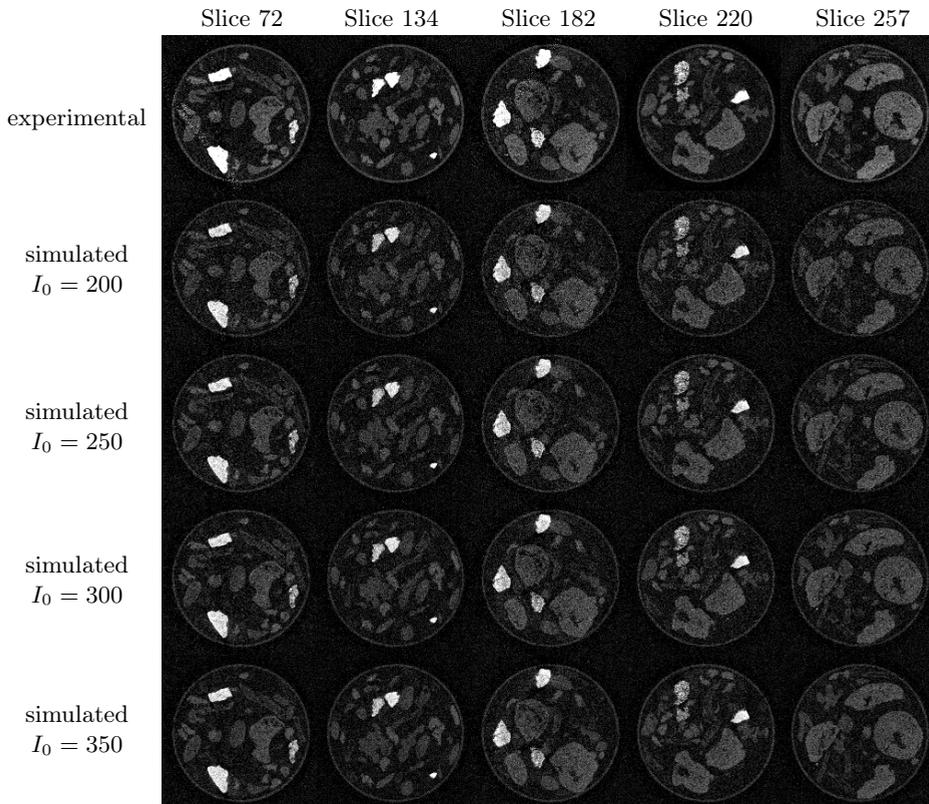


Figure 5.3: Visual comparison of the FBP-reconstructed images of the experimental and simulated noisy data with different levels of I_0 (200, 250, 300, 350) from the 2DeteCT dataset for the slices with indices 72, 134, 182, 220, 257.

Table 5.2: Empirical selection of the appropriate noise level I_0 to generate the simulated noisy training data based on the SSIM and PSNR values of the data with respect to the ground truth (wrt GT) data of "mode 2".

Type of Noisy Data	Noise Level	Evaluation in Sinogram Domain		Evaluation in Reconstruction Domain (FBP of Noisy Data)	
		SSIM (wrt GT)	PSNR (wrt GT)	SSIM (wrt GT)	PSNR (wrt GT)
Experimental Noise	-	0.2658 \pm 0.0963	19.8130 \pm 4.6583	0.1899 \pm 0.0987	21.8024 \pm 3.5996
Simulated Noise	$I_0 = 200$	0.2965 \pm 0.0409	25.7190 \pm 0.8567	0.1364 \pm 0.0307	21.4468 \pm 1.8307
Simulated Noise	$I_0 = 250$	0.3448 \pm 0.0435	26.6607 \pm 0.8634	0.1627 \pm 0.0356	22.3976 \pm 1.8311
Simulated Noise	$I_0 = 300$	0.3854 \pm 0.0451	27.4191 \pm 0.8678	0.1865 \pm 0.0397	23.1659 \pm 1.8317
Simulated Noise	$I_0 = 350$	0.4201 \pm 0.0459	28.0517 \pm 0.8725	0.2083 \pm 0.0432	23.8089 \pm 1.8325

performance of the learned denoiser on the test set images. The overall PSNR is 3dB lower and also the SSIM metric is 0.0522 lower than its application to simulated noisy test set data. For the MSD-Net this gap is even more significant. The MSD-Net trained on simulated noisy data applied to experimental noisy data yields a 21.7057 dB lower PSNR and a 0.1323 lower SSIM on the test set images compared to its application to simulated noisy test set data. This might be due to the much lower number of parameters of the MSD-Net which is not able to capture the experimental noise equally well as the simulated artificial noise.

However, CT reconstruction is an inverse problem that can exacerbate noise from the sinogram during the reconstruction process. Furthermore, applying the required sinogram pre-processing steps changes the nature of the noise model in a complex way. Therefore, evaluating the performance of the denoisers in the reconstruction domain is scientifically more relevant since even small errors in the sinogram domain might be larger in the reconstruction domain. For this reason, Table 5.3 also compares the performance of the sinogram denoisers in the reconstruction domain (FBP of model output).

In there we can observe that the high performance in denoising the sinograms does not carry over to the reconstruction domain. Both the structural similarity and the PSNR in this domain drop substantially. Additionally, the evaluation in the reconstruction domain shows that learned denoising of experimental noisy data performs best if the CNNs are trained on experimental noisy data, as it is expected. Furthermore, the U-Net architecture seems to pick up the image content in terms of structural similarity (SSIM) better than the MSD-Net when trained on experimental noisy data. The PSNR performance is better for the MSD-Net in all training settings except for the case of training on simulated noisy data and testing on experimental noisy data.

After the uninformative visual inspection of the simulated noise in the sinogram domain, and considering that the ultimate goal is to obtain better reconstructed images, the qualitative analysis of the model performances was only carried out in the reconstruction domain which can be found in Figure 5.4. The qualitative visual inspection, also in comparison to the reference images displayed in Figure 5.5, shows that the models for sinogram denoising (found within the first four rows of the

Table 5.3: Quantitative performance analysis with PSNR and SSIM of the differently trained models in the reconstruction domain for the two different testing data with respect to the ground truth data from the iterative reference reconstructions of "mode 2" from the 2DeteCT dataset.

Method	Training Data	Metric	Testing Data	
			Experimental Noisy Data	Simulated Noisy Data
Evaluation in Sinogram Domain				
U-Net ②	Experimental Noisy Data	SSIM	0.8126 ± 0.0194	0.8167 ± 0.0199
		PSNR	18.4966 ± 0.6278	19.3181 ± 0.5575
U-Net ①	Simulated Noisy Data	SSIM	0.8273 ± 0.0240	0.8795 ± 0.0206
		PSNR	33.4602 ± 0.9533	36.6016 ± 0.5616
MSD-Net ②	Experimental Noisy Data	SSIM	0.8613 ± 0.0211	0.8239 ± 0.0216
		PSNR	36.2182 ± 0.7214	20.4747 ± 0.6793
MSD-Net ①	Simulated Noisy Data	SSIM	0.7512 ± 0.0226	0.8835 ± 0.0198
		PSNR	16.3208 ± 1.3965	38.0265 ± 0.7412
Evaluation in Reconstruction Domain (FBP of model output)				
U-Net ②	Experimental Noisy Data	SSIM	0.6134 ± 0.0732	0.6273 ± 0.0717
		PSNR	26.7127 ± 1.9780	27.5290 ± 1.9405
U-Net ①	Simulated Noisy Data	SSIM	0.5504 ± 0.0677	0.6351 ± 0.0713
		PSNR	28.3307 ± 2.0810	32.5568 ± 2.0169
MSD-Net ②	Experimental Noisy Data	SSIM	0.5984 ± 0.0741	0.6152 ± 0.0723
		PSNR	30.9185 ± 1.9707	28.3031 ± 1.9314
MSD-Net ①	Simulated Noisy Data	SSIM	0.3854 ± 0.0469	0.6372 ± 0.0469
		PSNR	11.6366 ± 2.3636	32.6552 ± 2.0173
Evaluation in Reconstruction Domain (model output)				
FBP+U-Net ②	Experimental Noisy Data	SSIM	0.8161 ± 0.0592	0.7466 ± 0.0681
		PSNR	29.8398 ± 2.1834	28.0435 ± 2.0848
FBP+U-Net ①	Simulated Noisy Data	SSIM	0.5957 ± 0.0844	0.6693 ± 0.0820
		PSNR	26.8841 ± 3.4800	28.8134 ± 3.9418
FBP+MSD-Net ②	Experimental Noisy Data	SSIM	0.7829 ± 0.0749	0.7892 ± 0.0731
		PSNR	32.0684 ± 1.9309	32.0704 ± 1.9580
FBP+MSD-Net ①	Simulated Noisy Data	SSIM	0.7615 ± 0.0702	0.8204 ± 0.0567
		PSNR	30.6211 ± 2.0626	33.1053 ± 2.0120

figure) do not produce high-quality reconstructions, particularly regarding fine image features/details. The images exhibit lower noise than the FBP reconstructions of the noisy data directly, but there is a noticeable loss of image sharpness.

5.2.3 Optimization in the reconstruction domain: mapping directly from Sinogram to Reconstruction

Having observed that a good model performance in the sinogram domain does not necessarily carry over to the reconstruction domain we wondered whether training the denoising algorithms with an optimization in the reconstruction domain mapping directly from noisy sinograms to "clean" reconstructions, would prove more effective as well. The model performance w.r.t. clean target reconstructions can be found in the bottom third of Table 5.3 and in the bottom half of Figure 5.4. The relative performance of the networks for the respective combinations of training and testing settings is the same as before, but the results are substantially better. We observe an increase of 0.2027 in the SSIM for the best performing model in the constellation experimental noisy training data and experimental noisy testing data and an increase of 0.1832 in the SSIM for the best performing model in the constellation simulated noisy training data and simulated noisy testing data. Also the performance with respect to the PSNR for each corresponding constellation of training and testing data is better if the models are optimized in the reconstruction domain.

A qualitative analysis of the images in Figure 5.4, also in comparison to the reference images displayed in Figure 5.5, show that the performance drop of training on simulated noisy data but testing on experimental noisy data is more substantial than what the performance metrics would suggest, as these metrics capture global performance rather than local. In all of the slices inspected, this particular train/test case produces the worst images of the quadruplet, for both models. Increased "graininess" permeates the entire image, and the low-intensity objects appear more porous than expected.

5.3 Discussion and conclusions

In this work, we aimed to answer the question to which extent algorithms trained on simulated noisy data are applicable to real-world experimental noisy data. This was achieved through the implementation of a realistic yet computationally efficient simulation method and utilizing less-commonly available raw experimental measurement data.

After tuning the noise simulation to the experimentally measured noise level, our empirical selection of I_0 to set the noise level proved to be an adequate choice both in the qualitative and quantitative assessment (PSNR and SSIM) in the reconstruction domain.

Differences in the simulation were mainly observed in the presence of large or closely grouped high-attenuation samples in the respective image slices, i.e. when beam

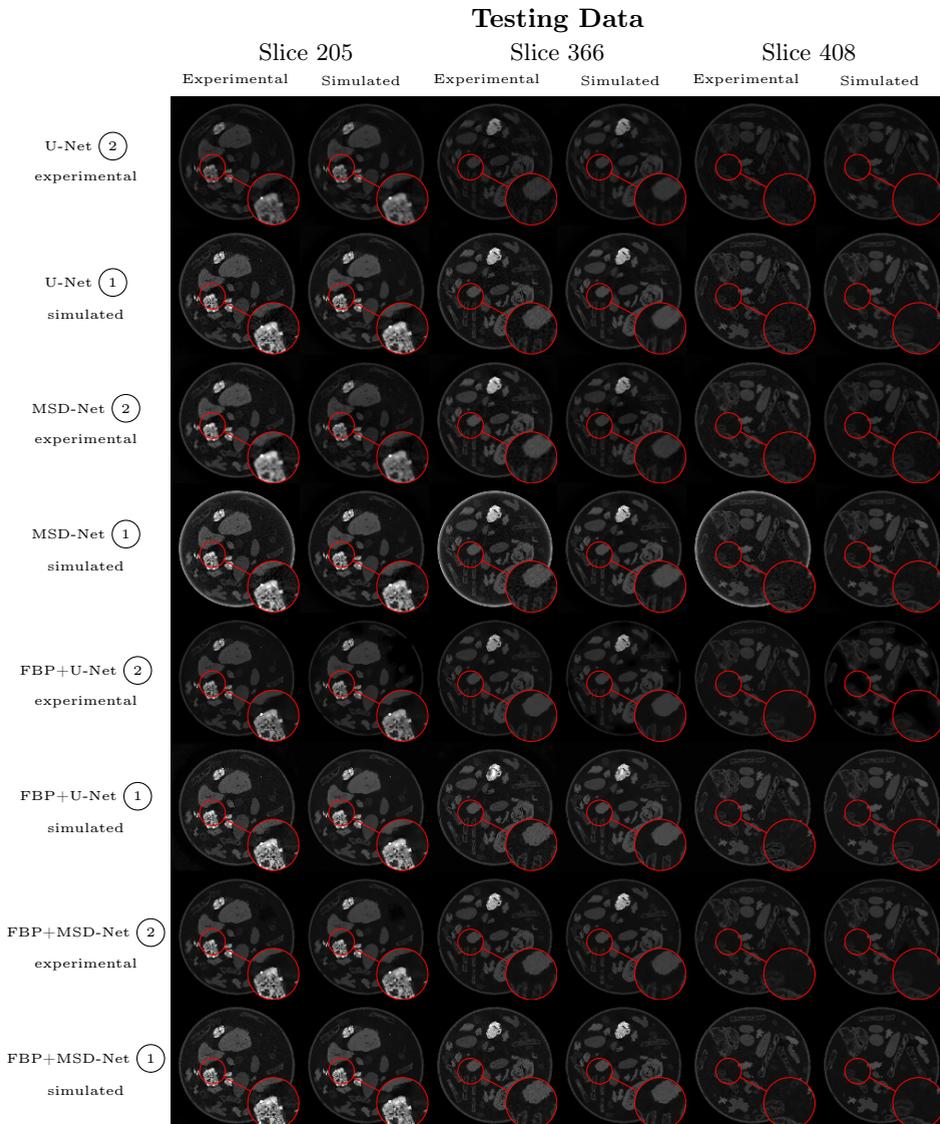


Figure 5.4: Qualitative performance analysis of the differently trained models in the reconstruction domain for the two different testing data (slices indices 205, 366, 408).

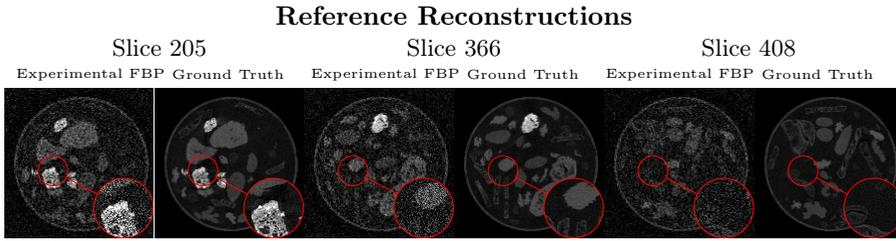


Figure 5.5: Reference reconstructions for the qualitative performance analysis of the differently trained models in the reconstruction domain for the two different testing data (slices indices 205, 366, 408).

5

hardening is present. This is expected, as the chosen noise model for simulation assumes monochromatic sources and thus cannot simulate highly non-linear effects such as beam hardening.

While sinogram denoising achieved better results with simulated noisy data when evaluated in the sinogram domain, the performance did not carry over to the reconstruction domain where training on experimental noisy data showed a higher performance in denoising experimental noisy data. As previously mentioned, this is caused by the inherent ill-posedness of CT reconstruction, that amplifies any remaining noise in the process. Therefore, training the denoising algorithms with an optimization in the reconstruction domain mapping directly from sinogram to reconstruction showed significant improvements in model performance. This is especially noticeable in terms of structural similarity and qualitative visual inspections of the reconstructions. It seems that the artifacts introduced by the FBP reconstruction are not too severe to mitigate via the subsequent post-processing network.

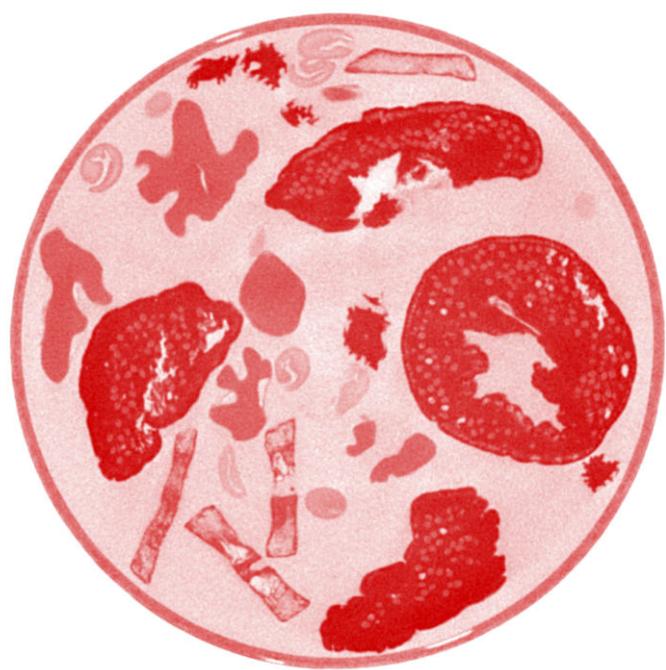
Our findings highlight the importance of carefully designing a noise simulation approach and choosing appropriate noise levels that match experimental data well. If possible the training should be conducted with an optimization in the reconstruction domain, i.e. mapping from raw measurement data to desired target reconstructions. In machine learning for computational imaging, simulated data can be quite different from experimental data, which can impact the transfer of learned systems to the real-world. In particular, the distributions of the training and testing data should be as close as possible and therefore training on experimental noisy data, if available, is preferable when the models are subsequently applied to experimental data. In our experiments, models trained on simulated data exhibit a measurable quantitative performance drop from simulated noisy testing data to experimental noisy testing data. This is even more noticeable by qualitative visual inspection, because these models produce the noisiest images from all the cases.

Ultimately, this research shows that appropriately simulating real noise is important in learned CT research. While computationally fast noise models, like the one presented in this work, will produce data that are close enough to experimental data to make

the models transferable to real-world applications, a drop in performance is expected. Hence, it is advisable to utilize real-world experimental data for training learned denoisers whenever feasible. Furthermore, one should be cautious with, presenting performance outcomes solely based on simple performance metrics when training only on simulated noisy data. As discussed before, our simulation model already captures much of the complexity of the experimental noise in the measurements. However, this work shows that the non-linearity of the imaging process is not captured well enough and that future work should investigate computationally efficient ways of including effects such as beam hardening or photon starvation. Possibly, generative models trained on experimental noisy and "clean" data could solve this challenge or alternatively simplified Monte Carlo particle simulations could be investigated. This study can serve as a starting point for crafting and testing even more sophisticated noise simulation approaches that might be able to close the sim-to-real gap [152, 205] for CT image denoising.

5.4 Code availability

Python scripts for setting up the neural network training as well as the evaluation of the noise reduction performance in the way described above are published on GitHub: https://github.com/CambridgeCIA/LIONscripts/paper_scripts/noise_paper. They make use of the ASTRA toolbox [1, 154, 190], which is openly available on (www.astra-toolbox.com) and tomosipo [81].



Benchmarking Learned Algorithms for Computed Tomography Image Reconstruction Tasks

Computed tomography (CT) is a widely used non-invasive diagnostic method which is applied in various fields such as medicine, materials science, industrial testing, and cultural heritage research. Based on X-ray projection images acquired from 360 degrees, cross-sectional images of an object or patient can be calculated using computer algorithms. For reconstructions of CT data that is, for example, badly sampled (limited- or sparse-angle), low-dose, or exhibits artifacts caused by e.g. metals or other dense materials, typically, a variety of image processing techniques are necessary for quality improvement.

The rise of deep learning [122] and widespread availability of large scale computing systems have led to substantial advances in computer vision, including tasks such as object detection, classification, segmentation or image denoising. A key enabler was the release of corresponding large-scale, open-source datasets such as MNIST [123], CIFAR [119] and ImageNet [46], which helped the research community to hill-climb on standardized benchmarks and continuously advance the state-of-the-art. The last five

This chapter is based on:

M. B. Kiss, A. Biguri, Z. Shumaylov, F. Sherry, K. J. Batenburg, C.-B. Schönlieb, and F. Lucka. “Benchmarking learned algorithms for computed tomography image reconstruction tasks”. *Applied Mathematics for Modern Challenges* 3.0 (2025), pp. 1–43.

years have also seen a rapid development of machine learning approaches specifically for CT image reconstruction [4, 11, 40, 88, 125, 126, 157, 179, 199, 213, 218], which hold great potential for further reducing patient dose, speeding up acquisitions, and improving image quality in challenging acquisition settings [29, 71].

Despite increasing research activity at the intersection of CT and machine learning, the field of CT image reconstruction still lacks large-scale, open-access, real-world datasets that employ standardized evaluation metrics and benchmarking baselines. Many CT studies use datasets that are either not openly available to the research community or largely consist of synthetic data, which may suffer from the broadly observed sim-to-real gap [152, 205]. Furthermore, most of them use different pre-processing pipelines and datasets of various sizes. This hinders the comparison of different state-of-the-art methods and makes reproducing as well as validating results a cumbersome and challenging task.

6 Early computer vision algorithms were often developed using small-scale datasets under lab conditions and showed a significant lack of generalization in the real world. The ability to generalize improved significantly with the emergence of large-scale datasets, which initially consisted of images sourced from the internet and later expanded to encompass increasingly unstructured data. Today, we have access to massive multi-modal (visual and language) datasets scrapped off the internet, enabling foundational models and large language models (LLMs) to achieve unprecedented levels of information processing and synthesis. Accordingly, developing a large-scale benchmarking dataset for 2D computed tomography may be a first step to enabling similar breakthroughs in data-driven CT image reconstruction.

In this paper, we utilize real-world experimental data instead of simulated CT data and design standardized experiments for various common CT reconstruction tasks, allowing for more systematic and standardized comparisons between learned algorithms on one unified dataset. The contributions of this paper are: (i) a benchmarking study for a fixed set of data-driven methods on a recently published dataset of real-world experimental measurements, the 2DeteCT dataset [98]; (ii) a toolbox for benchmarking that enables seamless addition of new methods; (iii) an option to load the 2DeteCT dataset differently within the toolbox for extensions to other problems and different CT reconstruction tasks. This work provides a starting point for the community to develop, test, and compare new methods on real-world experimental data in a straightforward and reproducible way, which can shorten the overall development time of new data-driven CT image reconstruction algorithms considerably.

In the remainder of this paper, we give a brief overview of related work in the field of CT datasets and present details about the mathematical foundation of data-driven CT reconstruction. We provide a short categorization of learning-based methods for solving inverse problems such as CT image reconstruction to give context for the information content of the benchmarking framework. Afterwards, we introduce the benchmarking design, including the various CT image reconstruction tasks, the data pipeline, and the performance metrics. Subsequently, we elaborate on the employed pre-processing of the benchmarking dataset, the evaluated methods of our numerical

experiments, and their training details. After presenting the benchmarking results we discuss the limitations of the dataset, its broader impact, and the code and data availability.

Table 6.1: A summary of publicly available CT datasets, supported tasks, their size, and their raw data availability. (✓) = possible through data generation.

Dataset	CT Image Reconstruction Tasks				Size (>100 samples)	Raw Data
	Low-Dose	Limited-Angle	Sparse-Angle	Beam-hardening reduction		
Mayo [141, 144]	✓	(✓)	(✓)	✗	✗[141] / ✓[144]	✗
LoDoPaB [124]	✓	(✓)	(✓)	✗	✓	✗
ICASSP GC8 [20]	✓	(✓)	(✓)	✗	✓	✗
Walnut CBCT [47]	✗	✓	✓	✗	✗	✓
2DeteCT [98]	✓	✓	✓	✓	✓	✓

6.1 Related work

Computer science has played a vital role in overcoming limitations of traditional imaging systems such as CT and magnetic resonance imaging (MRI). In combination with applied mathematics and advanced engineering, computer science forms the field of computational imaging. The main goal of computational imaging is to improve image quality, enhance resolution, enable novel imaging capabilities, and extract valuable information or hidden details and features that may not be directly visible by traditional imaging methods.

The field has undergone many technological advances throughout the last 15 years [68, 84, 171, 200] but the most recent focus has been on employing machine learning techniques [199]. Despite their clear necessity, the computational imaging field to date offers few large-scale datasets and benchmarks on real-world experimental data. Researchers of NYU and Facebook recently addressed this need for the case of MRI scans by publishing raw measurement data in their fastMRI [116] dataset, while the field of CT still lacked an open-access dataset of comparable scope (cf. Table 6.1). Acquiring such data in the medical sector presents particular challenges, including the radiation exposure patients receive from multiple CT scans and the lack of access to raw measurement data from commercial CT scanners. Previous attempts in the field of low-dose CT, such as the Mayo Clinic low-dose CT challenge of 2016 [141] and 2021 [144], the LoDoPaB dataset [124], and the IEEE ICASSP Grand Challenge 8 [20] have sought to bridge this gap, but relied on simulated data. These issues are further exacerbated due to the lack of raw projection data along with corresponding reconstructed image slices. Although the second release of the Mayo Clinic low-dose CT challenge of 2021 [144] already released raw projection data, the noise for these low-dose datasets remains simulated.

Only recently, the 2DeteCT dataset [98] overcame these shortcomings by providing raw measurement data with complementary features that can be used for a wide range of imaging tasks such as supervised or unsupervised denoising, limited- and sparse-angle scanning, beam-hardening reduction, super-resolution, region-of-interest

tomography or segmentation. In contrast to the clinical, in-vivo datasets such as the LIDC-IDRI [10] or the Mayo Clinic [141, 144], the 2DeteCT dataset would be categorized as an in-vitro dataset that only simulates the behavior of natural tissue.

Having one joint dataset instead of individual datasets of various research groups helps to train algorithms on a uniform set of data, to test them in a standardized way, and to compare them against other algorithms for different imaging tasks. Particularly, the problem of defining a ground truth or “gold standard” is prevalent in CT imaging. Usually, it involves some sort of choice or trade-off with respect to the image acquisition or generation whereas for the 2DeteCT dataset the “mode 2” acquisition provides clean data since its acquisition was designed in a high-resolution setting with an over-sampling in the number of angular projections, a high-dose tube setting, and with a beam filtration in place. Therefore, we treat the reference reconstructions of the 2DeteCT dataset as a ground truth or “gold standard” in this work. They utilize a Nesterov accelerated gradient descent (AGD) algorithm on a bigger reconstruction plane and subsequently crop the resulting reconstructions to their center region. These reconstructions can be used as target images for matching noisy or artifact-inflicted measurements of “mode 1” and “mode 3” respectively and for limited- or sparse-angle measurement data extracted from “mode 2”. A more detailed description of these acquisition modes can be found in section 6.5.1 and a visualization is presented in Figure 6.1. This figure also illustrates types of artifacts present in each reconstruction, highlighting the diverse challenges that data-driven CT reconstruction must address.

6

6.2 Data-driven CT reconstruction

In this section, we present the mathematical background of data-driven CT reconstruction, specifically focusing on 2D tomography, i.e. reconstructing 2D slices from 1D projection data. Furthermore, we briefly introduce how the four classes of methods explored in this work fit within this framework.

6.2.1 Tomographic reconstruction as a linear inverse problem

Tomographic reconstruction is an inverse problem which can be described as an image recovery task based on measurements obtained through the Radon transform: $y(\ell) = \int_{\ell} x(z) dz, \ell \in \mathcal{L}$. In this equation, \mathcal{L} represents the lines in \mathbb{R}^2 from the X-ray source to each detector pixel, defined by the scanner geometry and rotation. Typically, this problem is linearized and discretized as

$$Ax + \tilde{\epsilon} = y \tag{6.1}$$

where A represents the so-called *forward operator* which encapsulates the integral computations over these lines. Here, A is a matrix where each row corresponds to a line integral over the pixel grid of the object. In this context, x is a vector representing the pixel values of the image, y is a vector representing the measured sinogram values, and $\tilde{\epsilon}$ accounts for the noise or error, which may arise from the measurements themselves

or from the linearization of the operator.

Classically, to solve the inverse problem in Eq. 6.1 in a robust manner, a variational regularization approach [55, 176] is employed. The reconstruction is defined by the following minimization problem of the variational objective:

$$\hat{x} = \arg \min_x \{ \mathcal{D}(y, Ax) + \mathcal{R}(x) \}, \quad (6.2)$$

where \mathcal{D} measures the data fidelity between the measurement and the reconstructed image (most commonly the L^2 -distance in CT) and \mathcal{R} is a regularization function that promotes images of desired properties. The data fidelity term \mathcal{D} is usually chosen according to the noise distribution, and a good choice of regularizer \mathcal{R} is important for achieving accurate results. Traditionally, regularization functionals were hand-crafted to encourage the reconstruction x to have structures known to be realistic.

In practice, Eq. 6.2 is solved using iterative optimization schemes, and the quality of reconstructions is largely influenced by the choice of the regularization functional \mathcal{R} . A variety of methods have been proposed in the optimization literature to solve Eq. 6.2 with particular choices for the data fidelity term \mathcal{D} and the regularization term \mathcal{R} , often under the assumption that these functions are convex. In certain cases, these optimization methods yield superior reconstructions compared to the standard analytical approach of the inverse Radon transform, known as filtered backprojection, especially when suitable functions and parameters are selected.

While convexity of \mathcal{R} is analytically desirable for providing efficient optimization schemes with various guarantees, it is often observed that non-convex regularizers yield superior reconstructions in practice. However, this advantage comes at a price: finding global minima becomes generally infeasible, and sometimes even finding stationary points cannot be guaranteed.

6.2.2 Data-driven methods for tomographic reconstruction

In response to the limitations of classical knowledge-driven approaches, data-driven methods have rapidly advanced over the past decades. These methods can be categorized in different ways, including based on the amount of expert knowledge involved, which components are parameterized as neural networks, the domain of application, and the methodological approach employed (cf. Table 6.2).

In this work, we follow the general categorization of supervised learning methods of Arridge et al. [11] and consider the following methods in our benchmarking design: post-processing methods, learned/unrolled iterative methods, learned regularizer methods, plug-and-play methods.

While any strict categorization may overlook or misrepresent certain methodologies from the literature, such as self-supervised learning strategies, this framework effectively encompasses the majority of techniques found in the (weakly) supervised learning literature on data-driven CT reconstructions.

In the following section, we briefly introduce and describe each of the method cate-

Table 6.2: Method Categorizations

Article	Categorization
Zhang et al., 2020 [225]	Amount of expert knowledge involved: hand-crafted, hybrid approaches, mostly learned
Ye et al., 2023 [219]	Point of learned processing: pre-processing, post-processing, and raw-to-image
Ravishankar et al., 2019 [166]	Domain of application: image-domain, hybrid-domain, AUTOMAP [230], sensor-domain
Arridge et al., 2019 [11]	Methodological: post-processing methods, Learned / Unrolled Iterative Methods, Learned Regularizer Methods, Plug-and-Play Methods

gories for historical and methodological context. It is not intended as a state-of-the-art review of methods in the literature.

Direct solvers: Directly learning a reconstruction from measurement y as $\hat{x} = \mathcal{N}_\theta(y)$ has been proposed for MRI in AUTOMAP [230]. Recently, a similar direct method has been proposed for CT [64], but due to the limited success of this approach, we do not explore it further.

Post-processing methods: A straightforward approach to incorporating data-driven methods to overcome the ill-posed nature of CT reconstruction is to initially use classical method for reconstruction and subsequently train a network to learn the mapping from the manifold of inadequate reconstructions to the manifold ground truth reconstructions [75, 78, 88, 91, 228]:

$$\hat{x} = \mathcal{N}_\theta(\mathcal{F}(y)), \quad (6.3)$$

where \mathcal{F} is a reconstruction (e.g. FBP in this work), and \mathcal{N}_θ an appropriately parameterized neural network (NN).

Learned/Unrolled iterative methods: This method category arises from the realization that many iterative solvers, such as FISTA [14], have a resemblance to convolutional neural networks (CNNs) [74]. Notably, Barbu [12] introduced the idea of unrolling iterative methods. By unrolling a handcrafted iterative algorithm and using it as a building block of a deep neural network (DNN), the parameterized mathematical operators inherit hyperparameters, image priors, and data consistency constraints from the iterative methods [219]. A common way to design these methods is to find a particular step of an iterative solver, e.g. a *proximal* step, and replace it with an iteration-dependent shallow CNN. A broader overview of these methods is given in the review by Monga et al. [145].

The resulting unrolled iterative methods have more interpretable architectures compared to traditional “black-box” denoisers, commonly exhibit much fewer trainable parameters than standard DNNs, and enable combining domain knowledge with deep learning [225]. However, it is important to acknowledge that while these methods may

be more interpretable, they lose their mathematical guarantees when incorporating learned networks. Since these methods utilize the operator within the network, they are often referred to as model-based networks. In this context, the physics of the model, represented by the operator A , is provided to the network rather than learned.

Learned regularizer methods: This approach is based on learning the regularization functional \mathcal{R} in Eq. 6.2 from data. Traditionally, the regularization functional was hand-crafted for the problem, and over the past decades many hand-crafted functionals have been proposed; see Benning and Burger [16] for an overview. In contrast to standard DNNs, the minimization of the variational objective can be analyzed mathematically [149]. Examples include dictionary learning [39], generative [203], network Tikhonov [126], and adversarial regularization [183]. See Habring and Holler [76] or Dimakis et al. [48] for an overview.

Plug-and-Play methods: A special sub-case of learned regularizers is the field of Plug-and-Play (PnP) methods where proximal algorithms are used to optimize the inverse problem when either the data fidelity or regularization term is non-smooth. Two widely used iterative algorithms minimizing such composite functionals are the Alternating Direction Method of Multipliers (ADMM) [27] and the FISTA [14] which use proximal operators to avoid differentiating the non-smooth function. The proximal step in these algorithms can be replaced by a more general black-box denoiser (“plugged-in”) while the optimization algorithms run (“play”) as before. This approach of PnP methods was developed by Venkatakrishnan et al. [194] and an overview of theory, algorithms, and applications can be found in a recent review by Kamilov et al. [90]. Although PnP methods are heavily inspired by variational approaches, the study of their properties as convergent regularization methods is an area that is still under active development, with some initial work establishing results in this direction [51, 80].

6.3 Benchmark design

Following best practices on reproducibility for benchmarks [204] we define the purpose and scope of our benchmark as providing the research community with a benchmarking framework based on a real-world experimental dataset under CC BY 4.0 license. The framework consists of a versatile toolbox and a pipeline to evaluate and compare different algorithms. The toolbox can be used to set up reproducible, reusable experiments for different image reconstruction and processing tasks in X-ray computed tomography. The methods selected cover the full range of common categories of supervised learning methods for solving inverse problems. For each category we implement three well-established methods and evaluate their respective performance. The parameters for the CT image reconstruction tasks represent common choices in the field and all methods are implemented in recent Python and PyTorch versions. The evaluation is done with key quantitative performance metrics such as the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR). The results of the different methods are documented in Tables 6.5, 6.6, and 6.7 as well as in a visual overview for a selected image slice in Figure 6.2. All trained models are saved and made

available on GitHub. The LION toolbox <https://github.com/CambridgeCIA/LION/>, used for setting up the benchmarking experiments, enables future extensions by implementing other CT experiments or other ML-based methods. The codebase is open-source and licensed under GNU General Public License v3.0. The aim of the benchmarking design is threefold. Firstly, we give an overview of different categories of data-driven methods for CT image reconstruction. Secondly, we set-up an easy-to-use pipeline for implementing and testing algorithms on real-world experimental data. Thirdly, we provide a baseline comparison of the aforementioned data-driven methods on the most common CT image reconstruction tasks [219].

The benchmark design described in this work is unique in its combination of realism, dataset scale, variety of measurement settings and variety of reconstruction methods considered. Our major contribution is that we have developed a benchmarking framework that for the first time relies completely on real-world experimental data and investigates the whole range of common CT image reconstruction tasks. Existing studies on data-driven CT reconstruction usually focus on one of the method categories and a singular task for which a newly developed algorithm is compared to the most recent state-of-the-art and classical reconstruction methods. Often the data used for these assessments are not the same as the data the other method was tested on, e.g. the acquisition geometry, the sub-sampling, or the pre-processing might differ. All these factors limit their comparability and necessitate a benchmarking design with one common dataset and standardized CT reconstruction tasks as outlined below and visualized in Figure 6.1.

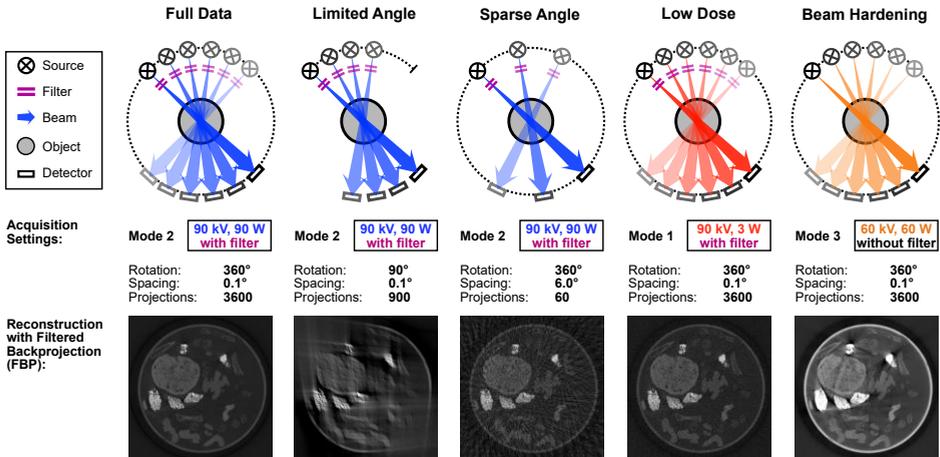


Figure 6.1: CT Image Reconstruction Tasks.

6.3.1 CT Image reconstruction tasks

2DeteCT encompasses raw experimental data of various acquisition modes for different CT image reconstruction tasks unified in one dataset. Defining and setting up CT image reconstruction tasks from the 2DeteCT dataset requires loading subsets of the experimental measurement data and defining the geometry for the reconstructions (see Table 6.3). For this work, we use the 2DeteCT sinograms of shape 956×3600 and reference reconstructions of shape 1024×1024 .

Table 6.3: Summary of the acquisition parameters of the 2DeteCT dataset, adapted from [98]. The Thoraeus filter is a compound filter made of Sn 0.1mm, Cu 0.2mm, Al 0.5mm. The SOD and SDD values are based on the motor readings of the FleX-ray scanner which get translated into physical quantities and are subject to alignment errors.

Acquisition parameter	Mode 1	Mode 2	Mode 3
Tube voltage	90.0 kV	90.0 kV	60.0 kV
Tube power	3.0 W	90.0 W	60.0 W
Filters used	Thoraeus	Thoraeus	No Filter
Exposure time	50.0 ms		
Binned detector pixel size	149.6 μm		
Number of binned detector pixels	956		
Source to object distance (SOD)	431.020 mm		
Source to detector distance (SDD)	529.000 mm		
Number of projections	3601		
Angular increment	0.1 deg		

Each of the tasks utilizes different properties of the 2DeteCT dataset, necessitates a corresponding data pairing and requires a pre-processing which we will discuss in Section 6.4.1. The “gold standards” set as target images for all tasks are the reference reconstructions of “mode 2” of the 2DeteCT dataset. The details of these tasks are described in the following and visualized in Figure 6.1.

Full data reconstruction: We use the complete raw projection data of the “mode 2” acquisition (full sinograms) and use the corresponding iterative reference reconstructions of the 2DeteCT dataset as target images. Since these were generated through an AGD algorithm on a bigger reconstruction plane and subsequently cropping the resulting reconstructions to their center region, the evaluated methods actually learn to mimic the final output of AGD for this task. However, the full data reconstruction serves as a reference for the performance of the evaluated methods on other CT image reconstruction tasks.

Limited-angle reconstruction: We limit the raw projection data of the “mode 2” acquisition to a smaller angular range. Depending on the used wedge of 120° , 90° , or 60° only the first 1200, 900 or 600 projection lines of the 956×3600 sinograms are extracted. With this missing information a standard reconstruction will show streaking, elongation, ghost tail, and missing boundaries artifacts of increasing severity with decreasing angular size of the wedge [13, 62]. The “gold standard” is a reconstruction of a complete sinogram and therefore we use the iterative reference reconstructions of the 2DeteCT dataset as target images.

Sparse-angle reconstruction: We sub-sample the raw projection data of the “mode 2” acquisition. Depending on the number of angles used 360, 120, or 60 projection lines evenly distributed over the full angular range are extracted from the original 3600 projections of the sinograms. Undersampling of this kind violates the Nyquist-Shannon sampling theorem and introduces aliasing artifacts [89]. The “gold standard” is a reconstruction of a fully-sampled sinogram and therefore we use the iterative reference reconstructions of the 2DeteCT dataset as target images.

Low-dose reconstruction: We use the complete raw projection data of the “mode 1” acquisition (full sinograms). Due to the low-dose setting of the acquisition, the measurements show very low photon counts and their corresponding iterative reconstruction slices show streaking artifacts and high granularity as manifestations of image noise. The “gold standard” for this task is a reconstruction of the corresponding high-dose acquisition of “mode 2” and we use the iterative reference reconstructions of the 2DeteCT dataset as target images.

Beam-hardening corrected reconstruction: We use the complete raw projection data of the “mode 3” acquisition (full sinograms). Due to the unfiltered beam spectrum of the X-ray source during the acquisition of the measurements, the corresponding iterative reconstruction slices show streaking and shadowing-like “cupping” artifacts as manifestations of beam-hardening and photon starvation [35, 158]. The “gold standard” for this task is a reconstruction of the corresponding filtered acquisition of “mode 2” and we use the iterative reference reconstructions of the 2DeteCT dataset as target images.

6.3.2 Pipeline

The benchmark framework was set up using LION (Learned Iterative Optimization Networks), an open-source Python toolbox for learned tomographic reconstruction. There are other open-source libraries that focus on providing a robust set of forward (and backward) operators, like ODL [2] or tomosipo [81] or libraries that focus on providing variational (non-ML) reconstructions, such as ASTRA [1, 154, 190] or TIGRE [18, 19]. Given such existing libraries, obtaining an operator for the experiments described above, can be straightforward, but creating such operators for real measured datasets is not always trivial and loading this data accordingly can be challenging.

LION focuses on using libraries such as tomosipo to build a toolbox for loading, pre-processing, simulating, and reconstructing CT data as well as training and evaluating data-driven methods in CT. As forward operators, LION currently supports the aforementioned tomosipo, which uses ASTRA’s ray-driven forward projector and an unmatched voxel-driven backprojector.

To our knowledge there is only one other maintained open-source library for deep learning that supports CT reconstruction, DeepInv [187]. DeepInv is a comprehensive library for inverse problems and learning. However, it does have limitations in terms of specificity, which is often overlooked by more generalist libraries. For instance, it does

not provide complex dataset definitions, topic-specific experiments, or application specific noise models that are crucial for certain applications.

Through this benchmarking study LION now features a designated data loader for the 2DeteCT dataset [98] and defines tasks and modes such that the data loader sources the corresponding data from this dataset. It performs a split into training (79.4%, 3930 slices), validation (11.11%, 550 slices), and test data (9.49%, 470 slices), defines the CT geometry and parameters of CT data processing, creates a forward operator, and loads and pre-processes sinograms and reconstructions. Furthermore, it contains designated experiment classes for the above CT image reconstruction tasks and PyTorch implementations of various models from the different method categories outlined above. A significant development effort was undertaken to ensure a consistent implementation of all deep learning methods using the same data and operator framework for the variety of CT image reconstruction experiments. Additionally, the toolbox defines metrics for training and evaluation, contains an optimizer for supervised learning settings and allows for saving all relevant information in a parameter file to completely reproduce models. These parameter files store among others, the used dataset parameters, training parameters, loss, epochs, optimizer, and the CT geometry. Lastly, the toolbox allows for saving trained models to compare against them and for storing scripts that have been used in papers to specifically reproduce experiments of that particular study.

To summarize, the motivation for this benchmarking design is providing the community with an easy pipeline to load real-world experimental data and conduct standardized experiments. It lays the foundation for the computational imaging community to easily implement and test methods on the 2DeteCT dataset using a custom data loader as well as tailored and standardized CT benchmarking experiments. This greatly extends the utility of the 2DeteCT dataset since researchers do not need to spend time on the implementation of their own data loaders or reconstruction tasks and can easily compare against other methods. It makes it easier than ever to start experimenting with deep-learning-based (and non-deep-learning-based) CT reconstruction in realistic settings, without the need for expert knowledge or simulating data. In particular, it allows users to avoid many of the pitfalls of trying to simulate appropriate measurement data and focus instead on the development of reconstruction methods.

6.3.3 Performance metrics

One common metric for evaluating CT reconstructions, especially in the case of limited or noisy data, is the peak signal-to-noise ratio (PSNR)[69]. It quantifies the ratio of the maximum possible value of a signal to the power of corrupting noise that affects the fidelity of the image. Furthermore, the structural similarity (SSIM) [180, 195, 202] indicates in a range from 0.0 to 1.0 how similar an evaluated image is to a reference image, where 1.0 means they are identical. For both metrics, higher scores indicate a better algorithm performance and a ground truth reference image is necessary. The ground truth reference images used in this work are the reference reconstructions

of “mode 2” of the 2DeteCT dataset which utilize an AGD algorithm on a bigger reconstruction plane which is then cropped to its center.

6.4 Numerical experiments

6.4.1 Pre-processing

All numerical experiments for the different CT image reconstruction tasks are set up as sinogram-to-reconstruction experiments. They do not perform sinogram-to-sinogram or reconstruction-to-reconstruction experiments such as sinogram denoising, artifact reduction, inpainting, beam-hardening reduction. This means that the ML-based algorithms take a sinogram as input data and corresponding iterative reference reconstructions from acquisition “mode 2” as target data. In principle, it is possible to also perform sinogram-to-sinogram or reconstruction-to-reconstruction experiments within the LION toolbox but this would make the comparison between e.g. post-processing methods and learned/unrolled iterative methods less fair. Therefore, we chose to only use sinogram-to-reconstruction experiments in this benchmarking study. The sinograms are pre-processed with LION using modules such as ASTRA [1, 154, 190] and tomosipo [81] according to the description in the original dataset publication and as outlined below.

The sinograms are pre-processed into a beam intensity loss image by subtracting detector offset counts (“dark currents”) from the measured photon counts per detector pixel and by dividing by so-called “flat fields”, the pixel-dependent sensitivities of the detector. To perform a CT reconstruction, the data is then transformed with the negative logarithm to follow the Beer-Lambert law. For more details please refer to the original dataset publication [98].

6.4.2 Evaluated methods

The method selection in this work focuses on (weakly) supervised learning methods, excluding self-supervised and unsupervised approaches to establish a foundation for benchmarking. To this end, we prioritize established supervised learning methods that can serve as reliable baselines, omitting some newer techniques based on transformers and generative models, as explained at the end of this subsection. In the following, we introduce which methods from the literature will be used in this benchmark. To limit the broad scope of this work, we use three methods from each subclass presented in Section 6.2.2.

Table 6.4 lists the different methods that have been evaluated for the benchmarking and in which of the described method categories they fall. The details of the networks’ training can be found in Section 6.4.3 and in the GitHub repository mentioned in the section “Code and Data Availability”. Additionally, we evaluate classical reconstruction methods on the test data. Analytical methods such as filtered backprojection (FBP) or iterative methods such as AGD [151] or regularized methods such as the Chambolle-

Pock (PDHG) [37] solver with total variation (TV) regularization are highly effective and still widely used in practice [15].

Table 6.4: Evaluated methods

Category	Method (Year and Reference)
Classical Methods	FBP [79], AGD [151], PDHG [37]
Post-Processing Methods	U-Net [173], MSD-Net [160], DnCNN [227]
Learned / Unrolled Iterative Methods	Learned Gradient [4], TV-regularized Learned Gradient, Learned Primal Dual [3]
Learned Regularizer Methods	AR [134], TDV [117], ACR [147, 148]
Plug-and-Play Methods	DnCNN-PnP [227], DRUNet-PnP [226], GS-PnP [85]

Post-processing methods: In this work, the evaluated post-processing methods are the U-Net [173], the MSD-Net [160], and the DnCNN [227]. The U-Net, originally designed for image segmentation tasks, is well known in many fields of machine learning and consists of a contracting path, which captures both low and high-frequency features, a bottleneck layer, and a symmetric expanding path. The expanding path integrates information from corresponding layers in the contracting path, effectively translating learned features back into the image space at each resolution. The MSD-Net has proven itself particularly effective for CT image reconstruction problems in the literature. Its neural network architecture, incorporating dense connections between layers at different scales, helps to effectively capture both local and global information in images. The DnCNN is a deep learning architecture specifically designed for image denoising tasks. It uses a series of convolutional layers with batch normalization to learn noise patterns and remove them from images. It has achieved state-of-the-art results in image denoising benchmarks.

Learned / Unrolled iterative methods: The first unrolled method we consider is Learned Gradient (LG) [4], which seeks to directly learn the update step in the gradient descent solver rather than relying solely on an additive step. The LG method parameterizes a fixed number of gradient steps to approximate the optimal direction based on the true gradient of the variational objective. The second method is an extension of this model including a TV regularization in its update rule (LGTV). In both of these methods, a small four-layer CNN is employed to replace the update step, taking the current image estimate and gradient(s) as inputs. The last method considered is the Learned Primal Dual (LPD) algorithm [3]. LPD solves the variational optimization problem by learning proximal-like steps in both primal and dual variables simultaneously. By jointly training two networks to update primal and dual variables, this algorithm can efficiently solve tasks such as image reconstruction and denoising and is known to produce high-quality results, using a minimal set of learned parameters. In LPD, each gradient step is substituted by a shallow four-layer CNN.

Learned regularizer methods: Learned regularization methods, which directly parameterize the regularization functional using a neural network, typically differ in either their training strategy, network architecture, or variational objective optimization scheme. For evaluation, the adversarial regularizer (AR)[134] and its convex coun-

terpart (ACR)[147, 148] are considered alongside total deep variation (TDV)[117]. Both adversarial regularizers are trained using a Wasserstein-1 distance-based loss, while TDV is trained by minimizing the distance between the ground truth and the reconstruction achieved via a fixed number of gradient steps on the variational objective. AR is parameterized using a standard CNN with a single dense layer, and the variational objective is optimized via accelerated gradient descent with early stopping. ACR utilizes an input convex neural network, and the variational objective is optimized with accelerated gradient descent and backtracking. TDV is parameterized using a multiscale convolutional neural network.

Plug-and-Play methods: There are various axes along which the settings of PnP methods can be varied, including the choice of splitting method and the architecture of the denoiser. In this work, we will fix the splitting to be a forward-backward splitting of a variational objective, and consider the effect of varying the denoiser architectures: two of them will be “unconstrained”, differing mainly in model capacity, while the last one has a structural constraint that allows for provable convergence. To be more specific, the first method (DnCNN-PnP) replaces the proximal operator of the regularization functional by DnCNN [227], while the second method (DRUNet-PnP) uses a DRUNet [226] instead, which gives improved denoising performance at the cost of significantly more parameters and increased computational time. Finally, the third method (GS-PnP) splits the variational objective in the opposite way, taking a gradient step on the regularization functional and a proximal step on the data discrepancy functional. It has been shown that it is possible to obtain high-quality PnP reconstructions in this way, while retaining the interpretation of minimizing a variational objective [85]. To compute the output of the denoiser in GS-PnP, it is necessary to perform an intermediate backpropagation on the backbone denoiser, resulting in significant extra computational cost, both in terms of memory and time. As in the work of Hurault, Leclaire, and Papadakis [85], we deal with this by scaling down (both in number of blocks and width of the blocks) the backbone DRUNet, as compared to the DRUNet used in DRUNet-PnP.

The field of deep learning methods is rapidly evolving, with new architectures and methods constantly being released. For this reason, we necessarily have had to omit some methods from consideration in this benchmark such as the most recent strides using transformers and diffusion models for CT reconstruction.

In the context of CT reconstruction, transformer-based reconstructions [198] generally take the form of what we have called a post-processing method in the benchmark, the only difference being the architecture of the “denoiser” used. In our benchmark, all of the architectures considered were convolutional. On the other hand, diffusion-model-based approaches [133, 186] are most similar to what we have called plug-and-play methods, as they alternately implement data-consistency steps, utilizing the forward model, and prior-consistency steps. In deterministic plug-and-play the latter involves applying a denoiser or executing a sampling step in stochastic restoration based on diffusion models.

In summary, we believe that our selection of method classes encompasses many methods of interest, even if some specific methods mentioned are not included in the presented comparison of twelve exemplary well-established approaches.

6.4.3 Training details

The basis for this benchmarking framework is the 2DeteCT dataset [98]. This dataset was split in a sophisticated way to ensure that no scanned sample mixes are shared between the training, validation, and test data. The data split is as follows: training data (79.4%, 3,930 slices), validation data (11.11%, 550 slices), test data (9.49%, 470 slices). The training was carried out without extensive hyperparameter tuning to achieve as good a result as possible. We prioritized adequate performance over extensive hyperparameter tuning to produce baseline results for a comparative analysis among techniques. The reported total training times for each method vary significantly, reflecting the differences in data size associated with each CT image reconstruction task.

Post-processing methods

The post-processing methods have all been trained with the same parameters: Adam optimizer [97] for 100 epochs with a learning rate of 10^{-4} and parameters $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The final models were chosen based on the minimum loss in the validation set. The total training times are dependent on the used machine, the CT image reconstruction task and the evaluated method:

- FBP+U-Net, total training time per CT image reconstruction task ranges between 26 – 73 hours,
- FBP+MSD-Net, total training time per CT image reconstruction task ranges between 83 – 120 hours,
- FBP+DnCNN, total training time per CT image reconstruction task ranges between 56 – 93 hours,

Learned / Unrolled iterative methods

These methods have been trained exactly the same way as the post-processing methods, however, LG and LGTV required a learning rate of 10^{-5} for stable training. The final models were again chosen based on the minimum loss in the validation set. The total training times are dependent on the used machine, the CT image reconstruction task and the evaluated method:

- LG, total training time per CT image reconstruction task ranges between 24 – 117 hours,
- LGTV, total training time per CT image reconstruction task ranges between 19 – 116 hours,

- LPD, total training time per CT image reconstruction task ranges between 23 – 153 hours,

Learned regularizer methods

All models were trained using an Adam optimizer with a learning rate of 10^{-4} and parameters $\beta_1 = 0.9$ and $\beta_2 = 0.99$. These hyperparameters were picked according to the original methods [117, 148], and were fixed for all tasks. Order of magnitude for the hyperparameters in minimization of the variational objective were found for one task, the sparse-angle reconstruction with 90 projections, and were adjusted for other tasks based on the operator norm. The adversarial regularization methods have been trained for 25 epochs, with a reduced validation set. Minimum validation loss model was then chosen. TDV was trained for 10 epochs, due to its computationally expensive training. The number of steps was chosen to be the maximal number allowing for the network to fit on a 24 GB GPU. Due to the relatively low SSIM numbers we hypothesize that the number of steps ideally would need to be increased for all experiments, but due to the sizes, remains infeasible. The total training times are dependent on the used machine, the CT image reconstruction task and the evaluated method:

- AR, total training time per CT image reconstruction task ranges between 50 – 75 hours,
- ACR, total training time per CT image reconstruction task ranges between 50 – 75 hours,
- TDV, total training time per CT image reconstruction task ranges between 90 – 110 hours,

Plug-and-Play methods

As above, we trained the denoisers for the PnP methods using the Adam optimizer, with a learning rate of 10^{-4} and parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We normalized the inputs by rescaling with the maximum pixel value found on the training set. We trained denoisers on Gaussian denoising tasks with a range of noise levels $\{0.001, 0.005, 0.01, 0.02, 0.03, 0.05, 0.07\}$, corresponding to (average) PSNRs of approximately 50 dB, 36 dB, 30 dB, 24 dB, 20 dB, 17 dB and 13 dB respectively. After training, we plugged the denoisers into the methods described in Section 6.4.2, selecting separately for each experiment the denoiser that performed best on the PnP reconstruction task on the validation set. Each denoiser was trained for 25 epochs, with total training times per denoiser being as follows:

- DnCNN, total training time 6 hours,
- DRUNet, total training time 10 hours,
- GS-DRUNet, total training time 15 hours.

6.5 Results and discussion

We report on the performances of different data-driven methods on the most common CT image reconstruction tasks in both a quantitative (see Tables 6.5, 6.6, and 6.7) and qualitative analysis (see Figure 6.2). The metrics are averaged over the whole test dataset and include their standard deviation whereas the qualitative analysis is for one specific slice (index 182) of the test dataset. While an extensive quantitative analysis of the performance describing trends in different performances is out of the scope of this paper, we provide a detailed but short quantitative and qualitative analysis below. In CT image reconstruction the qualitative analysis, i.e. the visual inspection of reconstructed images, is a crucial tool to augment the quantitative results reported in Tables 6.5, 6.6, and 6.7 and shall serve as a starting point for further analysis.

6.5.1 Relevance and difficulty of CT image reconstruction tasks

The basis of the benchmarking framework of this work are the selected CT image reconstruction tasks. Namely, Full Data, Limited-Angle, Sparse-Angle, Low-Dose and Beam-Hardening corrected reconstruction. In the following, we want to give a better insight about their respective relevance and difficulty.

Full data reconstruction: The Full Data CT image reconstruction task can be considered purely as a reference for each of the algorithms and does not pose any particular challenges. This is due to the fact that this task uses the full data of the “mode 2” acquisition of the 2DeteCT dataset which was designed in a high-resolution setting with an over-sampling in the number of angular projections, a high-dose tube configuration, and with a beam filtration in place. Classical methods such as FBP, AGD, and PDHG will perform well in these settings and there is no need for learned reconstruction methods.

Limited-angle reconstruction: If this data is limited or sparsified in the angular range, undersampling artifacts occur. Theoretically, acquisitions from 180° with sufficient angular sampling can produce an artifact-free image. When limiting the angular range further, the missing information causes more visible image artifacts such as streaking, elongation, ghost tail, and missing boundaries. The challenge of limited angle acquisition occurs for example in industrial product inspection and medical imaging mammography. During the selection of the angular span of the wedge, we tested a limited angle of 150° for a few algorithms and decided that the task at hand is not yet challenging enough. Therefore, we chose wedges of 120° , 90° , or 60° for our limited angle reconstruction tasks. Since classical reconstructions of 60° are already dominated by artifacts, a further undersampling was omitted. The difficulty of these tasks increases with decreasing available angular range.

Sparse-angle reconstruction: For undersampling in terms of sparsity, the Nyquist-Shannon sampling theorem [89] can give an approximation of how many projections are necessary for a well-sampled CT scan. For the experimental setup of the 2DeteCT dataset a minimal number of approximately 3000 projections is required for sufficient

sampling. Noticeable differences, however, only occur when undersampling by factors of five or more. Decreasing the number of projections is often used to speed up the CT acquisition process or to reduce dose for the scanned subject or sample. We tested undersamplings based on 720, 360, 180, 120, 90, and 60 projections for one classical, one post-processing, and one unrolled algorithm to decide which experiments to include in the benchmark. Since 720 and 360 projections both showed a relatively similar severity in artifacts we chose to include only the 360 projections case in our benchmarking. For the lower end of this undersampling range we concluded that 60 projections, so an undersampling of a factor 60 in comparison to the full data, was still feasible for some of the tested learned algorithms and should be considered as a challenging task for our benchmarking. To distribute the number of projections for our sparse-angle reconstruction we chose 360, 120, and 60 projections for our final experiments. Again, the difficulty of these tasks increases with decreasing the number of available projections.

6

Low-dose reconstruction: For the low-dose CT image reconstruction task, we use the acquisition data of “mode 1” which uses a $1/30$ tube current compared to “mode 2” acquisition that has been optimized for the best image quality. In medical imaging, a lower dose is typically chosen to achieve images of adequate quality for clinical purposes while minimizing radiation exposure. The “tube current”-“exposure time” products range from 50 to 400 mAs in clinical practice. “Mode 2” has a high-dose acquisition with a “tube current”-“exposure time” product of 18 mAs, while “mode 1” has a low-dose acquisition with a product of 0.6 mAs. It is important to note that the size of the scanning object and the setup geometry differ from a traditional medical CT scan, as the scanning object is much smaller with a circumference of approximately 35 cm compared to 50-100 cm for standard abdominal circumferences in children and adolescents [168]. Nevertheless, the low-dose CT reconstruction task can be viewed as having a similar or even higher noise-level than medical (extreme) low-dose CT data.

Beam-hardening corrected reconstruction: For the beam-hardening corrected CT image reconstruction task, we use the acquisition data of “mode 3”. Since the beam spectrum of the X-ray source remains unfiltered during the acquisition process, the corresponding iterative reconstruction slices show streaking and shadowing-like “cupping” artifacts as manifestations of beam-hardening and photon starvation [35, 158]. These artifacts are of a highly non-local and non-linear nature and corresponding data of beam-hardening afflicted CT images and physically filtered and corrected acquisition data are a novelty introduced by the 2DeteCT dataset [98]. In both medical and industrial settings, having high-attenuating areas in the region-of-interest causes severe artifacts and poses challenges. The severity of the challenge of learning to map between these data distributions was to date unknown and is reported in this work for the first time.

Overall, the limited-angle reconstruction from 60° and the beam-hardening corrected reconstruction can be considered the most difficult tasks.

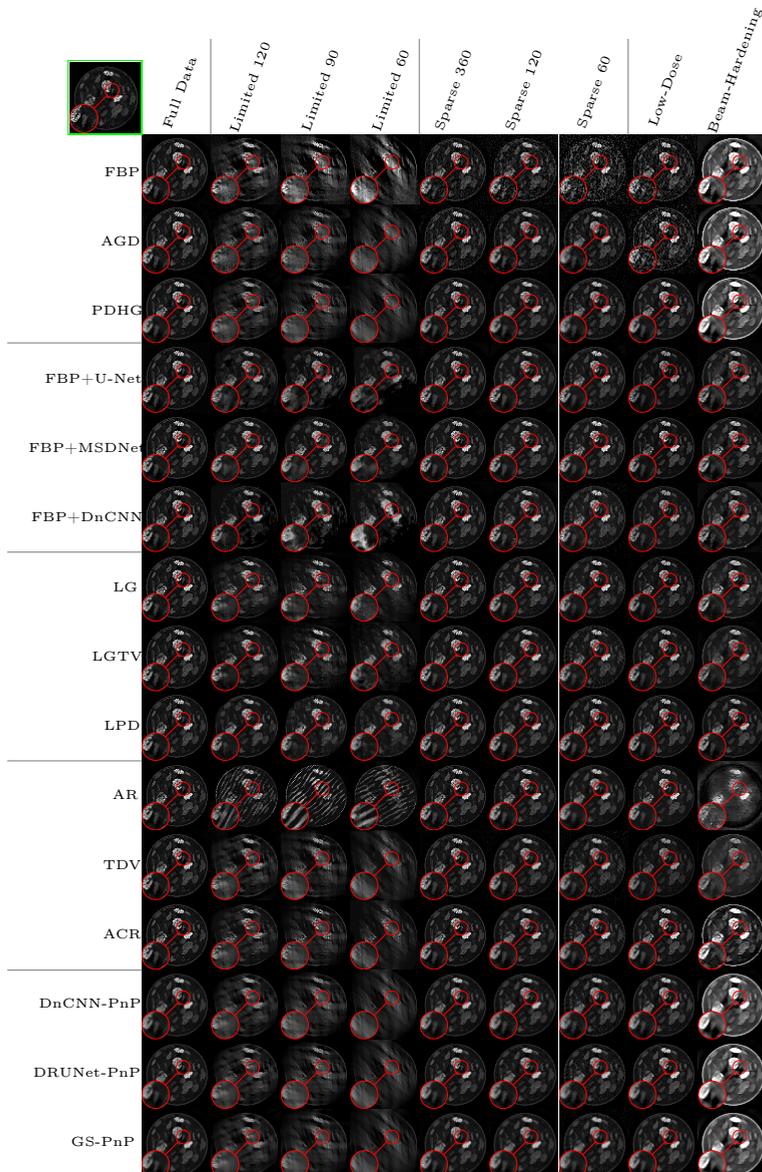


Figure 6.2: Qualitative analysis of all evaluated methods for slice 182 of the test dataset in comparison to the “gold standard” iterative reference reconstruction of the 2DeteCT dataset (green box).

Table 6.5: Quantitative analysis of all evaluated methods with respect to PSNR and SSIM on their performance in the CT image reconstruction tasks: Full Data “mode 2”, Low-Dose “mode 1”, and Beam-Hardening “mode 3”.

Method	Metric	CT Image Reconstruction Task		
		Full Data	Low-Dose	Beam-Hardening
Classical Methods				
FBP	SSIM	0.7463 ± 0.0296	0.0838 ± 0.0212	0.3367 ± 0.0464
	PSNR	35.0285 ± 2.0907	18.6437 ± 2.0508	14.5594 ± 1.9056
AGD	SSIM	0.7753 ± 0.0380	0.0727 ± 0.0182	0.3483 ± 0.0650
	PSNR	35.1006 ± 2.2801	17.7062 ± 2.0517	14.4294 ± 1.9255
PDHG	SSIM	0.7689 ± 0.0498	0.6820 ± 0.0749	0.4492 ± 0.0616
	PSNR	34.3251 ± 1.9703	31.9637 ± 1.9824	15.0090 ± 1.9094
Post-Processing Methods				
FBP+U-Net	SSIM	0.6499 ± 0.0681	0.7632 ± 0.0780	0.6336 ± 0.0760
	PSNR	32.6998 ± 1.9512	27.6772 ± 3.0133	28.0629 ± 3.8439
FBP+MSDNet	SSIM	0.8481 ± 0.0384	0.7253 ± 0.0864	0.7991 ± 0.0665
	PSNR	33.2999 ± 1.9492	31.6910 ± 1.9559	31.4389 ± 2.0988
FBP+DnCNN	SSIM	0.8324 ± 0.0403	0.7127 ± 0.0806	0.6328 ± 0.0751
	PSNR	32.2575 ± 1.9506	29.7645 ± 1.9748	28.2405 ± 2.0374
Learned / Unrolled Iterative Methods				
LG	SSIM	0.7498 ± 0.0708	0.6685 ± 0.0772	0.6025 ± 0.0747
	PSNR	32.4409 ± 1.9527	30.4679 ± 1.9945	28.4148 ± 1.9584
LGTV	SSIM	0.8221 ± 0.0532	0.7008 ± 0.0797	0.6656 ± 0.0774
	PSNR	33.3312 ± 1.9493	30.9991 ± 1.9328	29.5372 ± 2.0008
LPD	SSIM	0.8447 ± 0.0370	0.8282 ± 0.0519	0.8352 ± 0.0568
	PSNR	33.3086 ± 1.9466	32.6685 ± 1.9656	33.1382 ± 1.9759
Learned Regularizer Methods				
AR	SSIM	0.8196 ± 0.0385	0.8039 ± 0.0505	0.3125 ± 0.0479
	PSNR	32.3583 ± 1.9621	31.0472 ± 1.9763	19.5692 ± 1.9612
TDV	SSIM	0.7282 ± 0.0652	0.6047 ± 0.0815	0.5494 ± 0.0635
	PSNR	33.1204 ± 1.9496	28.2429 ± 1.9433	25.7832 ± 2.0109
ACR	SSIM	0.8518 ± 0.0362	0.8163 ± 0.0522	0.6742 ± 0.0505
	PSNR	33.7131 ± 1.9450	32.2621 ± 1.9689	18.5708 ± 1.7824
Plug-and-Play Methods				
DnCNN-PnP	SSIM	0.8585 ± 0.0402	0.7795 ± 0.0523	0.5989 ± 0.0375
	PSNR	32.8506 ± 1.9306	31.3986 ± 1.9424	15.4667 ± 1.9107
DRUNet-PnP	SSIM	0.8573 ± 0.0405	0.7984 ± 0.0517	0.5945 ± 0.0375
	PSNR	32.8935 ± 1.9327	31.5762 ± 1.9480	15.4543 ± 1.9102
GS-PnP	SSIM	0.7856 ± 0.0590	0.7727 ± 0.0683	0.5131 ± 0.0562
	PSNR	32.5734 ± 1.9331	31.7931 ± 1.9444	15.3466 ± 1.9128

Table 6.6: Quantitative analysis of all evaluated methods with respect to PSNR and SSIM on their performance in the CT image reconstruction tasks of Limited-Angle.

Method	Metric	CT Image Reconstruction Task		
		Limited-Angle 120	Limited-Angle 90	Limited-Angle 60
Classical Methods				
FBP	SSIM	0.3418 ± 0.0354	0.2369 ± 0.0323	0.1557 ± 0.0288
	PSNR	22.4188 ± 1.9240	19.4251 ± 1.9405	16.9057 ± 1.9579
AGD	SSIM	0.4904 ± 0.0550	0.4411 ± 0.0555	0.4146 ± 0.0559
	PSNR	25.7508 ± 1.9690	24.1128 ± 1.9528	22.8848 ± 1.9531
PDHG	SSIM	0.5923 ± 0.0550	0.5194 ± 0.0547	0.4658 ± 0.0510
	PSNR	26.3646 ± 1.9443	24.4392 ± 1.9321	22.7556 ± 1.9428
Post-Processing Methods				
FBP+U-Net	SSIM	0.7251 ± 0.0519	0.6338 ± 0.0659	0.5892 ± 0.0678
	PSNR	28.7931 ± 2.0052	27.3875 ± 2.0441	23.8511 ± 2.8370
FBP+MSDNet	SSIM	0.7840 ± 0.0641	0.7695 ± 0.0579	0.7148 ± 0.0726
	PSNR	30.7850 ± 1.9784	29.0111 ± 1.9740	27.4624 ± 1.9661
FBP+DnCNN	SSIM	0.5829 ± 0.0521	0.5661 ± 0.0528	0.5174 ± 0.0559
	PSNR	20.4433 ± 2.5345	23.4066 ± 2.2764	21.2132 ± 2.5140
Learned / Unrolled Iterative Methods				
LG	SSIM	0.6740 ± 0.0652	0.5746 ± 0.0655	0.5378 ± 0.0706
	PSNR	28.1639 ± 1.9380	26.3188 ± 1.9508	24.7228 ± 1.9630
LGTV	SSIM	0.6804 ± 0.0647	0.5590 ± 0.0639	0.5091 ± 0.0643
	PSNR	28.4131 ± 1.9300	26.0867 ± 1.9545	25.0225 ± 1.9687
LPD	SSIM	0.8296 ± 0.0410	0.8049 ± 0.0444	0.7724 ± 0.0571
	PSNR	31.1723 ± 1.9607	29.2534 ± 1.9941	28.0734 ± 1.9589
Learned Regularizer Methods				
AR	SSIM	0.6869 ± 0.0505	0.6100 ± 0.0543	0.5742 ± 0.0620
	PSNR	23.8496 ± 2.1578	21.1830 ± 2.1772	22.2350 ± 2.0260
TDV	SSIM	0.5940 ± 0.0595	0.5459 ± 0.0572	0.5282 ± 0.0584
	PSNR	26.3233 ± 1.9315	24.8127 ± 1.9399	23.3939 ± 1.9662
ACR	SSIM	0.7114 ± 0.0543	0.6575 ± 0.0541	0.5515 ± 0.0529
	PSNR	27.1792 ± 1.9441	25.3342 ± 1.9442	23.4915 ± 1.9539
Plug-and-Play Methods				
DnCNN-PnP	SSIM	0.7617 ± 0.0410	0.6981 ± 0.0441	0.6200 ± 0.0475
	PSNR	26.9997 ± 1.9330	25.0658 ± 1.9248	23.4108 ± 1.9521
DRUNet-PnP	SSIM	0.7634 ± 0.0411	0.7002 ± 0.0443	0.6149 ± 0.0511
	PSNR	27.0262 ± 1.9334	25.0829 ± 1.9254	23.4362 ± 1.9520
GS-PnP	SSIM	0.6396 ± 0.0670	0.5668 ± 0.0663	0.4989 ± 0.0625
	PSNR	26.3318 ± 1.9328	24.5129 ± 1.9282	22.8225 ± 1.9481

Table 6.7: Quantitative analysis of all evaluated methods with respect to PSNR and SSIM on their performance in the CT image reconstruction tasks of Sparse-Angle.

Method	Metric	CT Image Reconstruction Task		
		Sparse-Angle 360	Sparse-Angle 120	Sparse-Angle 60
Classical Methods				
FBP	SSIM	0.2947 ± 0.0453	0.1231 ± 0.0225	0.0611 ± 0.0112
	PSNR	24.9674 ± 2.0415	19.8769 ± 2.0124	16.6451 ± 1.9972
AGD	SSIM	0.3867 ± 0.0563	0.4142 ± 0.0630	0.4333 ± 0.0664
	PSNR	26.9629 ± 2.0444	27.5127 ± 1.9948	27.2796 ± 1.9553
PDHG	SSIM	0.6998 ± 0.0685	0.6712 ± 0.0718	0.5952 ± 0.0728
	PSNR	32.9158 ± 1.9739	31.7980 ± 1.9798	29.8020 ± 1.9326
Post-Processing Methods				
FBP+U-Net	SSIM	0.7449 ± 0.0801	0.7518 ± 0.0657	0.7728 ± 0.0592
	PSNR	30.4766 ± 3.4844	26.0785 ± 6.3779	19.5421 ± 9.9613
FBP+MSDNet	SSIM	0.8392 ± 0.0473	0.7993 ± 0.0650	0.7626 ± 0.0789
	PSNR	33.1188 ± 1.9820	32.2993 ± 1.9928	30.9931 ± 1.9875
FBP+DnCNN	SSIM	0.7864 ± 0.0618	0.6701 ± 0.0864	0.6180 ± 0.0796
	PSNR	31.7575 ± 2.0213	29.1817 ± 2.3053	28.6079 ± 2.3389
Learned / Unrolled Iterative Methods				
LG	SSIM	0.7846 ± 0.0628	0.6795 ± 0.0790	0.6428 ± 0.0777
	PSNR	32.5946 ± 1.9764	31.1950 ± 1.9658	29.9360 ± 1.9603
LGTV	SSIM	0.7811 ± 0.0659	0.7100 ± 0.0745	0.7081 ± 0.0689
	PSNR	32.9404 ± 1.9666	31.4072 ± 1.9647	29.9021 ± 1.9357
LPD	SSIM	0.8433 ± 0.0479	0.8300 ± 0.0500	0.8206 ± 0.0508
	PSNR	33.3809 ± 1.9513	32.7032 ± 1.9685	32.0583 ± 1.9789
Learned Regularizer Methods				
AR	SSIM	0.8309 ± 0.0447	0.8117 ± 0.0553	0.7949 ± 0.0595
	PSNR	32.8067 ± 1.9714	32.1030 ± 1.9577	30.8378 ± 1.9532
TDV	SSIM	0.6815 ± 0.0736	0.6235 ± 0.0741	0.5725 ± 0.0728
	PSNR	32.2673 ± 1.9641	30.6585 ± 1.9357	28.9451 ± 1.8995
ACR	SSIM	0.8271 ± 0.0494	0.8074 ± 0.0539	0.7849 ± 0.0524
	PSNR	33.1537 ± 1.9632	31.9181 ± 1.9666	30.5147 ± 1.9338
Plug-and-Play Methods				
DnCNN-PnP	SSIM	0.8405 ± 0.0432	0.8021 ± 0.0465	0.7637 ± 0.0484
	PSNR	32.4627 ± 1.9309	31.3847 ± 1.9271	29.9350 ± 1.8980
DRUNet-PnP	SSIM	0.8398 ± 0.0433	0.8000 ± 0.0465	0.7658 ± 0.0498
	PSNR	32.5065 ± 1.9324	31.3949 ± 1.9266	29.9518 ± 1.9085
GS-PnP	SSIM	0.7622 ± 0.0628	0.7588 ± 0.0684	0.6937 ± 0.0701
	PSNR	32.1977 ± 1.9353	31.2728 ± 1.9370	29.5791 ± 1.8960

6.5.2 Quantitative and qualitative analysis of the evaluated methods

For this benchmarking we compared a range of algorithms representative for different categories of learned reconstruction methods in several CT image reconstruction tasks and reported their performance with respect to SSIM and PSNR. However, it is particularly important to note that the quantitative analysis presented in Tables 6.5, 6.6, and 6.7 does not capture the nuances of the quality of the reconstruction upon visual inspection. This is particularly true in cases where the task at hand is more challenging, e.g. limited-angle reconstruction from 60° . In this case, the quantitative analysis indicates that both post-processing methods and learned/unrolled iterative methods perform similarly to the learned regularizer and PnP methods (e.g. ACR has better SSIM than LG). However, the qualitative analysis (visual inspection of the reconstructions) shows that for PnP and Learned Regularizer methods, limited angle reconstructions are not performing well, arguably producing images as bad as FBP. Thus, one should not fully trust the performance metrics when comparing models solving such challenging inverse problem scenarios.

Along the same lines, the performance of all evaluated methods on the Full Data reconstruction task should be considered carefully. The “gold standard” or ground truth for this reconstruction task are iterative reference reconstructions computed with an AGD algorithm on a larger field-of-view (2048×2048) and cropped to its center region of (1024×1024). Given that this is a relatively well-posed problem, AGD should have converged to the true minimizer of the data fidelity functional, and thus the resulting image should be a good target for data-driven methods. However, this is not strictly true, as noise in CT acquisition is much more complex than just Gaussian, and this target is not really the exact ground truth. One could argue that a solution from an explicitly regularized algorithm (e.g. Chambolle-Pock with TV) would be also an appropriate target to use, and this would change the numerical results of this work. This choice of using AGD as target, is likely not highly impacting the conclusions of this work, but it is important to clarify that this is a choice, and not a definition of the ground truth.

This explains why AGD performs excellently, and if the same scale of the reconstruction would be kept, the SSIM would be 1 and PSNR infinity. But, as explained, this means that all evaluated methods in the Full Data reconstruction are learning to produce AGD-like results, not the actual ground truth. PnP and learned regularizer methods produce a high SSIM because they are also optimized using Gradient Descent, but with a learned regularization step. Therefore, they can mimic AGD more appropriately.

The beam-hardening corrected reconstruction is a particularly interesting case study as the errors caused by beam-hardening are very non-local and non-linear. This presents a greater challenge than tasks which are more similar to “denoising” such as sparse-angle. Therefore all methods that do not learn to imitate the final results directly (classical methods, PnP and adversarial regularizers) fail to produce a good image. Beam-hardening corrected reconstruction appears to be a more challenging task for variational regularization methods and learned regularizers and PnP methods. This can be explained by the linearization of the forward operator which assumes

monochromatic X-ray sources. However, beam-hardening is ultimately a non-linear effect caused by wide beam spectra and their non-linear absorption. This operator mismatch pushes the minimization into the wrong direction, causing artifacts, as the forward model deviates from the physical process of acquiring the measurements.

As an overall discussion on performance, it is worth noting that post-processing methods, albeit lacking mathematical guarantees, consistently produce quantitatively and visually relatively good results on all CT image reconstruction tasks. As is seen in Figure 6.1, for example for FBP+DnCNN in the Limited Angle reconstruction from 60° , these methods, however, may suffer from “hallucinations”. This should not come as a surprise as post-processing methods do not enforce data consistency.

Generally, learned/unrolled iterative methods tend to be better at ensuring consistency in both data and image space, while (learned) regularization approaches provably achieve consistency [176]. However, we emphasize that data consistency is necessary but not sufficient for preventing hallucinations, especially as the ill-posedness of the reconstruction problem increases, and as a result these methods may also suffer from hallucinations. Indeed, even classical, model-based, approaches with theoretical guarantees of data consistency have been seen to exhibit a sensitivity to adversarial perturbations [67]. At the same time, training iterative unrolled methods requires much more time than post-processing methods, although the number of parameters are orders of magnitude smaller than standard networks like the U-Net.

Furthermore, we find that adversarial regularization and PnP approaches both tend to be well-performing in both sparse and full-data settings, oftentimes reaching performance of supervised learned/unrolled iterative methods, by the virtue of relying on the variational formulation and ultimately interpolating the missing image data. However, in the limited angle setting it is no longer an interpolation problem, but instead an in-painting problem. Neither PnP nor AR have been designed for in-painting directly and often rely on local image information. However, in-painting inherently requires non-local information in order to fill-in the missing data.

6.5.3 Limitations and broader impact

In our benchmarking study, we aimed to establish a foundational understanding of how learned algorithms of different method categories perform on standardized CT reconstruction tasks with real-world experimental data. We prioritized adequate performance over extensive hyperparameter tuning to produce baseline results for a comparative analysis. However, we recognize that the interplay of hyperparameters such as architecture, learning rates, regularization strengths, and iteration counts can significantly affect performance.

Consequently, the results shown in Figure 6.2 and Tables 6.5, 6.6, and 6.7 should be interpreted cautiously, as they reflect performance under limited tuning. Future users must conduct thorough hyperparameter optimization tailored to their specific applications to fully leverage each method’s potential.

While other CT image reconstruction tasks such as region-of-interest tomography, super-resolution, or segmentation are supported by the 2DeteCT dataset in principle, they have not yet been fully implemented in the benchmarking framework. Furthermore, while the dataset was designed to resemble abdominal CT scans, there are several remaining differences. However, 2DeteCT does provide realistic experimental data for a range of research fields such as manufacturing industry, food industry, and materials science. In an ideal case scenario, it would be possible to have raw measurement data for medical CT scanners and medically relevant subjects. However, medical CT manufacturers claim this data as proprietary and ethical concerns on both patients' privacy and radiation dose prohibit acquiring matching data pairs of e.g. high-dose and low-dose scans or other acquisition modes. Moreover, the mismatch between the 2DeteCT dataset and medical CT image characteristics and morphology could create a significant performance gap that remains unexamined both in this study and in the existing literature. As a result, there is no guarantee that the overall performance trends observed in this work are transferable to medical CT cases. To address this, future research will need to focus on acquiring medical datasets and investigating retraining or transfer learning techniques to confirm or challenge these findings.

Additionally, the dataset has been acquired using a specific acquisition geometry and a non-medical micro-CT scanner, which could limit the generalization of trained algorithms to other CT data. For that, the performance of models trained on 2DeteCT could be evaluated under out-of-distribution (OOD) conditions in both image distribution and forward operator. For instance, applying the trained models to a different dataset would allow for assessing the models' generalization capacity. This type of experiment is particularly critical in the medical field, where out-of-distribution changes are common [95].

Such an evaluation under OOD conditions should be conducted thoroughly, involving a wide range of OOD cases rather than just one or a few specific instances. While the extensive combinations of OOD tests and generalization scenarios pose significant computational challenges, making them impractical for the current study, we release the trained models and accompanying code on GitHub (see Section 6.5.4). We hope this will facilitate future research in this area and believe that our comparison framework lays a solid foundation for conducting comprehensive OOD studies.

There are also slight remaining beam-hardening artifacts in the filtered, clean acquisition of "mode 2", indicating a reduction but not complete removal of beam-hardening. The chosen performance metrics of SSIM and PSNR are common quality assessments, but meaningful quality metrics for reconstructed (medical) CT images should be clinically relevant, task dependent, and aware of unaltered image content [225]. Despite these challenges, trained models on this dataset could potentially be applied to other data through transfer learning, with potential benefits for the medical sector. Researchers must be aware of potential distribution shifts and validate their algorithms on suitable data for the intended application. However, unlike other existing datasets, the 2DeteCT dataset provides raw experimental measurement data and the presented

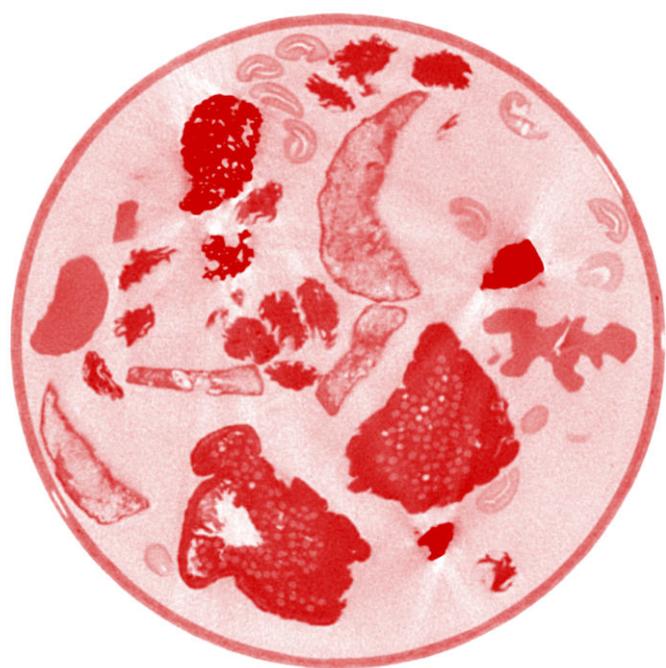
benchmarks show how well existing algorithms work on real-world experimental data. With this, we take a step towards closing the gap between developed algorithms and real-world applications by utilizing real-world experimental data instead of simulated data.

6.5.4 Code and data availability

The 2DeteCT dataset which serves as the foundation of this benchmarking paper is hosted on zenodo [102] and the LION toolbox is hosted on GitHub by the Cambridge Image Analysis group and maintained by Ander Biguri. This open-source toolbox allows for easy access to a variety of methods, tools, and resources. Additionally, new methods can be seamlessly added to the toolbox, with demos available to showcase how code should be organized and set up within the LION framework. This holds also true for future extensions of the available CT image reconstruction tasks mentioned in the paragraph “Limitations and broader impact”. By regularly updating and expanding the toolbox the benchmarking project remains relevant, efficient, and effective in advancing the field of ML-based CT image reconstruction. The models and the scripts to train and evaluate the models presented in this benchmark, will be made available in the GitHub repository above and tagged with ‘AMMC_benchmark’ upon publication of this work.

6.6 Conclusion

Our benchmarking study provides a comparison of a fixed set of data-driven CT reconstruction algorithms with real-world experimental data in a reproducible and reusable way. It provides a starting point to develop new methods significantly faster as time-consuming implementations of data loaders, reconstruction tasks, comparison methods and evaluation protocols do not have to be redone. The open-source toolbox allows for seamless addition of new state-of-the-art methods and for extensions towards other problems and different CT reconstruction tasks.



Conclusion and outlook

In this thesis, we highlighted the importance of tailoring the CT acquisition to the respective scanning samples and used this concept to acquire an unprecedented 2D expandable, trainable, experimental Computed Tomography dataset for machine learning, the 2DeteCT dataset. With this dataset we advanced the development of learned algorithms for 2D X-ray computed tomography, addressed the broadly observed sim-to-real gap, and laid the foundation for the reproducible comparison of machine learning based CT image reconstruction methods. In this chapter, we will summarize the contributions and limitations of the work presented in this thesis. We conclude with an outlook on subsequent work building on the contents of this thesis and share perspectives on future research and their possible impact on the research field.

7.1 Contributions and limitations

The contributions of this thesis can be classified as curiosity-driven, instructive, and future-oriented. Many of the findings presented in this work have been results of simple but fundamental questions such as: "How do we acquire an informative CT scan?", "What are application areas that the field of CT imaging works on with machine learning methods?", "Do we need real-world experimental noisy data or is artificially simulated data enough?", "How do machine learning methods from different method categories compare against each other on real-world experimental CT reconstruction tasks?". The research presented in this thesis focused not only on publishable results but also on how it can be instructive for other researchers and enable future work building on its results.

In **chapter 3**, we showed the importance of tailoring the CT acquisition to the scanned samples in the context of cultural heritage objects. To make the results

comprehensible not only to X-ray imaging specialists but a broad audience including museum professionals we introduced also several of the key factors that influence CT image formation. In particular, we discussed the use of beam filtration and the underlying physics of manipulating the emitted X-ray beam spectrum of a CT source to improve the image quality of CT scans of multi-material and multi-scale cultural heritage objects. We gave instructions on how this can be done with limited resources in a low-cost DIY fashion. We not only presented the improved final result but demonstrated the influence of the CT acquisition parameters on the radiographs and CT reconstructions. Illustrated by case study objects from the textile collection of the Rijksmuseum, Amsterdam, The Netherlands, we provided insights and intuitions on choosing suitable acquisition parameters and on how to design an informative CT scan. To enable future work of museum professionals building on these insights we extracted a general concept of steps to design an object-tailored CT scan for individual cases.

Despite the clear image quality improvement of the object-tailored method presented in chapter 3 compared to an untailored CT acquisition, there remain disadvantages. Firstly, that through the filtering the beam intensity is reduced, which results in a decreased SNR, and secondly, that the beam hardening problem cannot be completely eliminated [87]. To counter the degraded X-ray signal and greater image noise, longer exposure times and averaging can be used [92]. Depending on the available scanning time and capabilities of the scanning facility these disadvantages can therefore be mitigated. Another limitation of the results presented in chapter 3 is, of course, that they are from the usage of a specific CT system as well as specific objects which necessitates tailoring them to new individual cases. With the concept of steps for an object-tailored CT scan design we provided though a short guidance for this, while the two case studies provide insights and intuitions on choosing suitable acquisition parameters that take the objects' characteristics into consideration and improve image quality in CT reconstructions.

In **chapter 4**, we designed and acquired an unprecedented 2D expandable, trainable, experimental Computed Tomography dataset for machine learning, the 2DeteCT dataset. We deemed two-dimensional, reconstructed CT slices especially useful for method development since the corresponding learning and reconstruction tasks require less computational resources compared to their three-dimensional counterparts. We started our research with the assessment that mathematical and computational studies typically rely on artificial data simulated with varying degrees of realism. Furthermore, we investigated what computational imaging tasks we wanted to serve with our dataset - namely, low-dose reconstruction, limited or sparse angular sampling, beam-hardening artifact reduction, super-resolution, region-of-interest tomography or segmentation. To enable similar research we described not only the overall study design but also explained the experimental design of the acquisition and the data processing protocols. Additionally to the raw projection data in sinograms we provided also reference reconstructions and segmentations as well as an implementation of the complete computational pipeline based on open-source software. This makes the dataset accessible to a broad range of researchers including those who do not have high-performance computing facilities readily available.

Although the dataset presented in chapter 4 resembles abdominal CT scans, it has a limited selection of samples including dried fruits, nuts, coffee beans, and stones. This fixed selection may restrict its applicability to the medical sector. Additionally, the dataset has been acquired using a specific acquisition geometry and a non-medical micro-CT scanner, which could limit the generalization of algorithms trained on this dataset to other CT data. There are also slight remaining beam hardening artifacts in the filtered, clean acquisition of “mode 2”, indicating a reduction but not complete removal of beam hardening.

In **chapter 5**, we aimed to answer the important question whether it is enough to train learned denoising algorithms on simulated noisy data or whether it is necessary to use experimental noisy data. Although this question is simple, it targets the implicit assumption of many machine learning studies that simulating noisy data is sufficient. Without the 2DeteCT dataset and its matching data of both low-dose and high-dose CT acquisitions it would have not been possible to investigate this question. To observe the different performances of algorithms trained on simulated noisy data and on experimental noisy data we trained two common neural networks on both types of noisy data, experimental and simulated. The testing of the trained networks was carried out on the data that they have been trained on but also to their respective counterparts. The results were evaluated quantitatively in the sinogram and reconstructed image domain as well as qualitatively in the reconstructed image domain by visual inspection. We showed that it is important to choose appropriate noise levels that match experimental data well when designing a noise simulation approach. Leveraging simulated data for machine learning in computational imaging can be challenging, if this data is quite different from the experimental data, and can impact the transfer of learned systems to the real-world. An end-to-end training, i.e. a mapping from raw measurement data to desired target reconstructions, outperformed the sequential approach. Future work can build on our noise simulation model that produces data that are close enough to experimental data to make the models transferable to real-world applications or investigate more sophisticated noise simulation approaches for CT image denoising applications to bridge the remaining gap between simulated and experimental data.

Even though chapter 5 answers the question to which extent algorithms trained on simulated noisy data are applicable to real-world experimental noisy data, its results are limited to one simulation approach. This realistic yet computationally efficient simulation method utilized less-commonly available raw experimental measurement data and already captures much of the complexity of the experimental noise in the measurements. However, this work shows that the non-linearity of the imaging process is not captured well enough.

In **chapter 6**, we proposed a benchmarking framework of various machine learning algorithms for different image reconstruction tasks in X-ray computed tomography. We categorized these methods into post-processing networks, learned/unrolled iterative methods, learned regularizer methods, and plug-and-play methods. We created a pipeline for easy implementation and evaluation using key performance metrics,

including SSIM and PSNR, to showcase the effectiveness of various algorithms on tasks such as full data reconstruction, limited-angle reconstruction, sparse-angle reconstruction, low-dose reconstruction, and beam-hardening corrected reconstruction. To enable reproducibility and future extensions in machine learning based CT image reconstruction research, both the dataset and toolbox are published open source. This benchmarking framework also helps to develop new methods significantly faster and to compare against different state-of-the-art methods easier as time-consuming implementations of data loaders, reconstruction tasks, comparison methods and evaluation protocols do not have to be redone.

Overall, the performance of the different algorithms in the aforementioned four method categories can be summarized as follows: The post-processing methods, while lacking mathematical guarantees, generally produce strong quantitative and visual results across CT image reconstruction tasks. However, they can suffer from "hallucinations" due to the absence of data consistency enforcement. The learned/unrolled iterative methods excel in ensuring data and image consistency, making them less prone to hallucinations compared to post-processing techniques. Nonetheless, they are computationally intensive to train, despite having fewer parameters than standard networks like U-Net. The learned regularization methods are proven to achieve consistency and can mitigate hallucinations. However, they are not immune to issues, especially in highly ill-posed reconstruction scenarios where even model-based methods with theoretical guarantees can be sensitive to adversarial perturbations. Also, the Plug-and-Play (PnP) methods perform well in sparse and full-data contexts, often matching the performance of supervised learned/unrolled iterative methods. However, they struggle in limited angle settings, where the reconstruction task shifts from interpolation to in-painting, requiring non-local information, which their design does not readily facilitate.

While other CT image reconstruction tasks such as region-of-interest tomography, super-resolution, or segmentation are supported by the 2DeteCT dataset, in principle, they have not yet been fully implemented in the benchmarking framework presented in chapter 6. Furthermore, the chosen performance metrics of SSIM and PSNR might be, although being a common quality assessment, of limited significance since meaningful quality metrics for reconstructed (medical) CT images should be clinically relevant, task dependent, and aware of unaltered image content [225].

7.2 Outlook

Subsequent work of chapter 3 includes both work in the field of cultural heritage [25, 26, 153] but also imaging science [38, 162, 178]. We expect that the instructive and detailed description of an object-tailored CT acquisition including a low-cost realization of beam filtration techniques with DIY approaches will see increasing adoption and might help researchers in the future to acquire more informative CT scans.

Also the 2DeteCT dataset from chapter 4 has already been used outside of this thesis for a variety of publications [60, 201, 220] and its experimental design approach has been re-used [181]. Future work on this dataset can include expansions of its scope. Adding more slices with the same sample mix could be for example used to increase the size of the data collection and to host possible coding challenges. Other expansions could include various new samples in the sample mix or using an entirely different sample mix. Especially, adding more detailed multi-class segmentations to the dataset could enable work on deep learning based segmentation algorithms or methods for simultaneous tomographic image reconstruction and segmentation. To enable faster prototyping and testing of machine learning algorithms there has been calls for smaller versions of the 2DeteCT dataset within the research community. Therefore, we consider releasing a small-scale 2DeteCT version with smaller resolution (sinograms of shape 478×721 instead of 1912×3601 and reconstructions of 256×256 instead of 1024×1024).

Future work on learned denoisers and noise simulation approaches as discussed in chapter 5 could investigate how to include effects such as beam hardening or photon starvation in a computationally efficient way. Both, simplified Monte Carlo particle simulations or generative models trained on experimental noisy and "clean" data could potentially solve this challenge.

Generally, releases of large-scale, open-source datasets such as MNIST [123], CIFAR [119] and ImageNet [46] enabled machine learning researchers to establish standardized benchmarks and to continuously advance the state-of-the-art. Accordingly, our work in chapter 6 might be a first step to enabling similar breakthroughs in machine learning based CT image reconstruction. Future work in this direction can include extensions of the LION toolbox to incorporate tasks such as super-resolution, ROI-tomography, foreign object detection, or segmentation. Furthermore, it is desirable to expand the evaluated methods to transformers [197, 198, 214, 229], diffusion models [41, 42, 133, 186], and self- [189, 211] and unsupervised [91, 120, 128, 210] methods and compare their performance against the provided baseline of (weakly) supervised learning methods.

Lastly, the 2DeteCT dataset with its matching raw projection data and reference reconstructions and segmentations of three different acquisition modes provides an extensive basis for developing and testing machine learning methods and we hope the full depth of it is used in future research in the field of computational imaging.

Bibliography

- [1] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, and J. Sijbers. “Fast and flexible X-ray tomography using the ASTRA toolbox”. *Opt. Express* 24.22 (Oct. 2016), pp. 25129–25147 (cit. on pp. 99, 110, 112).
- [2] J. Adler, H. Kohr, and O. Öktem. *ODL 0.6.0*. Version v0.6.0. Apr. 2017 (cit. on p. 110).
- [3] J. Adler and O. Öktem. “Learned primal-dual reconstruction”. *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1322–1332 (cit. on p. 113).
- [4] J. Adler and O. Öktem. “Solving ill-posed inverse problems using iterative deep neural networks”. *Inverse Problems* 33.12 (2017), p. 124007 (cit. on pp. 102, 113).
- [5] J. Als-Nielsen and D. McMorrow. *Elements of modern X-ray physics*. Hoboken, NJ: John Wiley & Sons, 2011 (cit. on pp. 40, 69).
- [6] J. Alzubi, A. Nayyar, and A. Kumar. “Machine learning from theory to algorithms: an overview”. In: *Journal of physics: conference series*. Vol. 1142. IOP Publishing, 2018, p. 012012 (cit. on p. 13).
- [7] O. Amir, D. Braunstein, and A. Altman. “Dose optimization tool”. In: *Medical Imaging 2003: Visualization, Image-Guided Procedures, and Display*. Vol. 5029. SPIE, 2003, pp. 815–821 (cit. on p. 84).
- [8] Amsterdam museum. *KA 18660 - Phallus Beurs*. 1500 - 1600 (cit. on p. 47).
- [9] M. Andonova. “Ancient basketry on the inside: X-ray computed microtomography for the non-destructive assessment of small archaeological monocotyledonous fragments: examples from Southeast Europe”. *Heritage Science* 9 (Dec. 2021) (cit. on p. 38).
- [10] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, et al. “The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans”. *Medical physics* 38.2 (2011), pp. 915–931 (cit. on p. 104).
- [11] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb. “Solving inverse problems using data-driven models”. *Acta Numerica* 28 (2019), pp. 1–174 (cit. on pp. 17, 102, 105, 106).
- [12] A. Barbu. “Training an active random field for real-time image denoising”. *IEEE Transactions on Image Processing* 18.11 (2009), pp. 2451–2462 (cit. on p. 106).

- [13] S. Barutcu, S. Aslan, A. K. Katsaggelos, and D. Gürsoy. “Limited-angle computed tomography with deep image and physics priors”. *Scientific Reports* 11.1 (2021), p. 17740 (cit. on p. 109).
- [14] A. Beck and M. Teboulle. “A fast iterative shrinkage-thresholding algorithm for linear inverse problems”. *SIAM Journal on Imaging Sciences* 2.1 (2009), pp. 183–202 (cit. on pp. 106, 107).
- [15] M. Beister, D. Kolditz, and W. A. Kalender. “Iterative reconstruction methods in X-ray CT”. *Physica Medica* 28.2 (2012), pp. 94–108 (cit. on p. 113).
- [16] M. Benning and M. Burger. “Modern regularization methods for inverse problems”. *Acta Numerica* 27 (2018), pp. 1–111 (cit. on p. 107).
- [17] T. M. Benson and B. K. De Man. “Synthetic CT noise emulation in the raw data domain”. In: *IEEE Nuclear Science Symposium & Medical Imaging Conference*. IEEE. 2010, pp. 3169–3171 (cit. on p. 84).
- [18] A. Biguri, M. Dosanjh, S. Hancock, and M. Soleimani. “TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction”. *Biomedical Physics & Engineering Express* 2.5 (2016), p. 055010 (cit. on p. 110).
- [19] A. Biguri, R. Lindroos, R. Bryll, H. Towsyfyan, H. Deyhle, I. El khalil Harrane, R. Boardman, M. Mavrogordato, M. Dosanjh, S. Hancock, et al. “Arbitrarily large tomography with iterative algorithms on multiple GPUs using the TIGRE toolbox”. *Journal of Parallel and Distributed Computing* 146 (2020), pp. 52–63 (cit. on p. 110).
- [20] A. Biguri and S. Mukherjee. “Advancing the frontiers of deep learning for low-dose 3D cone-beam computed tomography (CT) reconstruction”. In: *2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*. IEEE. 2024, pp. 81–82 (cit. on pp. 85, 103).
- [21] F. E. Boas, D. Fleischmann, et al. “CT artifacts: causes and reduction techniques”. *Imaging Med* 4.2 (2012), pp. 229–240 (cit. on p. 41).
- [22] J. M. Boone and J. A. Seibert. “An accurate method for computer-generating tungsten anode x-ray spectra from 30 to 140 kV”. *Medical Physics* 24.11 (1997), pp. 1661–1670 (cit. on pp. 12, 34, 43, 70).
- [23] F. G. Bossema, S. B. Coban, A. Kostenko, P. van Duin, J. Dorscheid, I. Garachon, E. Hermens, R. van Liere, and K. J. Batenburg. “Integrating expert feedback on the spot in a time-efficient explorative CT scanning workflow for cultural heritage objects”. *Journal of Cultural Heritage* 49 (2021), pp. 38–47 (cit. on pp. 7, 38).
- [24] F. G. Bossema, M. Domínguez-Delmás, W. J. Palenstijn, A. Kostenko, J. Dorscheid, S. B. Coban, E. Hermens, and K. J. Batenburg. “A novel method for dendrochronology of large historical wooden objects using line trajectory X-ray tomography”. *Scientific Reports* 11.1 (2021), p. 11024 (cit. on p. 38).

- [25] F. G. Bossema, P. J. van Laar, K. Meechan, D. O’Flynn, J. Dyer, T. van Leeuwen, S. Meijer, E. Hermens, and K. J. Batenburg. “Inside out: Fusing 3D imaging modalities for the internal and external investigation of multi-material museum objects”. *Digital Applications in Archaeology and Cultural Heritage* 31 (2023), e00296 (cit. on p. 132).
- [26] F. G. Bossema, W. J. Palenstijn, A. Heginbotham, M. Corona, T. van Leeuwen, R. van Liere, J. Dorscheid, D. O’Flynn, J. Dyer, E. Hermens, et al. “Enabling 3D CT-scanning of cultural heritage objects using only in-house 2D X-ray equipment in museums”. *Nature Communications* 15.1 (2024), p. 3939 (cit. on p. 132).
- [27] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al. “Distributed optimization and statistical learning via the alternating direction method of multipliers”. *Foundations and Trends® in Machine learning* 3.1 (2011), pp. 1–122 (cit. on p. 107).
- [28] R. N. Bracewell and A. Riddle. “Inversion of fan-beam scans in radio astronomy”. *Astrophysical Journal*, vol. 150, p. 427 150 (1967), p. 427 (cit. on p. 4).
- [29] S. L. Brady, A. T. Trout, E. Somasundaram, C. G. Anton, Y. Li, and J. R. Dillman. “Improving image quality and reducing radiation dose for pediatric CT by using deep learning reconstruction”. *Radiology* 298.1 (2021), pp. 180–188 (cit. on p. 102).
- [30] A. Breger, A. Biguri, M. S. Landman, I. Selby, N. Amberg, E. Brunner, J. Gröhl, S. Hatamikia, C. Karner, L. Ning, et al. “A study of why we need to reassess full reference image quality assessment with medical images”. *arXiv preprint arXiv:2405.19097* (2024) (cit. on p. 90).
- [31] A. Breger, C. Karner, I. Selby, J. Gröhl, S. Dittmer, E. Lilley, J. Babar, J. Beckford, T. J. Sadler, S. Shahipasand, et al. “A study on the adequacy of common IQA measures for medical images”. *arXiv preprint arXiv:2405.19224* (2024) (cit. on p. 90).
- [32] D. J. Brenner and E. J. Hall. “Computed tomography—an increasing source of radiation exposure”. *New England journal of medicine* 357.22 (2007), pp. 2277–2284 (cit. on p. 81).
- [33] T. M. Buzug. *Computed tomography*. Springer, 2011 (cit. on p. 74).
- [34] F. Casali. “Chapter 2 X-ray and neutron digital radiography and computed tomography for cultural heritage”. In: ed. by D. Bradley and D. Creagh. Vol. 1. *Physical Techniques in the Study of Art, Archaeology and Cultural Heritage*. Elsevier, 2006, pp. 41–123 (cit. on p. 38).
- [35] E. Van de Casteele, D. Van Dyck, J. Sijbers, and E. Raman. “An energy-based beam hardening model in tomography”. *Physics in Medicine & Biology* 47.23 (2002), p. 4181 (cit. on pp. 11, 40, 69, 110, 118).

- [36] M. Cellina, M. Cè, G. Irmici, V. Ascenti, E. Caloro, L. Bianchi, G. Pellegrino, N. D’Amico, S. Papa, and G. Carrafiello. “Artificial intelligence in emergency radiology: where are we going?” *Diagnostics* 12.12 (2022), p. 3223 (cit. on p. 14).
- [37] A. Chambolle and T. Pock. “A first-order primal-dual algorithm for convex problems with applications to imaging”. *Journal of Mathematical Imaging and Vision* 40 (2011), pp. 120–145 (cit. on p. 113).
- [38] G. Chen, Z. Zhang, S. Xu, S. Jiang, X. Liu, P. Tang, S. Li, and X. Xiang. “An optimized filter design approach for enhancing imaging quality in industrial linear accelerator”. *Journal of X-Ray Science and Technology* Preprint (2024), pp. 1–14 (cit. on p. 132).
- [39] G. Chen and D. Needell. “Compressed sensing and dictionary learning”. *Finite Frame Theory: A Complete Introduction to Overcompleteness* 73 (2016), p. 201 (cit. on p. 107).
- [40] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang. “Low-dose CT with a residual encoder-decoder convolutional neural network”. *IEEE Transactions on Medical Imaging* 36.12 (2017), pp. 2524–2535 (cit. on p. 102).
- [41] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye. “Diffusion Posterior Sampling for General Noisy Inverse Problems”. In: *The Eleventh International Conference on Learning Representations, ICLR 2023*. The International Conference on Learning Representations. 2023 (cit. on p. 133).
- [42] H. Chung, B. Sim, D. Ryu, and J. C. Ye. “Improving diffusion models for inverse problems using manifold constraints”. *Advances in Neural Information Processing Systems* 35 (2022), pp. 25683–25696 (cit. on p. 133).
- [43] S. B. Coban, F. Lucka, W. J. Palenstijn, D. Van Loo, and K. J. Batenburg. “Explorative imaging and its implementation at the FleX-ray Laboratory”. *Journal of Imaging* 6.4 (2020), p. 18 (cit. on pp. 25, 43, 49, 64).
- [44] C. B. daveynin. *GE LightSpeed CT scanner at Open House, Monroeville, Pennsylvania*. URL: <https://www.flickr.com/photos/daveynin/7417096292/> (visited on Nov. 26, 2024) (cit. on p. 2).
- [45] B. De Man, S. Basu, N. Chandra, B. Dunham, P. Edic, M. Iatrou, S. McOlash, P. Sainath, C. Shaughnessy, B. Tower, et al. “CatSim: a new computer assisted tomography simulation environment”. In: *Medical Imaging 2007: Physics of Medical Imaging*. Vol. 6510. SPIE. 2007, pp. 856–863 (cit. on p. 85).
- [46] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. “ImageNet: A large-scale hierarchical image database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2009, pp. 248–255 (cit. on pp. 15, 62, 101, 133).

- [47] H. Der Sarkissian, F. Lucka, M. van Eijnatten, G. Colacicco, S. Coban, and K. J. Batenburg. “A cone-beam X-ray computed tomography data collection designed for machine learning”. *Scientific Data* 6 (Oct. 2019), p. 215 (cit. on pp. 62, 103).
- [48] A. G. Dimakis, A. Bora, D. Van Veen, A. Jalal, S. Vishwanath, and E. Price. “Deep generative models and inverse problems”. *Mathematical Aspects of Deep Learning* (2022), pp. 400–421 (cit. on p. 107).
- [49] M. Domínguez-Delmás, F. Bossema, B. Van der Mark, A. Kostenko, S. Coban, S. Van Daalen, P. Van Duin, and K. Batenburg. “Dating and provenancing the Woman with lantern sculpture – A contribution towards attribution of Netherlandish art”. *Journal of Cultural Heritage* 50 (2021), pp. 179–187 (cit. on p. 38).
- [50] J. Dorscheid, F. G. Bossema, P. van Duin, S. B. Coban, R. van Liere, K. J. Batenburg, and G. P. Di Stefano. “Looking under the skin: multi-scale CT scanning of a peculiarly constructed cornett in the Rijksmuseum”. *Heritage Science* 10.1 (2022), p. 161 (cit. on p. 38).
- [51] A. Ebner and M. Haltmeier. “Plug-and-play image reconstruction is a convergent regularization method”. *IEEE Transactions on Image Processing* (2024) (cit. on p. 107).
- [52] A. J. Einstein, M. J. Henzlova, and S. Rajagopalan. “Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography”. *Jama* 298.3 (2007), pp. 317–323 (cit. on p. 81).
- [53] I. A. Elbakri and J. A. Fessler. “Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography”. In: *Medical Imaging 2003: Image Processing*. Vol. 5032. SPIE. 2003, pp. 1839–1850 (cit. on p. 85).
- [54] I. A. Elbakri and J. A. Fessler. “Statistical image reconstruction for polyenergetic X-ray computed tomography”. *IEEE transactions on medical imaging* 21.2 (2002), pp. 89–99 (cit. on p. 85).
- [55] H. W. Engl, M. Hanke-Bourgeois, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications <Dordrecht> 375. Dordrecht: Kluwer Acad. Publ, 2000 (cit. on pp. 4, 105).
- [56] W. Ertel. *Introduction to artificial intelligence*. Springer Nature, 2024 (cit. on p. 13).
- [57] Ethan A Merritt. *X-ray Absorption Edges*. 1996-2010 (cit. on p. 34).
- [58] K Faulkner and B. Moores. “Noise and contrast detection in computed tomography images”. *Physics in Medicine & Biology* 29.4 (1984), p. 329 (cit. on p. 83).
- [59] L. Feldkamp, L. David, and J. Kress. “Practical cone-beam algorithm”. *J Opt Soc Am A* 1 (1984), pp. 612–619 (cit. on pp. 5, 6, 45).

- [60] D. Frank, J. Jelten, and T. Lasser. “elsa: an elegant framework for tomographic reconstruction”. *Journal of Open Source Software* 9.94 (2024), p. 6174 (cit. on p. 133).
- [61] P. Fried, J. Woodward, D. Brown, D. Harvell, and J. Hanken. “3D scanning of antique glass by combining photography and computed tomography”. *Digital Applications in Archaeology and Cultural Heritage* 18 (2020), e00147 (cit. on pp. 38, 49).
- [62] J. Friel and E. T. Quinto. “Characterization and reduction of artifacts in limited angle tomography”. *Inverse Problems* 29.12 (2013), p. 125007 (cit. on p. 109).
- [63] D. P. Frush, C. C. Slack, C. L. Hollingsworth, G. S. Bisset, L. F. Donnelly, J. Hsieh, T. Lavin-Wensell, and J. R. Mayo. “Computer-simulated radiation dose reduction for abdominal multidetector CT of pediatric patients”. *American Journal of Roentgenology* 179.5 (2002), pp. 1107–1113 (cit. on p. 84).
- [64] L. Fu and B. De Man. “Deep learning tomographic reconstruction through hierarchical decomposition of domain transforms”. *Visual Computing for Industry, Biomedicine, and Art* 5.1 (2022), p. 30 (cit. on p. 106).
- [65] F. Gao and L. Han. “Implementing the Nelder-Mead simplex algorithm with adaptive parameters”. *Computational Optimization and Applications* 51 (May 2012), pp. 259–277 (cit. on p. 75).
- [66] R. Garnett. “A comprehensive review of dual-energy and multi-spectral computed tomography”. *Clinical imaging* 67 (2020), pp. 160–169 (cit. on p. 3).
- [67] M. Genzel, J. Macdonald, and M. März. “Solving Inverse Problems With Deep Neural Networks – Robustness Included?” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.1 (2022), pp. 1119–1134 (cit. on p. 124).
- [68] D. T. Ginat and R. Gupta. “Advances in computed tomography imaging technology”. *Annual Review of Biomedical Engineering* 16 (2014), pp. 431–453 (cit. on p. 103).
- [69] B. Girod. “Psychovisual aspects of image communication”. *Signal Processing* 28.3 (1992), pp. 239–251 (cit. on pp. 90, 111).
- [70] L. W. Goldman. “Principles of CT: Radiation Dose and Image Quality”. *Journal of Nuclear Medicine Technology* 35.4 (2007), pp. 213–225. eprint: <https://tech.snmjournals.org/content/35/4/213.full.pdf> (cit. on pp. 38, 70).
- [71] H. Gong, L. Ren, S. S. Hsieh, C. H. McCollough, and L. Yu. “Deep learning enabled ultra-fast-pitch acquisition in clinical x-ray computed tomography”. *Medical physics* 48.10 (2021), pp. 5712–5726 (cit. on p. 102).
- [72] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016 (cit. on p. 13).
- [73] D. Gourdeau, S. Duchesne, and L. Archambault. “On the proper use of structural similarity for the robust evaluation of medical image synthesis models”. *Medical Physics* 49.4 (2022), pp. 2462–2474 (cit. on p. 90).

- [74] K. Gregor and Y. LeCun. “Learning fast approximations of sparse coding”. In: *Proceedings of the 27th International Conference on Machine Learning*. 2010, pp. 399–406 (cit. on p. 106).
- [75] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser. “CNN-based projected gradient descent for consistent CT image reconstruction”. *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1440–1453 (cit. on p. 106).
- [76] A. Habring and M. Holler. “Neural-network-based regularization methods for inverse problems in imaging”. *GAMM-Mitteilungen* (2024), e202470004 (cit. on p. 107).
- [77] H. Hamba, T. Nikaido, A. Sadr, S. Nakashima, and J. Tagami. “Enamel Lesion Parameter Correlations between Polychromatic Micro-CT and TMR”. *Journal of Dental Research* 91.6 (2012). PMID: 22476867, pp. 586–591 (cit. on pp. 30, 42).
- [78] Y. Han and J. C. Ye. “Framing U-Net via deep convolutional framelets: Application to sparse-view CT”. *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1418–1429 (cit. on p. 106).
- [79] P. C. Hansen, J. Jørgensen, and W. R. B. Lionheart. *Computed Tomography: Algorithms, Insight, and Just Enough Theory*. Ed. by P. C. Hansen, J. Jørgensen, and W. R. B. Lionheart. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2021 (cit. on pp. 2–4, 37, 113).
- [80] A. Hauptmann, S. Mukherjee, C.-B. Schönlieb, and F. Sherry. “Convergent Regularization in Inverse Problems and Linear Plug-and-Play Denoisers”. *Foundations of Computational Mathematics* (2024) (cit. on p. 107).
- [81] A. A. Hendriksen, D. Schut, W. J. Palenstijn, N. Viganó, J. Kim, D. M. Pelt, T. Van Leeuwen, and K. J. Batenburg. “Tomosipo: fast, flexible, and convenient 3D tomography for complex scanning geometries in Python”. *Optics Express* 29.24 (2021), pp. 40494–40513 (cit. on pp. 99, 110, 112).
- [82] J. Hsieh. “Adaptive streak artifact reduction in computed tomography resulting from excessive X-ray photon noise”. *Medical Physics* 25.11 (1998), pp. 2139–2147 (cit. on pp. 81, 84).
- [83] J. Hsieh. *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*. Washington: SPIE press, 2003 (cit. on pp. 1, 37, 83).
- [84] J. Hsieh and T. Flohr. “Computed tomography recent history and future perspectives”. *Journal of Medical Imaging* 8.5 (2021), pp. 052109–052109 (cit. on p. 103).
- [85] S. Hurault, A. Leclaire, and N. Papadakis. “Gradient Step Denoiser for convergent Plug-and-Play”. In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022 (cit. on pp. 113, 114).
- [86] A. Jansson, J. Ekengren, A.-R. Zekavat, and L. Pejryd. “Effects of x-ray penetration depth on multi material computed tomography measurements”. In: *6th Conference on Industrial Computed Tomography*. 2016 (cit. on p. 38).

- [87] R. J. Jennings. “A method for comparing beam-hardening filter materials for diagnostic radiology”. *Medical physics* 15.4 (1988), pp. 588–599 (cit. on pp. 30, 42, 57, 130).
- [88] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. “Deep convolutional neural network for inverse problems in imaging”. *IEEE Transactions on Image Processing* 26.9 (2017), pp. 4509–4522 (cit. on pp. 102, 106).
- [89] P. M. Joseph and R. A. Schulz. “View sampling requirements in fan beam computed tomography”. *Medical Physics* 7.6 (1980), pp. 692–702 (cit. on pp. 7, 110, 117).
- [90] U. S. Kamilov, C. A. Bouman, G. T. Buzzard, and B. Wohlberg. “Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications”. *IEEE Signal Processing Magazine* 40.1 (2023), pp. 85–97 (cit. on p. 107).
- [91] E. Kang, J. Min, and J. C. Ye. “A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction”. *Medical Physics* 44.10 (2017), e360–e375 (cit. on pp. 106, 133).
- [92] R. A. Ketcham and W. D. Carlson. “Acquisition, optimization and interpretation of X-ray computed tomographic imagery: applications to the geosciences”. *Computers & Geosciences* 27.4 (2001). 3D reconstruction, modelling & visualization of geological materials, pp. 381–400 (cit. on pp. 57, 130).
- [93] F. M. Khan and J. P. Gibbons. *Khan’s the physics of radiation therapy*. Philadelphia, PA: Lippincott Williams & Wilkins, 2014 (cit. on pp. 34, 43, 70).
- [94] F. Kharfi. “Mathematics and Physics of Computed Tomography (CT): Demonstrations and Practical Examples”. In: *Imaging and Radioanalytical Techniques in Interdisciplinary Research*. Ed. by F. Kharfi. Rijeka: IntechOpen, 2013. Chap. 4 (cit. on pp. 9, 33, 46, 71).
- [95] A. Khorashadizadeh, V. Debarnot, T. Liu, and I. Dokmanić. “GLIMPSE: Generalized Local Imaging with MLPs”. *arXiv preprint arXiv:2401.00816* (2024) (cit. on p. 125).
- [96] E. Kim, K. Muroi, T. Koike, and J. Kim. “Dose Reduction and Image Quality Optimization of Pediatric Chest Radiography Using a Tungsten Filter”. *Bioengineering* 9.10 (2022) (cit. on pp. 30, 42).
- [97] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. *arXiv preprint arXiv:1412.6980* (2014) (cit. on pp. 91, 115).
- [98] M. Kiss, S. Coban, K. Batenburg, T. van Leeuwen, and F. Lucka. “2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning”. *Scientific Data* 10.1 (2023), p. 576 (cit. on pp. 61, 82, 87, 102, 103, 109, 111, 112, 115, 118, 155).
- [99] M. B. Kiss, A. Biguri, C.-B. Schönlieb, K. J. Batenburg, and F. Lucka. “Learned denoising with simulated and experimental low-dose CT data”. *arXiv preprint arXiv:2408.08115* (2024) (cit. on pp. 81, 155).

- [100] M. B. Kiss, A. Biguri, Z. Shumaylov, F. Sherry, K. J. Batenburg, C.-B. Schönlieb, and F. Lucka. “Benchmarking learned algorithms for computed tomography image reconstruction tasks”. *Applied Mathematics for Modern Challenges* 3.0 (2025), pp. 1–43 (cit. on pp. 101, 155).
- [101] M. B. Kiss, F. G. Bossema, P. J. van Laar, S. Meijer, F. Lucka, T. van Leeuwen, and K. J. Batenburg. “Beam filtration for object-tailored X-ray CT of multi-material cultural heritage objects”. *Heritage Science* 11.1 (2023), p. 130 (cit. on pp. 37, 155).
- [102] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 1-1,000*. June 2023 (cit. on pp. 76, 82, 126, 155).
- [103] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 1-1,000 (reference reconstructions and segmentations)*. June 2023 (cit. on pp. 76, 82, 156).
- [104] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 1,001-2,000*. June 2023 (cit. on pp. 76, 82, 155).
- [105] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 1,001-2,000 (reference reconstructions and segmentations)*. June 2023 (cit. on pp. 76, 82, 156).
- [106] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 2,001-3,000*. June 2023 (cit. on pp. 76, 82, 155).
- [107] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 2,001-3,000 (reference reconstructions and segmentations)*. June 2023 (cit. on pp. 76, 82, 156).
- [108] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 3,001-4,000*. June 2023 (cit. on pp. 76, 82, 155).
- [109] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 3,001-4,000 (reference reconstructions and segmentations)*. June 2023 (cit. on pp. 76, 82, 156).

- [110] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 4,001-5,000*. June 2023 (cit. on pp. 76, 82, 155).
- [111] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 4,001-5,000 (reference reconstructions and segmentations)*. June 2023 (cit. on pp. 76, 82, 156).
- [112] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices OOD*. June 2023 (cit. on pp. 76, 82, 155).
- [113] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices OOD (reference reconstructions and segmentations)*. June 2023 (cit. on pp. 76, 82, 156).
- [114] M. B. Kiss, H. Der Sarkissian, and F. Lucka. *2DeteCT - Collection of Python scripts for loading, pre-processing, reconstructing and segmenting X-ray CT projection data of the 2DeteCT data collection*. <https://github.com/mbkiss/2DeteCTcodes/>. Accessed: 2025-09-26. June 2023 (cit. on p. 156).
- [115] A. Klein, S. Wallkötter, S. Silvester, A. Tanbakuchi, actions user, P. Müller, J. Nunez-Iglesias, M. Harfouche, Dennis, A. Lee, and et al. *imageio/imageio: v2.26.0*. Zenodo, Feb. 2023 (cit. on p. 77).
- [116] F. Knoll, J. Zbontar, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdalv, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui. “fastMRI: A Publicly Available Raw k-Space and DICOM Dataset of Knee Images for Accelerated MR Image Reconstruction Using Machine Learning”. *Radiology: Artificial Intelligence* 2.1 (2020). PMID: 32076662, e190007. eprint: <https://doi.org/10.1148/ryai.2020190007> (cit. on pp. 62, 103).
- [117] E. Kobler, A. Effland, K. Kunisch, and T. Pock. “Total deep variation for linear inverse problems”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 7549–7558 (cit. on pp. 113, 114, 116).
- [118] A. Kostenko, W. Palenstijn, S. Coban, A. Hendriksen, R. van Liere, and K. Batenburg. “Prototyping X-ray tomographic reconstruction pipelines with Flexbox”. *SoftwareX* 11 (2020), p. 100364 (cit. on pp. 45, 46).
- [119] A. Krizhevsky, G. Hinton, et al. “Learning multiple layers of features from tiny images”. <https://www.cs.toronto.edu/kriz/learning-features-2009-TR.pdf> (2009) (cit. on pp. 15, 101, 133).

- [120] S. Kuanar, V. Athitsos, D. Mahapatra, K. Rao, Z. Akhtar, and D. Dasgupta. “Low dose abdominal CT image reconstruction: an unsupervised learning based approach”. In: *2019 IEEE international conference on image processing (ICIP)*. IEEE. 2019, pp. 1351–1355 (cit. on p. 133).
- [121] G. M. Lasio, B. R. Whiting, and J. F. Williamson. “Statistical reconstruction for X-ray computed tomography using energy-integrating detectors”. *Physics in medicine & biology* 52.8 (2007), p. 2247 (cit. on pp. 84, 85).
- [122] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning”. *Nature* 521.7553 (2015), pp. 436–444 (cit. on pp. 13, 101).
- [123] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. “Gradient-based learning applied to document recognition”. *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324 (cit. on pp. 15, 62, 101, 133).
- [124] J. Leuschner, M. Schmidt, D. O. Baguer, and P. Maass. “LoDoPaB-CT, a benchmark dataset for low-dose computed tomography reconstruction”. *Scientific Data* 8.1 (2021), pp. 1–12 (cit. on pp. 62, 85, 103).
- [125] J. Leuschner, M. Schmidt, P. S. Ganguly, V. Andriiashen, S. B. Coban, A. Denker, D. Bauer, A. Hadjifaradji, K. J. Batenburg, P. Maass, et al. “Quantitative comparison of deep learning-based image reconstruction methods for low-dose and sparse-angle CT applications”. *Journal of Imaging* 7.3 (2021), p. 44 (cit. on pp. 89, 102).
- [126] H. Li, J. Schwab, S. Antholzer, and M. Haltmeier. “NETT: Solving inverse problems with deep neural networks”. *Inverse Problems* 36.6 (2020), p. 065005 (cit. on pp. 102, 107).
- [127] T. Li, X. Li, J. Wang, J. Wen, H. Lu, J. Hsieh, and Z. Liang. “Nonlinear sinogram smoothing for low-dose X-ray CT”. *IEEE Transactions on Nuclear Science* 51.5 (2004), pp. 2505–2513 (cit. on pp. 81, 84).
- [128] H. Liao, W.-A. Lin, S. K. Zhou, and J. Luo. “ADN: artifact disentanglement network for unsupervised metal artifact reduction”. *IEEE Transactions on Medical Imaging* 39.3 (2019), pp. 634–643 (cit. on p. 133).
- [129] P.-S. Liao, T.-S. Chen, and P. C. Chung. “A Fast Algorithm for Multilevel Thresholding”. *J. Inf. Sci. Eng.* 17 (2001), pp. 713–727 (cit. on p. 75).
- [130] Z. Liao, S. Hu, M. Li, W. Chen, et al. “Noise estimation for single-slice sinogram of low-dose X-ray computed tomography using homogenous patch”. *Mathematical Problems in Engineering* 2012 (2012) (cit. on p. 84).
- [131] D. Limited. *X-ray Detector Product Specifications DEXELA 1512 CMOS*. June 2010 (cit. on pp. 25, 44, 64).
- [132] *LION - Learned Iterative Optimization Networks*. <https://github.com/Cambria/LION/>. Accessed: 2024-16-06. 2024 (cit. on p. 89).

- [133] J. Liu, R. Anirudh, J. J. Thiagarajan, S. He, K. A. Mohan, U. S. Kamilov, and H. Kim. “DOLCE: A model-based probabilistic diffusion framework for limited-angle ct reconstruction”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 10498–10508 (cit. on pp. 114, 133).
- [134] S. Lunz, O. Öktem, and C.-B. Schönlieb. “Adversarial Regularizers in Inverse Problems”. In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Vol. 31. Curran Associates, Inc., 2018 (cit. on p. 113).
- [135] V. Macchi, E. Picardi, A. Porzionato, A. Morra, L. Tabarin, F. Gusella, B. Grignon, and R. Caro. “Friar Leopold Mandic (1866–1942): the computed tomography of the body of a saint”. *Surgical and Radiologic Anatomy* 40 (Aug. 2018), pp. 1–9 (cit. on p. 38).
- [136] S. Maruyama. “Properties of the SSIM metric in medical image assessment: correspondence between measurements and the spatial frequency spectrum”. *Physical and Engineering Sciences in Medicine* 46.3 (2023), pp. 1131–1141 (cit. on p. 90).
- [137] P. Massoumzadeh, S. Don, C. F. Hildebolt, K. T. Bae, and B. R. Whiting. “Validation of CT dose-reduction simulation”. *Medical physics* 36.1 (2009), pp. 174–189 (cit. on pp. 84, 85).
- [138] P. Massoumzadeh, O. A. Earl, and B. R. Whiting. “Noise simulation in X-ray CT”. In: *Medical Imaging 2005: Physics of Medical Imaging*. Vol. 5745. SPIE. 2005, pp. 898–909 (cit. on pp. 84, 85).
- [139] MATLAB. *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010 (cit. on p. 77).
- [140] J. R. Mayo, K. P. Whittall, A. N. Leung, T. E. Hartman, C. S. Park, S. L. Primack, G. K. Chambers, M. K. Limkeman, T. L. Toth, and S. H. Fox. “Simulated dose reduction in conventional chest CT: validation study.” *Radiology* 202.2 (1997), pp. 453–457 (cit. on pp. 81, 84).
- [141] C McCollough. “TU-FG-207A-04: overview of the low dose CT grand challenge”. *Medical physics* 43.6Part35 (2016), pp. 3759–3760 (cit. on pp. 62, 85, 103, 104).
- [142] A. Meloni, F. Frijia, D. Panetta, G. Degiorgi, C. De Gori, E. Maffei, A. Clemente, V. Positano, and F. Cademartiri. “Photon-Counting Computed Tomography (PCCT): Technical Background and Cardio-Vascular Applications”. *Diagnostics* 13.4 (2023) (cit. on p. 41).
- [143] F. Mihanović, I. Jerković, I. Kružić, S. Andjelinović, S. Janković, and Ž. Bašić. “From Biography to Osteobiography: An Example of Anthropological Historical Identification of the Remains of St. Paul”. *The Anatomical Record* 300.9 (2017), pp. 1535–1546 (cit. on p. 38).
- [144] T. R. Moen, B. Chen, D. R. Holmes III, X. Duan, Z. Yu, L. Yu, S. Leng, J. G. Fletcher, and C. H. McCollough. “Low-dose CT image and projection dataset”. *Medical physics* 48.2 (2021), pp. 902–911 (cit. on pp. 62, 85, 103, 104).

- [145] V. Monga, Y. Li, and Y. C. Eldar. “Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing”. *IEEE Signal Processing Magazine* 38.2 (2021), pp. 18–44 (cit. on p. 106).
- [146] L. Montaina, S. Longo, G. Galotta, G. Tranquilli, R. Saccuman, and S. Capuani. “Assessment of the panel support of a seventeenth-century Dutch painting by clinical multislice computed tomography”. *Studies in Conservation* 66.3 (2021), pp. 174–181 (cit. on p. 38).
- [147] S Mukherjee, S Dittmer, Z Shumaylov, S Lunz, O Öktem, and C.-B. Schönlieb. “Data-Driven Convex Regularizers for Inverse Problems”. In: *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2024, pp. 13386–13390 (cit. on pp. 113, 114).
- [148] S. Mukherjee, S. Dittmer, Z. Shumaylov, S. Lunz, O. Öktem, and C.-B. Schönlieb. “Learned convex regularizers for inverse problems”. *arXiv preprint arXiv:2008.02839* (2020) (cit. on pp. 113, 114, 116).
- [149] S. Mukherjee, A. Hauptmann, O. Öktem, M. Pereyra, and C.-B. Schönlieb. “Learned reconstruction methods with convergence guarantees: a survey of concepts and applications”. *IEEE Signal Processing Magazine* 40.1 (2023), pp. 164–182 (cit. on p. 107).
- [150] F. Natterer. *The mathematics of computerized tomography*. SIAM, 2001 (cit. on p. 5).
- [151] Y. E. Nesterov. “A method of solving a convex programming problem with convergence rate $O(k^2)$ ”. In: *Doklady Akademii Nauk*. Vol. 269. Russian Academy of Sciences. 1983, pp. 543–547 (cit. on pp. 74, 112, 113).
- [152] S. I. Nikolenko. “Synthetic-to-real domain adaptation and refinement”. In: *Synthetic data for deep learning*. Springer, 2021, pp. 235–268 (cit. on pp. 15, 99, 102).
- [153] D. Nykonenko, O. Yatsuk, L. Guidorzi, A. Lo Giudice, F. Tansella, L. P. Cesareo, G. Sorrentino, P. Davit, M. Gulmini, and A. Re. “Glass beads from a Scythian grave on the island of Khortytsia (Zaporizhzhia, Ukraine): insights into bead making through 3D imaging”. *Heritage Science* 11.1 (2023), p. 238 (cit. on p. 132).
- [154] W. Palenstijn, K. Batenburg, and J. Sijbers. “Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs)”. *Journal of Structural Biology* 176.2 (2011), pp. 250–253 (cit. on pp. 99, 110, 112).
- [155] X. Pan, E. Y. Sidky, and M. Vannier. “Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?” *Inverse problems* 25.12 (2009), p. 123009 (cit. on p. 83).
- [156] X. Pan, J. Siewerdsen, P. J. La Riviere, and W. A. Kalender. “Anniversary Paper: Development of x-ray computed tomography: The role of Medical Physics and AAPM from the 1970s to present”. *Medical Physics* 35.8 (2008), pp. 3728–3739 (cit. on p. 38).

- [157] H. S. Park, S. M. Lee, H. P. Kim, J. K. Seo, and Y. E. Chung. “CT sinogram-consistency learning for metal-induced beam hardening correction”. *Medical Physics* 45.12 (2018), pp. 5376–5384 (cit. on p. 102).
- [158] J. A. Patton and T. G. Turkington. “SPECT/CT physical principles and attenuation correction”. *Journal of Nuclear Medicine Technology* 36.1 (2008), pp. 1–10 (cit. on pp. 11, 40, 69, 110, 118).
- [159] D. M. Pelt, K. J. Batenburg, and J. A. Sethian. “Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks”. *Journal of Imaging* 4.11 (2018), p. 128 (cit. on p. 89).
- [160] D. M. Pelt and J. A. Sethian. “A mixed-scale dense convolutional neural network for image analysis”. *Proceedings of the National Academy of Sciences* 115.2 (2018), pp. 254–259 (cit. on pp. 21, 82, 89, 91, 113).
- [161] A. N. Primak, J. C. R. Giraldo, C. D. Eusemann, B. Schmidt, B. Kantor, J. G. Fletcher, and C. H. McCollough. “Dual-source dual-energy CT with additional tin filtration: Dose and image quality evaluation in phantoms and in-vivo”. *AJR. American journal of roentgenology* 195.5 (2010), p. 1164 (cit. on p. 70).
- [162] T. Prohaszka, L. Neumann, and M. Haltmeier. “Derivative-Free Iterative One-Step Reconstruction for Multispectral CT”. *Journal of Imaging* 10.5 (2024), p. 98 (cit. on p. 132).
- [163] J. Radon. “1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten”. *Classic papers in modern diagnostic radiology* 5.21 (2005), p. 124 (cit. on p. 4).
- [164] J. Radon. “On the determination of functions from their integral values along certain manifolds”. *IEEE transactions on medical imaging* 5.4 (1986), pp. 170–176 (cit. on p. 4).
- [165] G. Ramachandran and A. Lakshminarayanan. “Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms”. *Proceedings of the National Academy of Sciences* 68.9 (1971), pp. 2236–2240 (cit. on p. 4).
- [166] S. Ravishankar, J. C. Ye, and J. A. Fessler. “Image reconstruction: From sparsity to data-adaptive methods and machine learning”. *Proceedings of the IEEE* 108.1 (2019), pp. 86–109 (cit. on pp. 16, 61, 106).
- [167] A. Re, F. Albertin, C. Avataneo, R. Brancaccio, G. Corsi J. ad Cotto, S. De Blasi, G. Dughera, E. Durisi, W. Ferrarese, A. Giovagnoli, N. Grassi, A. Lo Giudice, P. Mereu, G. Mila, M. Nervo, N. Pastrone, F. Prino, L. Ramello, M. Ravera, C. Ricci, A. Romero, R. Sacchi, A. Staiano, L. Visca, and L. Zamprota. “X-ray tomography of large wooden artworks: the case study of “Doppio corpo” by Pietro Piffetti”. *Heritage Science* 2 (2014), p. 19 (cit. on p. 38).

- [168] J. Reid, J. Gamberoni, F. Dong, and W. Davros. “Optimization of kVp and mAs for Pediatric Low-Dose Simulated Abdominal CT: Is It Best to Base Parameter Selection on Object Circumference?” *American Journal of Roentgenology* 195.4 (2010). PMID: 20858833, pp. 1015–1020. eprint: <https://doi.org/10.2214/AJR.09.3862> (cit. on pp. 71, 118).
- [169] Rijksmuseum Amsterdam. *BK-KOG-29 - Beurs van groen fluweel met trekkoord en afgezet met gouddraad op een doosje in de vorm van een schaar (fallus?) van dito fluweel en passement, voorzien van twee metalen ogen*. na 1580 (cit. on pp. 46, 47).
- [170] Rijksmuseum Amsterdam. *BK-NM-3086 - Langwerpig, taps toelopend foedraal met dop, van leer verdeeld in drie compartimenten, bekleed met paars fluweel en geborduurd met goud- en zilverdraad, zoetwaterparels, en zilveren lovers*. c. 1600 - c. 1625 (cit. on pp. 46, 47).
- [171] E. L. Ritman. “Current status of developments and applications of micro-CT”. *Annual Review of Biomedical Engineering* 13 (2011), pp. 531–552 (cit. on p. 103).
- [172] Á. Rodríguez-Sánchez, A. Thompson, L. Körner, N. Brierley, and R. Leach. “Review of the influence of noise in X-ray computed tomography measurement uncertainty”. *Precision Engineering* 66 (2020), pp. 382–391 (cit. on pp. 33, 88).
- [173] O. Ronneberger, P. Fischer, and T. 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 (cit. on pp. 82, 89, 113).
- [174] M. Salamon, M. Arzig, P. J. Wellmann, and N. Uhlmann. “Comparison of achievable contrast features in computed tomography observing the growth of a 4H-SiC bulk crystal”. *Materials* 12.22 (2019), p. 3652 (cit. on p. 41).
- [175] A. L. Samuel. “Some studies in machine learning using the game of checkers”. *IBM Journal of research and development* 3.3 (1959), pp. 210–229 (cit. on p. 13).
- [176] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*. Springer, 2009 (cit. on pp. 4, 105, 124).
- [177] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. “NIH Image to ImageJ: 25 years of image analysis”. *Nature methods* 9.7 (2012), pp. 671–675 (cit. on p. 77).
- [178] M. Seracini, M. Bettuzzi, R. Brancaccio, and M. P. Morigi. “Calibration-free 3D ray-tracing beam hardening correction in computed tomography”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1062 (2024), p. 169226 (cit. on p. 132).

- [179] H. Shan, A. Padole, F. Homayounieh, U. Kruger, R. D. Khera, C. Nitiwarangkul, M. K. Kalra, and G. Wang. “Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction”. *Nature Machine Intelligence* 1.6 (2019), pp. 269–276 (cit. on p. 102).
- [180] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. “A statistical evaluation of recent full reference image quality assessment algorithms”. *IEEE Transactions on image processing* 15.11 (2006), pp. 3440–3451 (cit. on p. 111).
- [181] J. Shi, O. Elkilany, A. Fischer, A. Suppes, D. M. Pelt, and K. J. Batenburg. “LoDoInd: Introducing A Benchmark Low-dose Industrial CT Dataset and Enhancing Denoising with 2.5 D Deep Learning Techniques”. *preprint* (2024) (cit. on p. 133).
- [182] Z. Shumaylov, J. Budd, S. Mukherjee, and C.-B. Schönlieb. “Provably Convergent Data-Driven Convex-Nonconvex Regularization”. In: *NeurIPS 2023 Workshop on Deep Learning and Inverse Problems*. 2023 (cit. on p. 5).
- [183] Z. Shumaylov, J. Budd, S. Mukherjee, and C.-B. Schönlieb. *Weakly Convex Regularisers for Inverse Problems: Convergence of Critical Points and Primal-Dual Optimisation*. arXiv:2402.01052. 2024. arXiv: 2402.01052 [math.OA] (cit. on p. 107).
- [184] S. S. Skiena. *The data science design manual*. Springer, 2017 (cit. on p. 15).
- [185] N. Sodini, D. Dreossi, A. Giordano, J. Kaiser, F. Zanini, and T. Zikmund. “Comparison of different experimental approaches in the tomographic analysis of ancient violins”. *Journal of Cultural Heritage* 27 (2017), S88–S92 (cit. on p. 38).
- [186] Y. Song, L. Shen, L. Xing, and S. Ermon. “Solving Inverse Problems in Medical Imaging with Score-Based Generative Models”. In: *International Conference on Learning Representations*. 2022 (cit. on pp. 114, 133).
- [187] J. Tachella, D. Chen, S. Hurault, M. Terris, and A. Wang. *DeepInverse: A deep learning framework for inverse problems in imaging*. Version latest. June 2023 (cit. on p. 110).
- [188] M. F. Tarplee and N. Corps. “Acquiring optimal quality X-ray μ CT scans”. *Application note* 1 (2008), pp. 93–99 (cit. on p. 42).
- [189] E. Valat, A. Hauptmann, and O. Öktem. “Self-Supervised Denoiser Framework”. *arXiv preprint arXiv:2411.19593* (2024) (cit. on p. 133).
- [190] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers. “The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography”. *Ultramicroscopy* 157 (2015), pp. 35–47 (cit. on pp. 99, 110, 112).
- [191] G. Van Rossum and F. L. Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995 (cit. on p. 77).

- [192] M. Vandenbeusch and B. O’Flynn D. abd Moreno. “Layer by Layer: The Manufacture of Graeco-Roman Funerary Masks”. *The Journal of Egyptian Archaeology* 107 (2021), 281–98 (cit. on p. 38).
- [193] A. Vandivere, J. Wadum, K. J. Van den Berg, and A. van Loon. “From ‘Vermeer Illuminated’ to ‘The Girl in the Spotlight’: approaches and methodologies for the scientific (re-)examination of Vermeer’s Girl with a Pearl Earring”. *Heritage Science* 7 (Aug. 2019), pp. 1–14 (cit. on p. 38).
- [194] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg. “Plug-and-play priors for model based reconstruction”. In: *2013 IEEE Global Conference on Signal and Information Processing*. IEEE. 2013, pp. 945–948 (cit. on p. 107).
- [195] A. K. Venkataramanan, C. Wu, A. C. Bovik, I. Katsavounidis, and Z. Shahid. “A hitchhiker’s guide to structural similarity”. *IEEE Access* 9 (2021), pp. 28872–28896 (cit. on p. 111).
- [196] H. Villarraga-Gómez, E. Herazo, and S. Smith. “X-ray computed tomography: from medical imaging to dimensional metrology”. *Precision Engineering* 60 (Nov. 2019), pp. 544–569 (cit. on p. 37).
- [197] C. Wang, K. Shang, H. Zhang, Q. Li, and S. K. Zhou. “DuDoTrans: dual-domain transformer for sparse-view CT reconstruction”. In: *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer. 2022, pp. 84–94 (cit. on p. 133).
- [198] D. Wang, F. Fan, Z. Wu, R. Liu, F. Wang, and H. Yu. “CTformer: convolution-free Token2Token dilated vision transformer for low-dose CT denoising”. *Physics in Medicine & Biology* 68.6 (2023), p. 065012 (cit. on pp. 114, 133).
- [199] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler. “Image reconstruction is a new frontier of machine learning”. *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1289–1296 (cit. on pp. 102, 103).
- [200] G. Wang, H. Yu, and B. De Man. “An outlook on x-ray CT research and development”. *Medical Physics* 35.3 (2008), pp. 1051–1064 (cit. on p. 103).
- [201] Z. Wang, Z. Deng, F. Liu, Y. Huang, H. Yu, and J. Cui. “OSNet & MNetO: Two Types of General Reconstruction Architectures to Transform DBP Images for Linear Computed Tomography in Multi-Scenarios”. *IEEE Transactions on Instrumentation and Measurement* (2024) (cit. on p. 133).
- [202] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. “Image quality assessment: from error visibility to structural similarity”. *IEEE transactions on image processing* 13.4 (2004), pp. 600–612 (cit. on pp. 90, 111).
- [203] G. Webber and A. J. Reader. “Diffusion models for medical image reconstruction”. *BJR/ Artificial Intelligence* 1.1 (2024), ubae013 (cit. on p. 107).
- [204] L. M. Weber, W. Saelens, R. Cannoodt, C. Sonesson, A. Hapfelmeier, P. P. Gardner, A.-L. Boulesteix, Y. Saeyns, and M. D. Robinson. “Essential guidelines for computational method benchmarking”. *Genome Biology* 20 (2019), pp. 1–12 (cit. on p. 107).

- [205] M. Weiss, N. Brierley, M. von Schmid, and T. Meisen. “Simulation Study: Data-Driven Material Decomposition in Industrial X-ray Computed Tomography”. *NDT 2.1* (2024), pp. 1–15 (cit. on pp. 15, 99, 102).
- [206] B. R. Whiting. “Signal statistics in X-ray computed tomography”. In: *Medical Imaging 2002: Physics of Medical Imaging*. Vol. 4682. SPIE. 2002, pp. 53–60 (cit. on pp. 84, 85).
- [207] B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O’Sullivan, D. L. Snyder, and J. F. Williamson. “Properties of preprocessed sinogram data in X-ray computed tomography”. *Medical physics* 33.9 (2006), pp. 3290–3303 (cit. on pp. 78, 82, 84, 85).
- [208] B. R. Whiting and E. Muka. “Image quantization: statistics and modeling”. In: *Medical Imaging 1998: Physics of Medical Imaging*. Vol. 3336. SPIE. 1998, pp. 260–271 (cit. on pp. 84, 85).
- [209] P. Wilson, M. Smith, J. Hay, J. Warnett, A. Attridge, and M. Williams. “X-ray computed tomography (XCT) and chemical analysis (EDX and XRF) used in conjunction for cultural conservation: the case of the earliest scientifically described dinosaur *Megalosaurus bucklandii*”. *Heritage Science* 6 (Oct. 2018) (cit. on p. 38).
- [210] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum. “Generative adversarial networks for noise reduction in low-dose CT”. *IEEE transactions on medical imaging* 36.12 (2017), pp. 2536–2545 (cit. on p. 133).
- [211] D. Wu, H. Ren, and Q. Li. “Self-supervised dynamic CT perfusion image denoising with deep neural networks”. *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.3 (2020), pp. 350–361 (cit. on p. 133).
- [212] M. Wu, P. FitzGerald, J. Zhang, W. P. Segars, H. Yu, Y. Xu, and B. De Man. “XCIST—an open access X-ray/CT simulation toolkit”. *Physics in Medicine & Biology* 67.19 (2022), p. 194002 (cit. on pp. 85, 89).
- [213] T. Würfl, M. Hoffmann, V. Christlein, K. Breininger, Y. Huang, M. Unberath, and A. K. Maier. “Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems”. *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1454–1463 (cit. on p. 102).
- [214] W. Xia, Z. Yang, Q. Zhou, Z. Lu, Z. Wang, and Y. Zhang. “A transformer-based iterative reconstruction model for sparse-view ct reconstruction”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pp. 790–800 (cit. on p. 133).
- [215] L. Xing, E. A. Krupinski, and J. Cai. “Artificial intelligence will soon change the landscape of medical physics research and practice”. *Medical physics* 45.5 (2018), p. 1791 (cit. on p. 13).
- [216] XRE TESCAN. *TESCAN micro-CT webpage*. 2023 (cit. on pp. 25, 43).

- [217] J. Xu and B. M. Tsui. “Electronic noise modeling in statistical iterative reconstruction”. *IEEE Transactions on Image Processing* 18.6 (2009), pp. 1228–1238 (cit. on p. 85).
- [218] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang. “Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss”. *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1348–1357 (cit. on p. 102).
- [219] J. C. Ye, Y. C. Eldar, and M. A. Unser. *Deep Learning for Biomedical Image Reconstruction*. Cambridge University Press, 2023 (cit. on pp. 16, 106, 108).
- [220] H. Yu, C. Liu, M. Thies, F. Wagner, A. Wang, G. Zhang, Y. Huang, F. Liu, and A. Maier. “Reducing penumbral blur in computed tomography by learning the inverse finite focal spot model”. *Optics Express* 32.13 (2024), pp. 23674–23686 (cit. on p. 133).
- [221] L. Yu, M. Shiung, D. Jondal, and C. H. McCollough. “Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols”. *Journal of computer assisted tomography* 36.4 (2012), pp. 477–487 (cit. on p. 83).
- [222] S. Zabić, Q. Wang, T. Morton, and K. M. Brown. “A low dose simulation tool for CT systems with energy integrating detectors”. *Medical physics* 40.3 (2013), p. 031102 (cit. on p. 85).
- [223] E. Zainulina, A. Chernyavskiy, and D. V. Dylov. “Self-supervised Physics-based Denoising for Computed Tomography”. *arXiv preprint arXiv:2211.00745* (2022) (cit. on pp. 84, 87).
- [224] D. Zeng, J. Huang, Z. Bian, S. Niu, H. Zhang, Q. Feng, Z. Liang, and J. Ma. “A simple low-dose X-ray CT simulation from high-dose scan”. *IEEE transactions on nuclear science* 62.5 (2015), pp. 2226–2233 (cit. on p. 90).
- [225] H.-M. Zhang and B. Dong. “A review on deep learning in medical image reconstruction”. *Journal of the Operations Research Society of China* 8 (2020), pp. 311–340 (cit. on pp. 15, 106, 125, 132).
- [226] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte. “Plug-and-Play Image Restoration With Deep Denoiser Prior”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.10 (2022), pp. 6360–6376 (cit. on pp. 113, 114).
- [227] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising”. *IEEE Transactions on Image Processing* 26.7 (2017), pp. 3142–3155 (cit. on pp. 113, 114).
- [228] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao. “A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution”. *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1407–1417 (cit. on p. 106).

-
- [229] Z. Zhang, L. Yu, X. Liang, W. Zhao, and L. Xing. “TransCT: dual-path transformer for low dose computed tomography”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*. Springer. 2021, pp. 55–64 (cit. on p. 133).
- [230] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen. “Image reconstruction by domain-transform manifold learning”. *Nature* 555.7697 (2018), pp. 487–492 (cit. on pp. 16, 106).
- [231] E. A. Zwanenburg, M. A. Williams, and J. M. Warnett. “Review of high-speed imaging with lab-based x-ray computed tomography”. *Measurement Science and Technology* 33.1 (Nov. 2021), p. 012003 (cit. on p. 39).

List of publications

Publications that are part of this thesis:

1. M. B. Kiss, F. G. Bossema, P. J. van Laar, S. Meijer, F. Lucka, T. van Leeuwen, and K. J. Batenburg. “Beam filtration for object-tailored X-ray CT of multi-material cultural heritage objects”. *Heritage Science* 11.1 (2023), p. 130.
2. M. Kiss, S. Coban, K. Batenburg, T. van Leeuwen, and F. Lucka. “2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning”. *Scientific Data* 10.1 (2023), p. 576.
3. M. B. Kiss, A. Biguri, C.-B. Schönlieb, K. J. Batenburg, and F. Lucka. “Learned denoising with simulated and experimental low-dose CT data”. *arXiv preprint arXiv:2408.08115* (2024).
4. M. B. Kiss, A. Biguri, Z. Shumaylov, F. Sherry, K. J. Batenburg, C.-B. Schönlieb, and F. Lucka. “Benchmarking learned algorithms for computed tomography image reconstruction tasks”. *Applied Mathematics for Modern Challenges* 3.0 (2025), pp. 1–43.

Published datasets and code related to this thesis:

1. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 1-1,000*. June 2023
2. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 1,001-2,000*. June 2023
3. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 2,001-3,000*. June 2023
4. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 3,001-4,000*. June 2023
5. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 4,001-5,000*. June 2023
6. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices OOD*. June 2023

7. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 1-1,000 (reference reconstructions and segmentations)*. June 2023
8. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 1,001-2,000 (reference reconstructions and segmentations)*. June 2023
9. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 2,001-3,000 (reference reconstructions and segmentations)*. June 2023
10. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 3,001-4,000 (reference reconstructions and segmentations)*. June 2023
11. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices 4,001-5,000 (reference reconstructions and segmentations)*. June 2023
12. M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka. *2DeteCT - A large 2D expandable, trainable, experimental Computed Tomography dataset for machine learning: Slices OOD (reference reconstructions and segmentations)*. June 2023
13. M. B. Kiss, H. Der Sarkissian, and F. Lucka. *2DeteCT - Collection of Python scripts for loading, pre-processing, reconstructing and segmenting X-ray CT projection data of the 2DeteCT data collection*. <https://github.com/mbkiss/2DeteCTcodes/>. Accessed: 2025-09-26. June 2023

Samenvatting

This chapter contains a layman's summary of this thesis in Dutch, English, and German. Thanks to Francien Bossema for her help with the summary in Dutch.

Samenvatting in het Nederlands: Ontwikkeling van zelflerende algoritmen voor tweedimensionale computertomografie (CT)

Dit proefschrift behandelt de ontwikkeling van zelflerende algoritmen voor tweedimensionale computertomografie (CT). CT is een beeldvormingsproces dat digitale dwarsdoorsnedes berekent uit een groot aantal röntgenfoto's. Het verschil met eenvoudige röntgenfoto's is dat je niet alleen door objecten of lichamen heen kunt kijken, maar ook hun interne structuur in doorsnedes kunt bekijken zonder dat deze structuren overlappen. De digitale beeldreconstructie is vaak gebaseerd op het gefilterde terugprojectie algoritme, dat de mate van absorptie voor elk volume-element (voxel) van het object berekent. De kwaliteit van deze beeldreconstructies is sterk afhankelijk van de kwaliteit van de individuele röntgenfoto's waaruit de doorsnedes worden berekend.

In hoofdstuk 2 behandelen we het maken van CT-scans van kunstobjecten/culturele artefacten. De beeldvorming van kunstobjecten is meestal bijzonder uitdagend, omdat deze objecten vaak bestaan uit veel verschillende materialen met verschillende dichtheden, diktes en afmetingen. Vooral metaalstructuren in kunstobjecten kunnen leiden tot een slechte visualisatie van de objecten in CT-scans. Het continue spectrum van de stralingsbron/röntgenbuis speelt hier een doorslaggevende rol. De hoogenergetische lichtdeeltjes - de röntgenstraling - die voor de CT-scan door het te meten object worden gestuurd, hebben een statistische energieverdeling. Laag-energetische lichtdeeltjes worden geabsorbeerd door het te meten object (vooral metalen structuren) en kunnen niet aan de andere kant worden gemeten. Dit leidt tot problemen bij de voorbewerking van de röntgenfoto's en de berekende beeldreconstructies vertonen fouten die artefacten worden genoemd. Een oplossing voor dit probleem ligt in het filteren van het laag-energetische deel van het röntgenspectrum - de laag-energetische lichtdeeltjes - zodat vooral hoger-energetische straling het object binnendringt. Deze filtratie wordt uitgevoerd met dunne metaalfolies van aluminium, koper of tin van verschillende diktes, die ook met elkaar gecombineerd kunnen worden. Hoewel deze filtering leidt tot een hoger gemiddeld energiespectrum, vermindert dit proces ook de intensiteit van de straling. Net als bij een foto die bij weinig licht is genomen, kunnen de röntgenbeelden er daardoor slecht belicht en ruisachtig uitzien, wat ook een effect heeft op de beeldreconstructies. Tegelijkertijd dragen lichtdeeltjes met een lagere energie in de beeldvorming bij aan een beter contrast. De uitdaging bij CT-scans van kunstobjecten is daarom om een balans te vinden tussen voldoende sterke filtratie en een voldoende meetsignaal en contrast in de beeldreconstructies.

In hoofdstuk 3 gebruiken we onze ervaring uit hoofdstuk 2 over het uitvoeren van op objecten afgestemde en geoptimaliseerde CT-scans om een grote dataset te verkrijgen. De basis van machinaal leren en de ontwikkeling van zelflerende algoritmen ligt in het verwerven van betekenisvolle en gevarieerde data waarvan de algoritmen kunnen leren. In het geval van tweedimensionale CT-beeldvorming bestaat dit uit de ruwe data (een groot aantal röntgenfoto's) en de bijbehorende beeldreconstructies. Met deze combinatie van gegevens kunnen algoritmen worden getraind om de beelden op een slimme manier te reconstrueren in plaats van klassieke wiskundige iteratieve reconstructiemethoden te gebruiken. Dit is niet noodzakelijk een voordeel voor meetgegevens van hoge kwaliteit, maar wordt relevanter als de verkregen data niet perfect is. Voor het verkrijgen van deze dataset gebruikten we een geavanceerde meetopstelling waarmee we data van zowel hoge als lage kwaliteit van hetzelfde object konden verkrijgen. We besloten om daarvoor data op te nemen met een lage stralingsdosis en dus een verhoogde beeldruis, evenals data waarbij we het röntgenspectrum niet van tevoren hadden gefilterd en waarvan de reconstructies sterke beeldartefacten vertoonden als gevolg van de spectraal verschillende absorptie (de zogeheten beam-hardening). Aangezien we corresponderende hoge kwaliteit en lage kwaliteit opnames en hun respectievelijke beeldreconstructies hebben, kunnen we algoritmes trainen om deze in elkaar om te zetten.

In hoofdstuk 4 onderzoeken we de vraag of het voldoende is om zelflerende algoritmen voor het verwijderen van ruis te trainen op kunstmatig toegevoegde ruis of dat het noodzakelijk is om hiervoor experimentele data die ruis bevat te gebruiken. Voor dit doel koppelen we geoptimaliseerde data uit hoofdstuk 3 aan de corresponderende ruisgevoelige "lage-dosis" data en trainen we twee veel gebruikte neurale netwerken (zelflerende algoritmen voor beeldverwerking) om ze in elkaar om te zetten. Tegelijkertijd nemen we de geoptimaliseerde data als basis om er kunstmatig ruis aan toe te voegen. De gesimuleerde ruis wordt ook gekoppeld aan de geoptimaliseerde data en de neurale netwerken worden getraind om ze in elkaar om te zetten. In het onderzoek laten we zien dat de neurale netwerken die zijn getraind op de gesimuleerde ruis beter zijn in het verwijderen van de ruis op experimentele datasets. Bij het trainen van data direct naar beeldreconstructies is het echter beter om de neurale netwerken te trainen op experimenteel gegenereerde ruis. De simulatie voor kunstmatige ruis lijkt niet complex genoeg om niet-lineaire artefactoorzaken zoals de hierboven genoemde "beam-hardening" vast te leggen. Daarom kunnen de zelflerende algoritmen, die deze artefacten niet zien tijdens hun training op data met kunstmatige ruis, deze achteraf niet verwijderen wanneer ze worden toegepast op experimentele data met ruis.

In hoofdstuk 5 voeren we een groot vergelijkend onderzoek uit naar zelflerende algoritmen voor CT-beeldreconstructie op basis van de dataset uit hoofdstuk 3. Hiervoor gebruiken we de verschillende data om gestandaardiseerde reconstructie-experimenten te definiëren. Als standaard gebruiken we de volledige geoptimaliseerde data van het eerste CT-experiment. Daarnaast gebruiken we deze data voor zes andere CT-experimenten door ten eerste de hoek van de gebruikte röntgenfoto's te beperken tot slechts 120°, 90° en 60° in plaats van 360° en ten tweede het aantal gebruikte foto's te reduceren tot 360, 120 en 60 röntgenfoto's en deze te verdelen over 360°. Voor

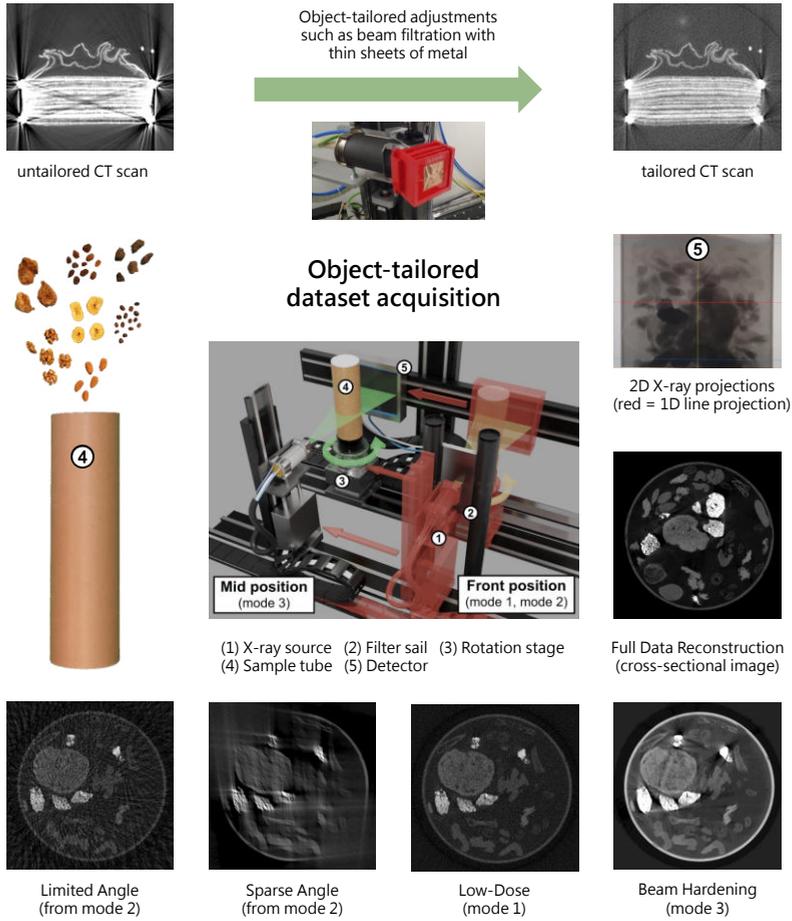
de laatste twee CT-experimenten gebruiken we de “low-dose” en “beam-hardening” data die ruis en beam hardening artefacten bevatten. In totaal worden 3 klassieke wiskundige reconstructiemethoden vergeleken met twaalf zelflerende algoritmen. Deze zijn onderverdeeld in vier categorieën: Post-processing netwerken, geleerde iteratieve methoden, geleerde regularisatiemethoden en plug-and-play methoden. In elke categorie worden drie algoritmen vergeleken voor alle CT-experimenten. De implementatie van de experimenten en algoritmen is uitgevoerd met open source code en kan worden uitgebreid met meer CT-experimenten en algoritmen voor vergelijkingsdoeleinden.

Met deze onderzoeksprojecten hebben we een fundamentele bijdrage kunnen leveren aan op objecten afgestemde CT-scans en de toekomstige gestandaardiseerde ontwikkeling en onderlinge vergelijkbaarheid van zelflerende algoritmen voor CT-beeldreconstructie.

Summary in English: Advancing learned algorithms for 2D X-ray computed tomography

This doctoral thesis deals with the advancement of learned algorithms for two-dimensional computed tomography (CT). This is an imaging process that calculates digital cross-sectional images from a large number of X-ray projections. The difference to simple X-ray images or radiographs, is that it is not only possible to see through objects or bodies, but also to view their internal structure in cross-sections. Computer-aided image reconstruction is often based on the filtered back projection algorithm and calculates the degree of absorption for each volume element (voxel) of the object. The quality of these image reconstructions depends heavily on the quality of the individual X-ray projections from which the cross-sections are calculated.

In Chapter 2, we look at the acquisition of CT scans tailored to cultural heritage objects. The imaging of cultural heritage objects is usually particularly challenging, as these objects often consist of many different materials with different densities, thicknesses and sizes. In particular, metal structures in cultural heritage objects can lead to a poor visualisation of the objects in CT scans. The continuous spectrum of the radiation source/X-ray tube plays an important role in this. The high-energy photons - the X-rays - that are sent through the scanning object for the CT scan have a statistical energy distribution. Low-energy photons are preferentially absorbed by the scanning object (especially by metallic structures within) and cannot be measured on the other side. This leads to problems in the pre-processing of the X-ray projections and the calculated image reconstructions show heavy visual errors known as image artifacts. One solution to this problem lies in filtering the low-energy part of the X-ray spectrum - the low-energy photons - so that mostly higher-energy radiation penetrates the object. This filtration is carried out using thin sheets of metal made of aluminium, copper, or tin of different thicknesses, which can also be combined with each other. Although this beam filtration leads to a higher average energy spectrum, this process also reduces the intensity of the radiation. Similar to a photograph taken in low light



Figuur S1: Visual summary.

conditions, the X-ray images can therefore appear noisy, which also has an effect on the image reconstructions. At the same time, lower-energy photons in the imaging process contribute to a higher contrast ratio. The challenge for object-tailored CT scans therefore is to find a balance between sufficiently strong beam filtration and a sufficiently large measurement signal and contrast ratio in the image reconstructions.

In Chapter 3, we use our experience from Chapter 2 on how to perform object-tailored CT scans to acquire a large dataset. The basis of machine learning and the development of learned algorithms lies in meaningful and diverse data from which the algorithms can learn. In the case of two-dimensional CT imaging, this is given by raw measurement data, a large number of X-ray projections, and corresponding image reconstructions. With this combination of data, algorithms can be trained to reconstruct the images in a learned way instead of using classical mathematical iterative reconstruction methods. This is not necessarily an advantage for high-quality measurement data, but becomes increasingly more relevant when the input data is not ideal. For the acquisition of this dataset, we developed a sophisticated measurement setup that would allow us to acquire both high-quality and inadequate data from the same object. For this, we decided to additionally record measurement data with low radiation dose (low-dose) and thus increased image noise as well as measurement data in which we did not filter the X-ray spectrum in advance and thus the images show strong image artefacts due to the spectrally different absorption (beam-hardening). Since we have corresponding optimised as well as inadequate measurement data and their respective image reconstructions, we can train algorithms to convert them into each other.

In Chapter 4, we investigate the question of whether it is sufficient to train learned algorithms for image denoising on artificially noisy measurement data or whether it is necessary to use experimentally noisy measurement data for this purpose. For this purpose, we pair optimised ‘clean’ measurement data from the dataset from chapter 3 with the corresponding noisy ‘low-dose’ measurement data and train two common neural networks (learned algorithms for image processing) to convert them into each other. At the same time, we take the clean measurement data as a basis to simulate and apply artificial noise to it. The simulated noisy data is also paired with the clean measurement data and the neural networks are trained to convert them into each other. We show in our studies that the neural networks trained on the simulated noisy data are better at denoising experimentally noisy measurement data. However, when training them to convert measurement data into image reconstructions, it is better to train the neural networks on experimentally noisy measurement data. The simulation approach for artificial noise does not seem to be complex enough to capture non-linear artifact causes such as beam-hardening mentioned above. Accordingly, the learned algorithms that do not see these artifacts during their training on simulated noisy data cannot remove them afterwards when applied to experimental noisy data.

In Chapter 5, we perform a large comparative study of learned algorithms for CT image reconstruction based on the dataset from Chapter 3. For this purpose, we use the different measurement data of the dataset to define standardized reconstruction ex-

periments. We use the complete ‘clean’ measurement data for the first CT experiment as a general comparison value. In addition, we use the ‘clean’ measurement data for six further CT experiments by firstly limiting the used X-ray projections to only 120°, 90° and 60° instead of 360° and secondly reducing the number of X-ray projections to 360, 120 and 60 distributed sparsely over 360°. For the last two CT experiments, we use the noisy ‘low-dose’ and artifact-inflicted ‘beam-hardening’ measurement data. A total of three classical mathematical reconstruction methods are compared with twelve learned algorithms. These are divided into four categories: Post-processing networks, unrolled/learned iterative methods, learned regulariser methods and plug-and-play methods. In each category, three algorithms are compared for all CT experiments. The implementation of the experiments and learned algorithms is done in an open-source code collection and can be extended by further CT experiments and algorithms for further comparisons.

With these research studies, we were able to make fundamental contributions to object-tailored CT scans and the future, standardized development and comparability of learned algorithms for CT image reconstruction.

Zusammenfassung auf Deutsch: Weiterentwicklung von Lernalgorithmen für zweidimensionale Computertomografie (CT)

Diese Doktorarbeit befasst sich mit der Weiterentwicklung von Lernalgorithmen für zweidimensionale Computertomografie (CT). Dabei handelt es sich um ein bildgebendes Verfahren, das digitale Schnittbilder aus einer Vielzahl an Röntgenprojektionen errechnet. Man könnte es alternativ auch Schichtröntgen nennen. Der Unterschied zu einfachen Röntgenaufnahmen besteht darin, dass man nicht nur durch Objekte oder Körper hindurchsehen kann, sondern auch deren innere Struktur in überlagerungsfreien Querschnitten betrachten kann. Die computergestützte Bildrekonstruktion basiert häufig auf dem Algorithmus der gefilterten Rückprojektion und errechnet den Absorptionsgrad der Röntgenstrahlen für jedes Volumenelement (Voxel) des Objektes. Die Qualität dieser Bildrekonstruktionen hängt dabei stark von der Qualität der einzelnen Röntgenprojektionen ab, aus denen die Querschnitte errechnet werden.

In Kapitel 2 beschäftigen wir uns mit der Aufnahme von CT-Scans zugeschnitten auf Kunstobjekte/Kulturgüter. Die Bildgebung von Kunstobjekten ist zumeist besonders herausfordernd, da diese Objekte häufig aus vielen verschiedenen Materialien mit unterschiedlichen Dichten, Dicken und Größen bestehen. Insbesondere Metallstrukturen in Kunstobjekten können zur mangelhaften Darstellung der Objekte in CT-Scans führen. Dabei spielt insbesondere das kontinuierliche Spektrum der Strahlungsquelle/Röntgenröhre eine entscheidende Rolle. Die hochenergetischen Lichtteilchen – die Röntgenstrahlung – die für den CT-Scan durch das Messobjekt geschickt werden, haben eine statistische energetische Verteilung. Niederenergetische Lichtteilchen werden dabei besonders häufig vom Messobjekt (vor allem von metallischen Strukturen)

absorbiert und können auf der anderen Seite nicht gemessen werden. In der Vorverarbeitung der Röntgenprojektionen führt dies zu Problemen und die errechneten Bildrekonstruktionen zeigen Bildfehler sogenannte Artefakte. Eine Lösung für dieses Problem liegt in der Filtrierung des niederenergetischen Teils des Röntgenspektrums – der niederenergetischen Lichtteilchen – sodass größtenteils höherenergetische Strahlung das Objekt durchdringt. Diese Filtrierung erfolgt durch dünne Metallfolien aus Aluminium, Kupfer oder Zinn von unterschiedlicher Dicke, die auch miteinander kombiniert werden können. Wenngleich diese Filtrierung zu einem durchschnittlich höherenergetischen Spektrum führt, verringert dieser Vorgang auch die Intensität der Strahlung. Ähnlich zu einer Fotografie bei schwachem Licht können die Röntgenaufnahmen demnach schlecht belichtet und verrauscht erscheinen, was sich auch auf die Bildrekonstruktionen auswirkt. Gleichzeitig tragen niederenergetischere Lichtteilchen in der Bildgebung zu einem höheren Kontrastverhältnis bei. Die Herausforderung bei auf Objekte zugeschnittenen CT-Scans ist entsprechend, eine Balance zu finden zwischen genügend starker Filtrierung und ausreichendem Messsignal sowie Kontrastverhältnis in den Bildrekonstruktionen.

In Kapitel 3 nutzen wir unsere Erfahrungen aus Kapitel 2, wie man auf Objekte zugeschnittene und optimierte CT-Scans durchführt, um einen großen Datensatz aufzunehmen. Die Basis von maschinellem Lernen und der Entwicklung von Lernalgorithmen liegt in aussagekräftigen und vielfältigen Daten, von denen die Algorithmen lernen können. Für den Fall von zweidimensionaler CT-Bildgebung sind dies zum einen die rohen Messdaten, eine Vielzahl an Röntgenprojektionen, und zum anderen deren entsprechende Bildrekonstruktionen. Mit dieser Kombination an Daten kann man entsprechend Algorithmen darauf trainieren, Rekonstruktionen der Bilder auf eine gelernte Art und Weise zu errechnen anstatt klassische, mathematische iterative Rekonstruktionsmethoden zu verwenden. Für hochqualitativ erfasste Messdaten ist dies nicht unbedingt von Vorteil, aber wird umso relevanter, wenn die eingehenden Daten nicht ideal sind. Für die Aufnahme dieses Datensatzes überlegten wir uns eine ausgeklügelte Messanordnung mit der wir sowohl hochqualitative als auch eingeschränkte Daten von dem gleichen Objekt erfassen können. Wir entschieden uns dabei dafür zusätzlich Messdaten mit geringer Strahlendosis (sog. low-dose) und damit erhöhtem Bildrauschen aufzunehmen als auch Messdaten, bei denen wir das Röntgenspektrum nicht vorab filterten und die Bildrekonstruktionen demnach starke Bildartefakte aufgrund der spektral unterschiedlichen Absorption (sog. beam-hardening) aufweisen. Dadurch dass wir korrespondierende optimierte als auch unzureichende Messdaten und deren jeweilige Bildrekonstruktionen besitzen, können wir Algorithmen darauf trainieren, diese ineinander umzuwandeln.

In Kapitel 4 untersuchen wir die Fragestellung, ob es ausreicht Lernalgorithmen für Bild-Entrauschen auf künstlich verrauschten Messdaten zu trainieren oder ob es notwendig ist, experimentell verrauschte Messdaten dafür zu verwenden. Dafür paaren wir optimierte „saubere“ Messdaten aus dem Datensatz aus Kapitel 3 mit den korrespondierenden verrauschten „low-dose“ Messdaten und trainieren zwei übliche neuronale Netzwerke (Lernalgorithmen für Bildverarbeitung) um diese ineinander umzuwandeln. Gleichzeitig nehmen wir die sauberen Messdaten als Grundlage um

künstliches Rauschen zu simulieren und auf diese anzuwenden. Auch die simuliert verrauschten Daten werden mit den sauberen Messdaten gepaart und die neuronalen Netzwerke darauf trainiert, diese ineinander umzuwandeln. Wir zeigen in unseren Untersuchungen, dass die neuronalen Netzwerke, die auf den simuliert verrauschten Daten trainiert wurden, besser darin sind, experimentell verrauschte Messdaten zu entrauschen. Wenn das Training aber von Messdaten zu Bildrekonstruktionen erfolgt, ist es besser die neuronalen Netzwerke auf experimentell verrauschten Messdaten zu trainieren. Der Simulationsansatz für das künstliche Rauschen scheint nicht komplex genug zu sein, um nicht-lineare Artefakt-Ursachen wie das oben erwähnte „beam-hardening“ zu erfassen. Entsprechend können die Lernalgorithmen, die diese Artefakte nicht im Verlauf ihres Trainings auf künstlich verrauschten Daten gesehen haben, diese auch nicht im Nachgang entfernen, wenn sie auf experimentell verrauschte Daten angewendet werden.

In Kapitel 5 führen wir eine große Vergleichsstudie von Lernalgorithmen für CT-Bildrekonstruktion auf Basis des Datensatzes aus Kapitel 3 durch. Hierfür nutzen wir die verschiedenen Messdaten des Datensatzes, um standardisierte Rekonstruktionsexperimente zu definieren. Als allgemeiner Vergleichswert nutzen wir für das erste CT-Experiment die vollständigen „sauberen“ Messdaten. Zudem nutzen wir die „sauberen“ Messdaten für sechs weitere CT-Experimente, indem wir zum einen die genutzten Röntgenprojektionen auf nur 120°, 90° bzw. 60° statt 360° limitieren und zum anderen die Anzahl auf 360, 120 und 60 Röntgenprojektionen reduzieren und spärlich auf 360° verteilen. Für die letzten beiden CT-Experimente nutzen wir die verrauschten und Artefakt-behafteten „low-dose“ und „beam-hardening“ Messdaten. Insgesamt werden drei klassische, mathematische Rekonstruktionsmethoden mit zwölf Lernalgorithmen verglichen. Diese sind in vier Kategorien aufgeteilt: Nachbearbeitungs-Netzwerke, ausgerollte/erlernte iterative Verfahren, erlernte Regularisierungsmethoden und Plug-and-Play Methoden. In jeder Kategorie werden drei Algorithmen für alle CT-Experimente verglichen. Die Implementierung der Experimente und Lernalgorithmen erfolgt in einer quelloffenen Code-Sammlung und kann um weitere CT-Experimente und Algorithmen für Vergleichszwecke erweitert werden.

Mit diesen Forschungsstudien konnten wir einen grundlegenden Beitrag für auf Objekte zugeschnittene CT-Scans und die zukünftige, standardisierte Entwicklung und Vergleichbarkeit von Lernalgorithmen für CT-Bildrekonstruktionen leisten.

Curriculum Vitae



Maximilian B. Kiss was born in 1995 in Regensburg, Germany. He obtained his bachelor's and master's degree in Physics from the Eidgenössische Technische Hochschule (ETH) Zurich in 2018 and 2020, respectively. From 2020 - 2024 he undertook a PhD degree under supervision of prof. dr. K. J. Batenburg and co-supervision of dr. F. Lucka and prof. dr. T. van Leeuwen at the Centrum Wiskunde & Informatica (CWI), the national research institute for mathematics and computer science in Amsterdam, The Netherlands. During his PhD studies, he took courses on inverse problems in imaging, on optimisation techniques in inverse problems, and in scientific conduct, among others. During this time, he was also active in voluntary work, co-founded the social startup "youmocracy - democracy needs you" (Berlin, Germany) and served as its Head of Nationwide Tasks & Acquisitions. Furthermore, he

was an active member (2021) and co-chair (2022) of CWI's Activity Committee. In 2023 and 2024, he was a Visiting Student at Clare Hall (Cambridge, United Kingdom), and performed part of his research at the Cambridge Image Analysis (CIA) group of prof. dr. C.-B. Schönlieb at the Department of Applied Mathematics and Theoretical Physics (DAMTP) of the University of Cambridge (Cambridge, United Kingdom) during a four-month and three-month research placement. More information about his career can be found on LinkedIn: <https://www.linkedin.com/in/maximilian-kiss-986a881a/>.

Photograph by Lukas Linder.

Acknowledgements

First and foremost, I want to thank my PhD supervisor, Joost Batenburg. Your provoking questions and our strong back and forth discussions enabled me to distill my research down to publishable results. I am glad you entertained my non-standard way of approaching projects and strong opinions on changing academic practice, giving way where it was possible and insisting on your approach where it was necessary. Since our first meeting you have transformed the trajectory of my PhD and I deem our meetings the most efficient part of my PhD studies.

At the same time, I want to thank my co-promoter and daily supervisor, Felix Lucka, who has secured funding for my position/project. I am grateful for your kind, caring supervision smoothing the start of my PhD studies during Corona. I appreciate your social nature and the opportunities you gave me, such as presenting on your behalf at international conferences, participating in selected workshops, and supporting two transformative research stays at the University of Cambridge.

Furthermore, I would like to thank all my co-authors for their contributions to the publications within this thesis and especially acknowledge Johannes Krauß for creating many of its illustrations.

Thank you, Tristan van Leeuwen, for being my second co-promoter and a supportive Computational Imaging (CI) group leader. Thanks also to my CI colleagues for the useful lunches, cake, tea and coffee discussions: Sophia, Allard, Poulami, Dzemila, Mathé, Roozbeh, Willem Jan, Francien, Richard, Vlad, Dirk, Alex, Ezgi, Hamid, Tianyuan, Marcos, Wallace, and Xianyuan.

I would like to thank the CWI staff, whose impact on a pleasant, supportive, and smooth work environment shall never be overlooked: Remco, Rob, Bikkie, Vera, Karima, Alma, Nada, Duda, Erik, Dick, and most importantly, the legendary Minnie.

Thanks also, to all foosball players at the institute, esp. Dyon, Léo, Fran, Llorenc and to all Activity Committee members who organized many fun events before, with, and after me being a member and its co-chair.

A special thank you to my closest people in Amsterdam: Francien and Paul, for your collaborative attitude, your social nature, and your emotional support. Vlad, for being a wonderful colleague and friend. Lisa and Daniel, for most entertaining after-work drinks; Hema and Nikolaj, for all the foodie experiences and being my paranymphs; Isa and Chris, for always having an open ear and heart.

In Cambridge I had the privilege to call Clare Hall my academic home. Thanks, Holly Hedgeland, for accepting my application as a Visiting Student and to all my rowing teammates. Thanks also to all my colleagues at the Cambridge Image Analysis group and my host Carola Schönlieb for providing such a social and collaborative environment.

I would like to express my gratitude to my informal and formal mentors, that have inspired and guided me in my academic life: Markus Anthofer, Henrik Becker, Alexandre Réfrégier, Peter Manser, Michael Fix, Lei Xing, and Karl Bush.

Thank you to all my friends afar that in one way or another supported me in my PhD journey.

My utmost gratitude to my dear friend and co-founder, Florentin Siegert. Working on youmocracy was always a welcome diversion from my academic work, and I am proud to have built a six-figure social start-up over three years with you.

To my best friend, Vincent Casser, thank you for always being there through countless hours of video and audio calls and for the wonderful weekends and vacations spent together during my PhD studies.

To my parents, Dr. Katharina and Peter Kiss, thank you for the unwavering support all these years. You have raised me to always give my best and aim the highest, although that meant for most of my adult life to live far away from you. Big thanks also to my brother, Florian Kiss, for being my childhood idol and for your understanding and balancing outside perspectives.

Sarah, my love, it was a great fortune to have met you in Cambridge. You were my safe haven in the stressful last months of my PhD, and I am looking forward to spending our lives together.

