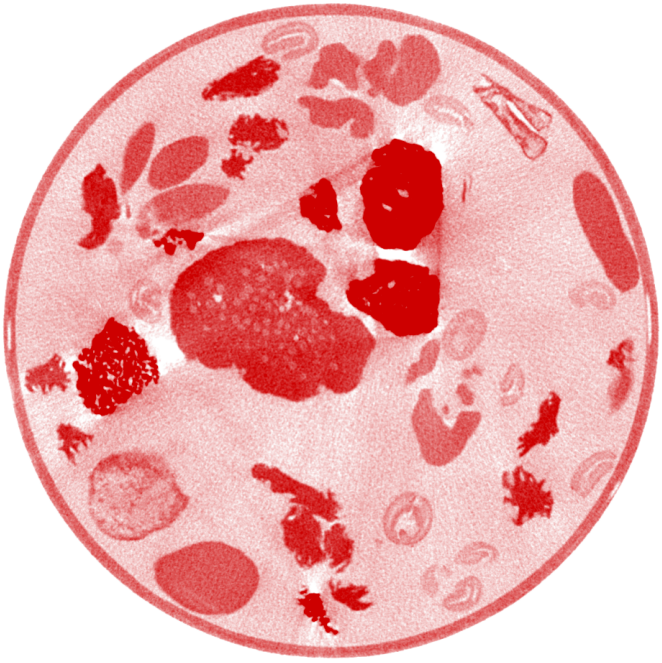
Advancing learned algorithms for 2D X-ray computed tomography

Kiss, M.B.



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

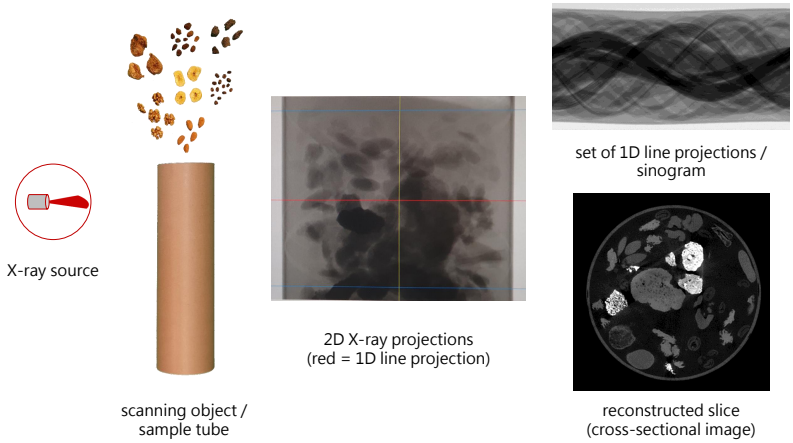


Figure 1.2: Basic terminology of (2D) Computed Tomography.

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 $y(\ell) = \int_\ell x(z)\,\mathrm{d}z, \ell \in \mathcal{L}$, 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:

$$Ax = y + \tilde{e} \tag{1.1}$$

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:

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

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

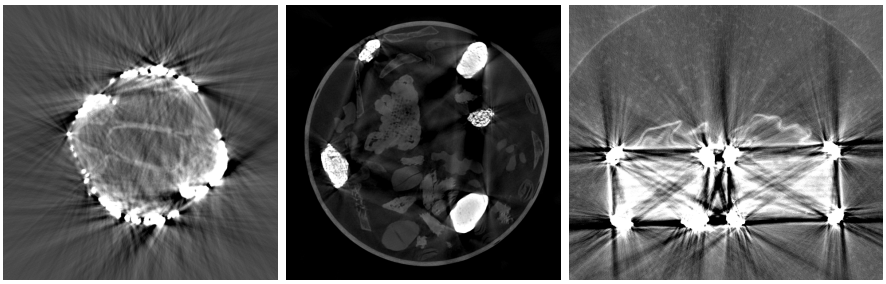


Figure 1.3: Image artifacts in CT imaging of various multi-material and multi-scale objects. From left to right: a velvet knife holder with gold and silver embroidery causing metal artifacts; a cardboard tube filled with dried fruits and nuts, coffee powder, and lava stones causing beam hardening artifacts, 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 µm resolution, features of the object that have a size of one pixel are 60 µm 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).

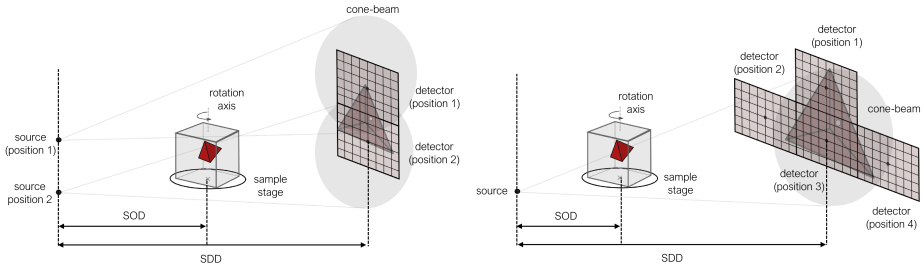


Figure 1.4: Schematic representation of a CT setup illustrating two options for tiling with different magnification factors $\frac{SDD}{SOD}$; left: vertical tiling with multiple source and detector positions; right: tiling with fixed source position and multiple detector positions to virtually create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

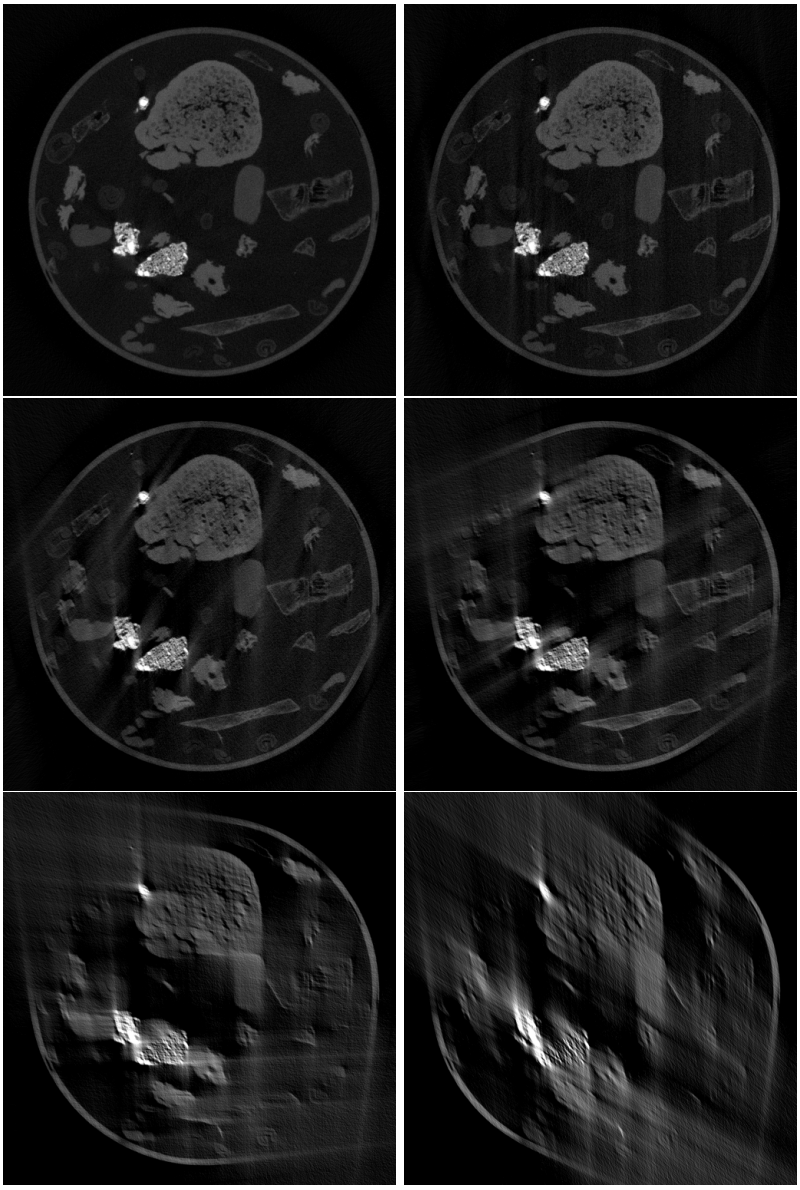


Figure 1.5: From top to bottom, left to right: FBP reconstructions of a limited-angle CT acquisition based on projection data with a limited angle of 180°, 150°, 120°, 90°, or 60° and a fully sampled CT acquisition.

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 an 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

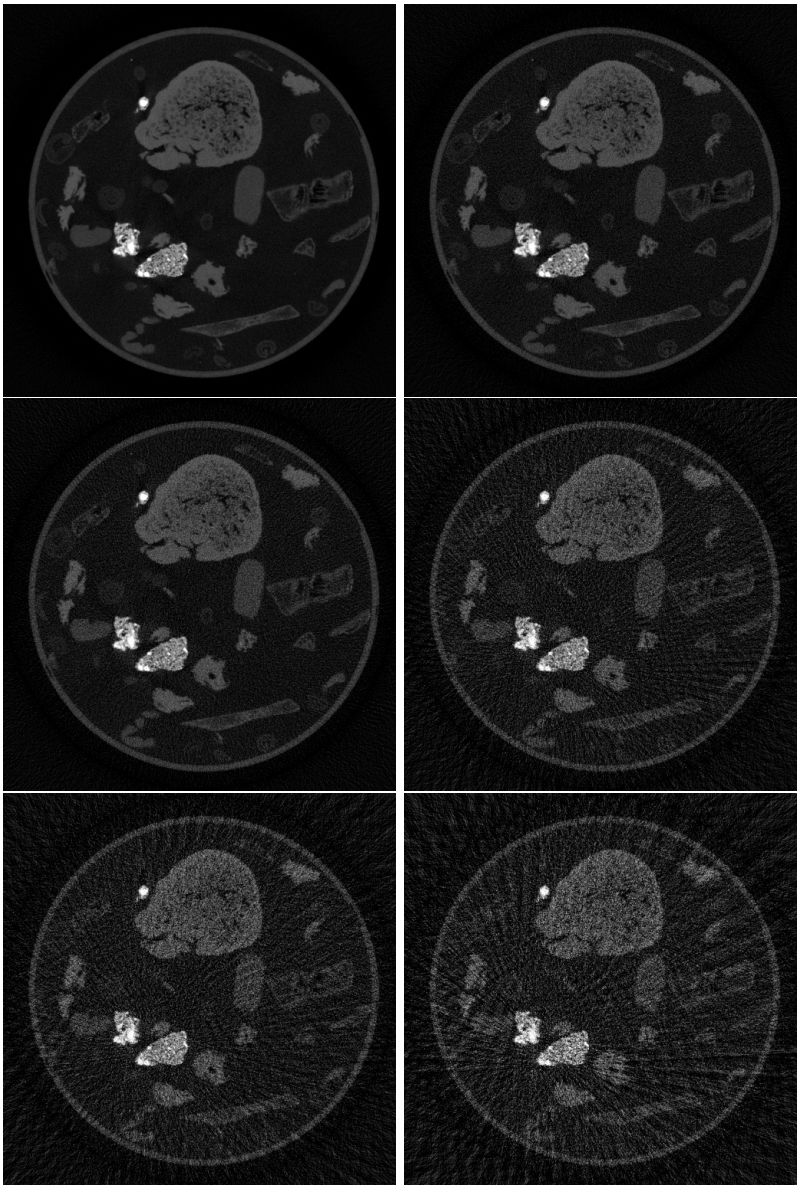


Figure 1.6: From top to bottom, left to right: FBP reconstructions of a sparse-angle CT acquisition based on 720, 360, 180, 120, 90, and 60 projections and a fully sampled CT acquisition.

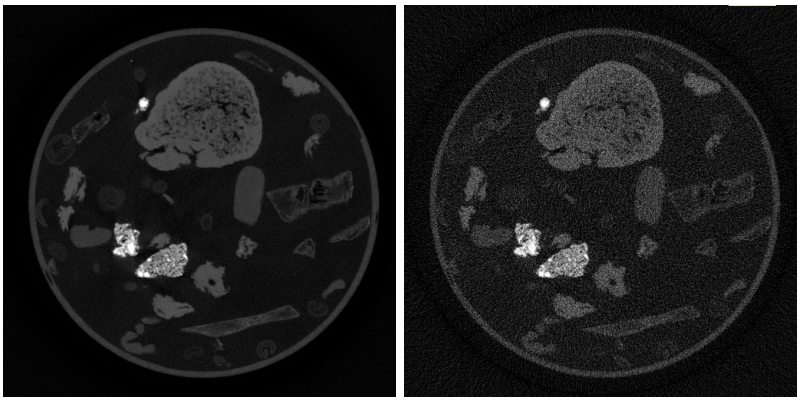


Figure 1.7: From left to right: FBP reconstruction of a high-dose and low-dose CT acquisition.

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.

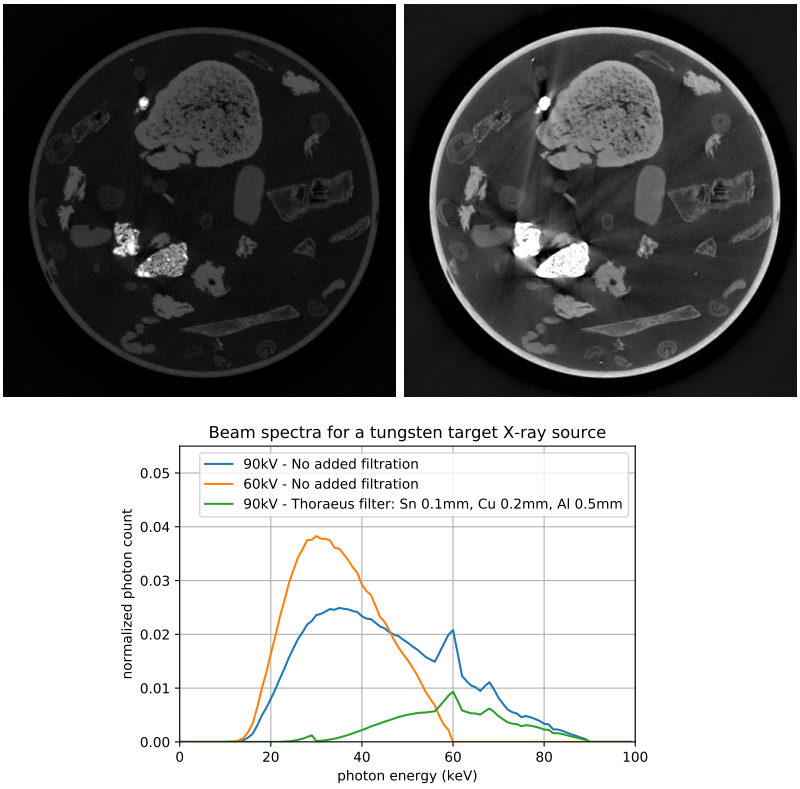


Figure 1.8: From left to right: FBP reconstruction of a filtered and an unfiltered CT acquisition exhibiting beam-hardening artifacts and the corresponding beam spectra (green and orange) simulated with the TASMIP software [22].

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 $i = 1, \ldots, m$ (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:

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

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.

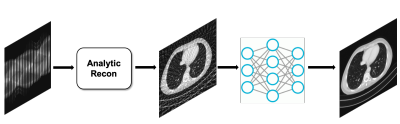
1.4 Machine learning for computed tomography

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].

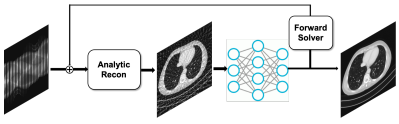
These models, however, require large-scale, versatile, experimental datasets to be trained on, and the field of X-ray CT lacked such datasets for developing machine learning methods. Especially two-dimensional, reconstructed CT slices would play an important role in advancing method development since 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.

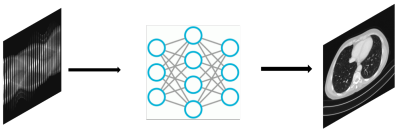
To further clarify the landscape of deep learning approaches for image reconstruction, a domain-based categorization was proposed by Ravishankar et al. [166], which classify methods into image-domain learning, hybrid-domain learning, AUTOMAP [230] (which learns a direct mapping between measurement and 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.



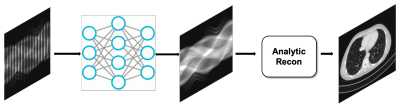
(**a**) Image-domain Learning.



(**b**) Hybrid-domain Learning



(**c**) AUTOMAP



(**d**) Sensor-domain Learning

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 untailored 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.

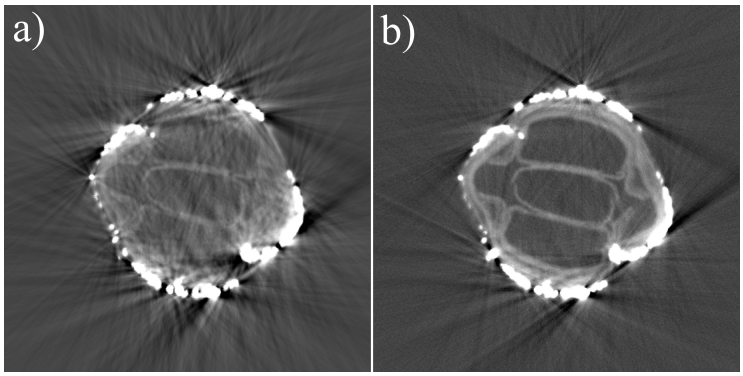


Figure 1.10: Reconstruction slice of a standard acquisition (50 kV, 700 µA, no filter) a) and an object-tailored acquisition (70 kV, 1000 µA, Thoraeus filter (Sn 0.25mm, Cu 0.5mm, Al 0.5mm) b) for the case study 2 "Purple velvet knife holder" from chapter 3

1

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).

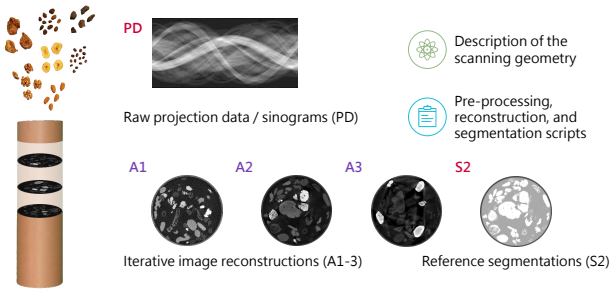


Figure 1.11: Overview of the scope of the 2DeteCT dataset.

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.

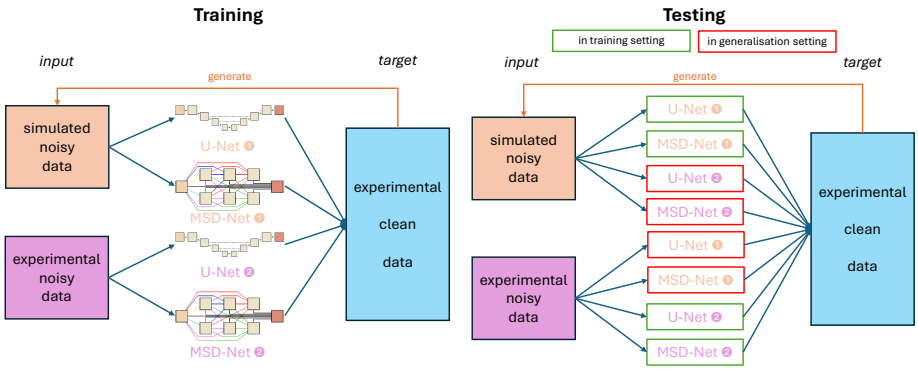


Figure 1.12: Training and testing scenarios for learned denoising networks (U-/MSD-Net illustrations adopted from [160]).

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).

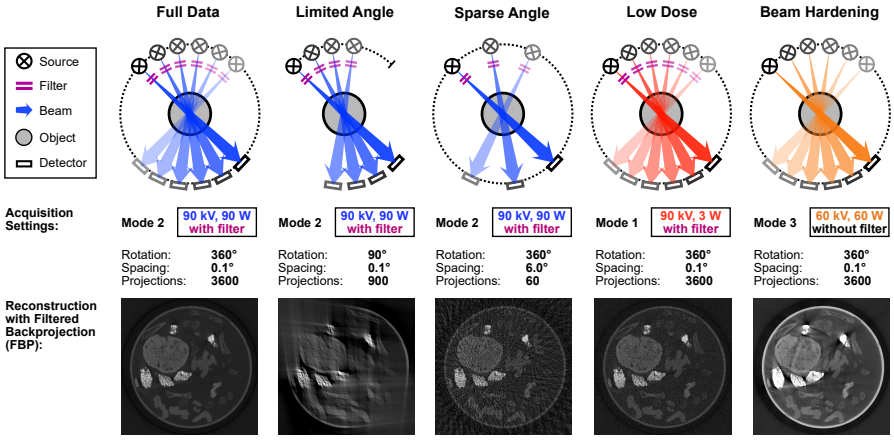


Figure 1.13: CT Image Reconstruction Tasks.