Universiteit Leiden
The Netherlands

Deep learning for tomographic reconstruction with limited data

Hendriksen, A.A.

1 Introduction

Tomography is a powerful technique to non-destructively determine the interior structure of an object. Usually, a series of projection images (e.g. X-ray images) is acquired from a range of different positions. from these projection images, a reconstruction of the object's interior is computed. Many advanced applications require fast acquisition, effectively limiting the number of projection images and imposing a level of noise on these images. These limitations result in artifacts (deficiencies) in the reconstructed images. Recently, deep neural networks have emerged as a powerful technique to remove these limited-data artifacts from reconstructed images, often outperforming conventional state-of-the-art techniques. To perform this task, the networks are typically trained on a dataset of paired low-quality and high-quality images of similar objects. This is a major obstacle to their use in many practical applications. In this thesis, we explore techniques to employ deep learning in advanced experiments where measuring additional objects is not possible.

In this chapter, we first describe several applications of tomography and the importance of accurate reconstructed images. Next, we describe the factors that determine the quality of the reconstructed images and how they are related. In Section 1.1.3, the image acquisition process is introduced with a focus on noise statistics that are relevant to proposed techniques in later chapters. In Sections 1.1.4 and 1.1.5, the tomographic inverse problem is formulated and relevant reconstruction algorithms are reviewed. Next, deep learning techniques for improving the quality of reconstructed tomographic images are introduced. In Section 1.1.7, these deep learning techniques are considered from a Bayesian viewpoint that allows predicting their properties. Next, practical obstacles to the collection of a training dataset are discussed. Finally, an overview is given of the main research questions and the chapters that aim to answer them.

1.1 Background

1.1.1 Applications of tomography

Tomography is a powerful technique to non-destructively determine the interior structure of an object. Since the 1960s, tomography has steadily gained popularity in a variety of applications, ranging from protein structure determination at the nanometer scale [17] to inspection of airplane assemblies of up to several meters [52]. In this section, we discuss tomography as it is used in three places: the hospital, the laboratory, and at a synchrotron facility (particle accelerator). For each, we discuss their characteristics and example use cases, as well as the importance of obtaining high-quality reconstructed images.

Hospital. The use of X-rays in medical applications is perhaps the most familiar to the public and can serve as an introduction to the topic. Investigation using only a single 2D X-ray image is known as radiography and is widely used to diagnose easily visible injuries such as a broken bone. Here, X-rays pass through the body and are collected on a detector. As different parts of the body absorb X-rays to a varying degree, a gray-scale image can be formed that corresponds to the absorption of the X-rays by the body.

Multiple X-ray images are required to investigate an object using tomography, which is known as CT (computed tomography) in medical applications. Here, multiple 2D X-ray images are acquired from varying positions, usually in a continuous fashion. From these images, a 3D reconstruction of the interior of the patient's body can be computed, enabling the diagnosis of diseases that require precise localization of low-contrast features in the patient's body, such as a tumor [85]. Commonly, the X-ray source and detector rotate in a helical pattern around a patient that lies flat, as displayed in Figure 1.1. On the right, a reconstructed image is shown, displaying a cross-section of the hip area. The achievable resolution is often quite coarse, making medical CT unsuitable for many non-medical applications.

Laboratory. Laboratory-based micro-CT is used in a variety of industrial and scientific applications. It is used to scan smaller objects where it provides substantially finer resolution than medical X-ray scanners. Typically, the sample under investigation is mounted on a rotation stage between an X-ray point source and a detector. This type of acquisition is known as circular cone beam geometry, and is illustrated in Figure 1.1. More flexible acquisition is possible, for instance, using the micro-CT scanner at the FleX-ray Laboratory at the CWI [42]. This scanner has ten degrees of freedom, enabling the user to zoom in (by moving the object closer to the source), or to obtain a larger virtual detector by repeating an acquisition with the detector at multiple positions and stitching the projection images together afterwards.

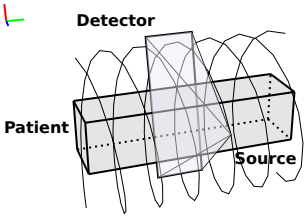
Micro-CT is used to investigate a wide variety of objects, of which we give a few examples here. One application of micro-CT is in art history, where it can be used to assign a date to wooden objects using the tree ring pattern hidden in their interior structure of the wood [24]. Another application is geology, where micro-CT is used to examine one-of-a-kind fossils and to quantify geological properties such

Medical CT

Photon flux: low

Resolution: 500 - 1000 µm

Acquisition time: seconds to minutes



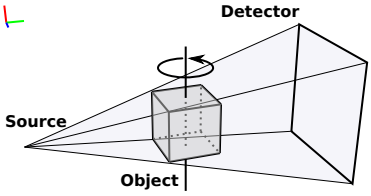
Helical cone beam

Laboratory micro-CT

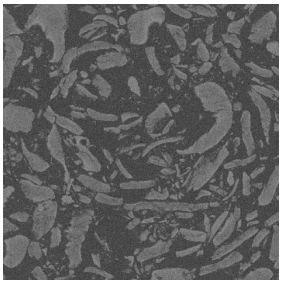
Photon flux: High

Resolution: 10 - 100 µm

Acquisition time: minutes to hours



Circular cone beam

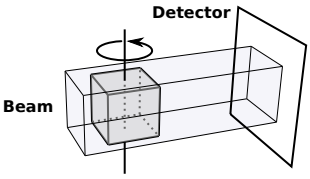


Synchrotron X-ray tomography

Photon flux: Extreme

Resolution: 0.1 - 1 µm

Acquisition time: milliseconds to hours



Single-axis parallel beam

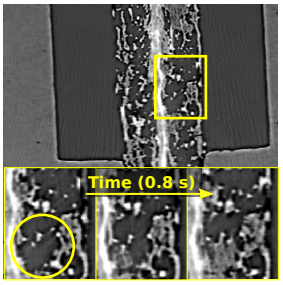


Figure 1.1: Three application areas of tomography. Medical CT is used to diagnose patients using little radiation and at a relatively coarse resolution. Often, the acquisition proceeds using a helical cone beam geometries, where the source-detector pair rotates around the patient. A reconstructed image is show on the right, containing a cross-section of the hip area. Laboratory micro-CT provides higher rates of radiation to investigate smaller objects a higher resolution than medical CT. Typically, the circular cone beam geometry is used, where an objects rotates between a point source and a flat panel detector. On the right, a reconstructed image of oatmeal grains is shown with a voxel size of 17 µm. Synchrotron-based X-ray tomography provides extremely high photon flux and even smaller resolution. The commonly used single axis parallel beam acquisition geometry is shown in the middle. On the right, a 4D (space + time) reconstruction of a hydrogen fuel cell is shown. The region in the yellow square is shown at three time steps, demonstrating that a water bubble is forming inside the fuel cell in the area indicated by the yellow circle.

as porosity and permeability of rock samples [92]. Properties that are important for the manufacturing and development of new batteries can also be quantified using micro-CT [164]. Because X-rays do not easily penetrate the heavy materials present in batteries, however, scanning can take multiple hours [48]. Faster scanning is possible at more advanced facilities.

Synchrotron. Many important scientific and industrial advances are enabled by tomography at one of 50 X-ray synchrotron facilities worldwide [190]. These facilities generate extremely bright X-ray beams using a particle accelerator, enabling acquisition at substantially higher speeds and smaller scales than laboratory-based micro-CT setups. In a typical setup, the sample is mounted on a rotation stage and the photons in the X-ray beam move along parallel lines, resulting in a single-axis parallel beam geometry, which is illustrated in Figure 1.1. In this way, scanning a battery sample takes 5 minutes instead of 4 hours [48].

Because it is a non-destructive technique, synchrotron-based tomography is ideally suited to track internal structural dynamics over time [115]. Here, a full 3D reconstruction of a dynamically evolving process inside an object is made at several time steps, resulting in a 3D movie. This enables examining batteries while they are discharging, for example, and has important implications for understanding and optimizing battery performance [12, 111]. Another application is the the investigation of water dynamics inside a hydrogen fuel cell [195]. This is illustrated in Figure 1.1, which shows the formation of a water bubble inside a fuel cell. The frame rate of these experiments (where each frame is one 3D volume) is typically around 10 Hz, but some experiments have demonstrated frame rates of up to 208 Hz — three times the frame rate of a regular computer monitor [56].

In all of these applications, the quality of the reconstructed images is important. In medical settings for instance, visually identifying a lesion (tumor) is complicated, as the contrast of the lesion with respect to the surrounding tissue is small [123]. Similar considerations arise in art and geology applications where fine-scale features, such as tree rings, can be extremely important. In addition to the visual inspection, quantitative analysis is used to automatically extract parameters of interest out of the reconstructed images, such as the porosity of a rock sample [115]. Here, the images are usually segmented to identify separate features or parts. Preferably, coarse-grained techniques, such as thresholding, flood-filling and clustering are used that can easily be manually tuned, inspected, and applied to large volumes at once. These techniques can be sensitive to imperfections in the reconstructed image, and can therefore benefit from high-quality reconstructions.

1.1.2 Factors that influence image quality

The quality of the reconstructed image depends on several factors in the acquisition process and they are all to some degree interconnected. The relationship between these factors is displayed schematically in Figure 1.2. We first discuss how these factors contribute to the quality of the reconstructed image and then discuss limits that occur in practice.

First of all, measurement noise on the projection images is an important factor

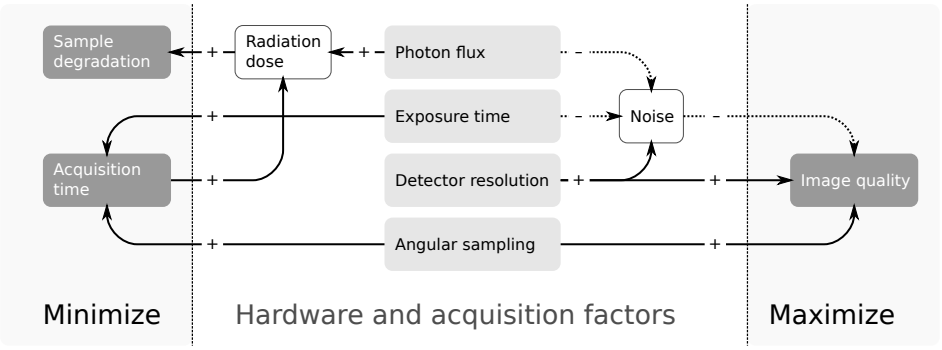


Figure 1.2: The relation between the aims of tomographic imaging and factors that can be influenced during acquisition. In the dark gray boxes, we identify three aims: preventing radiation damage, minimizing acquisition time, and maximizing image quality. The acquisition parameters are shown in the middle column. Solid arrows with a "+" indicate a positive correlation between factors and dashed arrows with a "−" indicate a negative correlation.

as noise carries over into the reconstructed images. The noise level of a projection image is influenced by the exposure time, detector resolution, and photon flux, the rate of photon emission from the source. As more photons are measured, the projection image becomes less noisy. Therefore, the noise level can be reduced by increasing the photon flux or the exposure time of each projection image. Detector resolution plays an important role as well, as doubling the pixel size in each dimension quadruples the number of photons measured per pixel, reducing the level of noise.

Second, the pixel resolution of the detector determines the maximally attainable voxel resolution of the reconstructed volume. In parallel beam setups, the detector pixel width and height determine the minimal resolvable width and height of the voxel. In cone beam setups, the detector and reconstruction resolution are additionally related by the magnification factor that results from moving the object closer to the source. In addition, blurring can be induced on the detector by cross-talk between nearby detector pixels and a large focal spot size, decreasing the effective resolution [31, 161].

Finally, the resolution of the reconstructed images is influenced by the angular sampling rate. As projection images are acquired from more angles, the resolution of the reconstructed images improves. The angular sampling rate is related to the size of the reconstructed volume by the Crowther criterion [95]. In parallel beam setups, for instance, achieving full resolution on an N^3 voxel grid requires $\pi N/2$ projection images to be acquired over an angular range of 180°. If the angular sampling rate is too low compared to the rotation speed, the projection images may suffer from motion blur. This occurs when the sample rotates too much in the course of one exposure [41]. Often, this effect is negligible in practice.

Other factors may also cause a deterioration of image quality. These typically involve a mismatch between physical reality the mathematical model used for

reconstruction. For instance, beam hardening artifacts may be introduced into the reconstructed images when the X-ray beam is not monochromatic [31]. We do not consider these problems in this thesis.

Practical challenges. In practice, the ability to obtain high-quality measurements can be restricted by the wish to prevent radiation damage and minimize acquisition time. In addition, other factors come in play that vary from application to application. In medical tomography, the photon flux and acquisition time are limited to prevent radiation damage. Because the diagnostic use of CT continues to increase, radiation damage is a growing concern: it has been estimated that the percentage of cancer cases in the United States that may be attributable to the radiation from CT studies has increased from 0.4% in 1996 to $1.5 - 2.0\%$ in 2007 [25]. Radiation damage is not limited to human studies and is a frequent concern in many applications.

The photon flux may also be limited by the available hardware. As previously mentioned, the photon flux in laboratory micro-CT is limited, which causes the scan of a battery to take hours. At synchrotron facilities, experiments may also encounter flux limits, for instance during a scan of an operating battery, where radiation heats up the battery and can disrupt the process under investigation.

In order to complete more scans faster, ways of minimizing the acquisition time are always sought for in the hospital, laboratory, and the synchrotron. The acquisition time can also face a hard limit. In 4D tomography, the speed of the interior dynamics of the object restricts the acquisition time of a single time step, as the reconstruction becomes blurred if too much movement occurs during its acquisition. An example of interior dynamics is shown in Figure 1.1, where a water bubble appears in a fuel cell in the course of a second, and intermediate reconstructions are necessary to track its growth.

The acquisition of a single time step is usually accelerated by reducing the angular sampling rate (resulting in undersampling) or by reducing the exposure time of each projection image. Decreasing the exposure time has a limit, however, as most detectors have a maximum frame rate [131]. This imposes a maximum on the angular sampling rate when the acquisition time is fixed.

In laboratory micro-CT, the achievable resolution in the reconstruction can be limited by the object size. Typically, the projection images can be magnified by moving the object closer to the source. If the object is too large however, its projection may not fall entirely within the detector, which can lead to truncation artifacts in the reconstruction. This imposes a maximum magnification, and therefore a tension exists between object size and achievable reconstruction resolution.

In summary, several factors influence the quality of the reconstructed images that must be balanced to take into account hardware limits, total scan time, and acceptable radiation dose.

1.1.3 Noise statistics in absorption-contrast tomography

As discussed above, noise and noise statistics play an important role in tomography, and specifically in the algorithms proposed in this thesis. We use the properties of

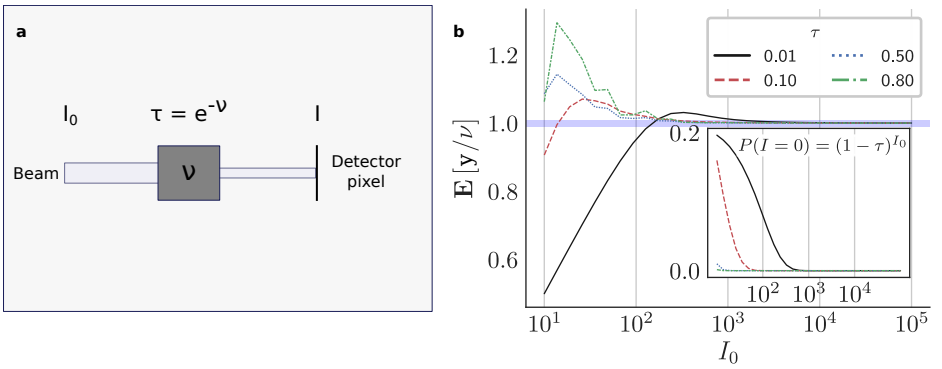


Figure 1.3: Absorption of an X-ray beam in a single-pixel detector setup. (**a**), An X-ray beam with initial intensity I_0 passes through a uniform mass with absorption coefficient ν. Each photon has a probability τ of being transmitted through the mass. (**b**), The relative bias in y due to non-linear effects in post-processing as a function of I_0, the emitted photon count. The horizontal blue bar indicates a 1% offset from ν. In the inset, the effect of photon starvation is plotted for the same values of I_0.

the noise in tomographic imaging to improve reconstructions using deep learning. One specific feature that the noise is assumed to have is that it is unbiased. A random variable a is said to be an unbiased estimate of b if

$$b = \mathbb{E}[\mathsf{a}] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} a_i, \tag{1.1}$$

i.e., if the average value of an increasing number of measurements $a_i \sim \mathsf{a}$ converges to b [178], where the second equality is a consequence of the strong law of large numbers. In this section, using an example of a single-pixel detector, we show that under reasonable conditions the bias of the noise is very mild.

First, we introduce the single-pixel detector setup and describe how absorption can be modeled as a random process. Next, we describe how the absorption process can become biased as a result of photon starvation and Jensen's inequality and describe conditions under which the bias is negligible. Finally, we discuss other sources of randomness in the acquisition process.

Suppose that an X-ray beam is measured on a single pixel detector, after it has passed through a uniform mass. Furthermore, suppose that the X-ray source emits I_0 photons and that the mass of unit width has an attenuation coefficient ν, describing its tendency to absorb photons. Each individual photon has a probability $\tau = e^{-\nu}$ to be transmitted through the object as a result of Beer-Lambert's law of attenuation [31]. The number of photons detected at the detector, I, is therefore $B(I_0, \tau)$ binomially distributed, with I_0 the number of emitted photons. The goal is to recover the attenuation coefficient ν from the measured number of photons I. This setup is illustrated in Figure 1.3 (**a**).

The expection of I has an exponential relation to ν, i.e.,

$$\mathbb{E}\left[\mathsf{I}\right] = \sum_{k=0}^{I_0} \tau^k (1-\tau)^{I_0-k} \binom{I_0}{k} k = I_0 \tau = I_0 e^{-\nu}. \tag{1.2}$$

In practice, the attenuation coefficient ν is therefore recovered by taking the logarithm, i.e,

$$\nu \approx \mathsf{y} := -\ln \frac{\mathsf{I}}{I_0}, \tag{1.3}$$

where we define the attenuation projection value y as the logarithmic fraction on the right-hand side.

The question is whether y is an unbiased estimate of ν. Since Equation (1.3) is non-linear, it is not self-evident that the noise is unbiased, i.e., $\mathbb{E}\left[\mathsf{y}\right] = \nu$. Two phenomena exert an opposite bias on this expectation: photon starvation and Jensen's inequality.

Photon starvation occurs when a pixel does not collect any photons, i.e., the event $\{\mathsf{I} = 0\}$. This event occurs with probability

$$P(\mathsf{I} = 0) = (1-\tau)^{I_0}. \tag{1.4}$$

When this happens, calculating y is more difficult, since the logarithm of zero is undefined. A practical solution is to set the photon count to 1 before taking the logarithm. This introduces a downward bias in the statistics of y.

The other effect on the expected attenuation projection function is purely mathematical. Since the negative logarithm is a convex function, Jensen's inequality [66] states that we have

$$\nu = -\ln \frac{\mathbb{E}\left[\mathsf{I}\right]}{I_0} \leq \mathbb{E}\left[-\ln \frac{\mathsf{I}}{I_0}\right] = \mathbb{E}\left[\mathsf{y}\right]. \tag{1.5}$$

Here, the equality on the left-hand side is a direct result of Equation (1.2). Hence, Jensen's inequality exerts an upward bias on y and the photon starvation correction exerts a downward bias.

The effect of photon starvation and Jensen's inequality is illustrated in Figure 1.3. Here, the measurement process has been repeated a million times to obtain an accurate estimate of $\mathbb{E}\left[\mathsf{y}\right]$. For small values of the transmittance τ, photon starvation correction exerts a strong downward bias at smaller values of I_0. For larger values of τ, the upward bias slowly reduces, until at roughly 1000 emitted photons, the bias is mostly removed. As a rule of thumb, for non-highly attenuation objects, at 1000 emitted photons per pixel there appears to be little bias in the noise.

For the sake of argument, some aspects of the tomographic acquisition pipeline have been simplified. In the standard treatment, the measurement of a photon is a statistical process that can be divided into three parts [31, 192]: (1) the probability of a photon being emitted from the source, (2) the probability of an emitted photon being transmitted through the object, and (3) the probability of

a transmitted photon being detected by the detector. The emitted number of photons I_0 per time step is typically Poisson-distributed [192]. The binomial process describing the absorption of photons by the object commonly approximated using a Poisson distribution as emission rate I_0 is typically large [31]. The measurement process on the detector can be modeled using a Poissonian-Gaussian model that is influenced by the quantum efficiency of converting photons to electrons (Poisson) and electronic noise (Gaussian) due to thermal noise or bias currents [55, 192]. In the regime where the absorption process is unbiased, however, these effects are minor.

1.1.4 Formulation of the inverse problem

Many common tomography setups can be modeled as a collection of line integrals through space where the ith measurement $y_i \in \mathbb{R}$ is obtained as a line integral

$$y_i = \int_{\mathbb{R}} x(\mathbf{s}_i + t\boldsymbol{\eta}_i)\ \mathrm{d}t \tag{1.6}$$

through a point $\mathbf{s}_i \in \mathbb{R}^3$ with direction $\boldsymbol{\eta}_i \in \mathbb{R}^3$ [173].

The canonical case is absorption contrast tomography, which is also used in medical applications. In absorption contrast tomography, the reconstruction problem can be posed as a linear discrete inverse problem. Suppose measurements $\mathbf{y} \in \mathcal{Y} = \mathbb{R}^{N_\theta \times N_p^2}$ are acquired from N_θ positions using a square detector that is divided into N_p^2 pixels. Define the cubic reconstruction volume $\mathbf{x} \in \mathcal{X} = \mathbb{R}^{N_v^3}$ on a voxel grid and let $\mathbf{A}$ denote the projection matrix such that $\mathbf{A}_{ij}$ describes the absorption by object voxel j of the ray to measurement i. The goal is to determine the value of $\mathbf{x}$ that gave rise to the measurement

$$\mathbf{A}\mathbf{x} = \mathbf{y}. \tag{1.7}$$

An algorithm, $\mathbf{R} : \mathcal{Y} \to \mathcal{X}$, that recovers the volume $\mathbf{x}$ from the measurements $\mathbf{y}$ is known as a tomographic reconstruction algorithm.

1.1.5 Tomographic reconstruction algorithms

Several approaches to tomographic reconstruction can be distinguished. Each approach can be characterized in terms of computational speed, sensitivity to noise and undersampling, and their ability to be applied to different acquisition geometries.

Fast filtered backprojection (FBP)-type reconstruction algorithms have been developed over the years for common acquisition geometries, such as single-axis parallel-beam (FBP) [134], circular cone beam (FDK) [53], and the helical cone beam (Katsevich's algorithm) geometry [89]. Each algorithm is specific to a single acquisition geometry. They are well-suited for fast, parallel computation [140] and are derived from a continous formulation of the inverse problem. These algorithms typically consist of two steps. First, the acquired projection data $\mathbf{y}$ is convolved

with a filter $\mathbf{h}$. Next, the filtered data is backprojected ($\mathbf{A}^T$) into the volume, leading to

$$\mathbf{x}_{\mathrm{FBP}} = \mathbf{A}^T (\mathbf{h} * \mathbf{y}) . \tag{1.8}$$

Reconstructions computed by FBP-type methods can suffer from significant artifacts when the number or quality of the measurements decreases. Artifacts can be remedied to an extent by altering the filter $\mathbf{h}$, making it more similar to a low-pass filter [134]. An algorithm that is similarly fast, specific to one acquisition geometry, and sensitive to noise, is the GridRec algorithm [118]. Doing away with the backprojection step, the algorithm can be considered as a non-uniform fast fourier transform [54], which aids its computational performance as images become larger [141]. The GridRec and FBP-type methods are collectively referred to as direct methods, distinguishing them from iterative and variational methods that are described next.

Iterative reconstruction algorithms aim to solve Equation 1.7 by treating it as a minimization problem that is typically formulated as

$$\mathbf{x}_{\mathrm{iter}} = \underset{\mathbf{x} \in \mathcal{X}}{\arg\min} \, \|\mathbf{A}\,\mathbf{x} - \mathbf{y}\|_2^2. \tag{1.9}$$

As the name implies, the methods operate by refining the reconstruction over a number of iterations. An elementary example is the Landweber iteration [105]

$$\mathbf{x}_0 = 0 \in \mathcal{X}, \tag{1.10}$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \eta \mathbf{A}^T (\mathbf{y} - \mathbf{A}\,\mathbf{x}_i) . \tag{1.11}$$

In each iteration, the residual error $\mathbf{y} - \mathbf{A}\,\mathbf{x}_i$ is backprojected and added to the current estimate after multiplication by a step size η. Common iterative reconstruction algorithms include ART [59], SART [6], CGLS [76], and SIRT [57, 60]. An advantage of iterative reconstruction methods is that they can be applied to any (non-standard) acquisition geometry where the forward operator $\mathbf{A}$ can be computed. Compared to direct methods however, they are substantially more computationally intensive. In addition, like direct methods, they can suffer from artifacts in case of noise or undersampling, which is typically remedied by early stopping, i.e., stopping the iteration before convergence has been reached. This effectively retains the low-frequency components of the reconstruction and moderates the high-frequency components.

Variational reconstruction algorithms aim to maximize not just the agreement with respect to the measured data, but also impose prior knowledge on the reconstruction through regularization. A common varational method in tomography is Total-Variation Minimization (TV-MIN) [168] that obtains a reconstruction through minimizing the objective

$$\mathbf{x}_{\mathrm{var}} = \underset{\mathbf{x} \in \mathcal{X}}{\arg\min} \, \|\mathbf{A}\,\mathbf{x} - \mathbf{y}\|_2^2 + \lambda g(x), \tag{1.12}$$

where the regularization function is defined by $g(x) = \|\|\nabla \mathbf{x}\|\|_1$ with ∇ the discrete gradient operation. This regularization function promotes piecewise constant reconstructions. The degree to which the solution must satisfy prior knowledge is modulated by the regularization parameter λ. Algorithms for minimizing the objective in Equation (1.12) include FISTA [15] and the Chambolle-Pock algorithm [33]. Compared to the previously described iterative reconstruction methods, variational methods tend to be slower to compute, although the addition of regularization can make them more robust to noise and undersampling. As the prior knowledge that informs the regularization function g depends on the object class being scanned, manual tuning of both g and λ may be required to obtain optimal results on new classes of objects [122]. Deep learning methods hold the promise of learning prior knowledge by example.

1.1.6 Convolutional neural networks (CNNs) in tomography

Among techniques for removing artifacts from reconstructed images, deep convolutional neural network (CNN)-based methods have shown strong results, often outperforming conventional state-of-the-art techniques [5, 81, 85, 140]. A substantial subset of these techniques are known as post-processing techniques. Here, a reconstruction is first computed using a fast reconstruction algorithm and a CNN is used to post-process the reconstructed image. Because the post-processing approach has proven to be applicable to large-scale tomographic problems [140], it is the main focus of this thesis.

CNNs are usually organized in layers: the input image is copied into the first layer, the output image is the final layer, and computations are performed in the intermediate layers. The CNN computes convolutions that are parameterized by thousands to millions of small filters (typically 3×3). In each intermediate layer, the images in the previous layer are convolved with these filters, after which a pixel-wise non-linear function is applied. A common non-linearity is the ReLU function, which is the identity for positive arguments and zero otherwise [79, 133]. A prototypical CNN is illustrated in Figure 1.4. In practice, results can be improved substantially by using more complex network structures. Throughout this thesis, we use the UNet [38, 156], DnCNN [199], and MS-D networks [140, 143], which are introduced in the relevant chapters.

A CNN is usually prepared to perform a specific image-to-image translation task by supervised training. During training, the network is presented with low-quality input images and high-quality target images from a training dataset. On each image pair, the parameters of the convolutions are optimized to minimize the training loss, i.e., the difference between the output and target image. The training loss is commonly quantified by the pixel-wise mean square error [66] and minimized using stochastic gradient descent [153].

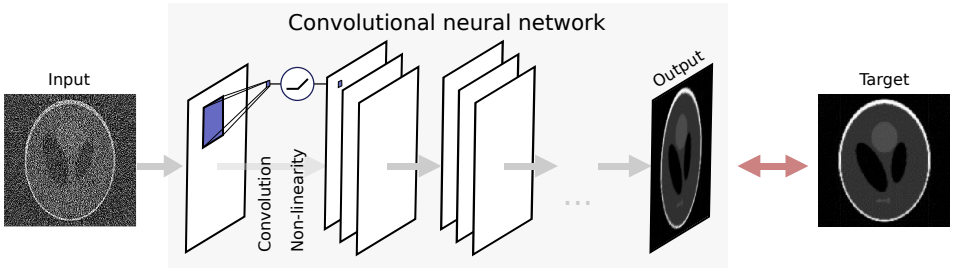


Figure 1.4: A prototypical convolutional neural network (CNN) in a supervised training arrangement for denoising. The CNN is composed of multiple layers. To compute an image in the next layer, images in the previous layer are convolved with a small filter (depicted in blue), after which a non-linear function is applied. To train a CNN to perform denoising, the output image is compared to a target image, and the convolution filters are updated accordingly.

1.1.7 CNN training as Bayesian function estimation

When formulated in a Bayesian statistical framework [9], the solution of the training minimization problem can be described in closed form. This solution can be used to predict important properties of the trained neural network, which we use in the methods developed in this thesis.

Suppose the pairs of input and target images are drawn from a prior distribution as follows

$$\mathrm{u}_i, \mathrm{v}_i \sim (\mathsf{u}, \mathsf{v}), \quad i = 1, \ldots, N, \tag{1.13}$$

where u and w are image-valued random variables. Then the solution of the training minimization problem on this dataset can be considered as approximating the regression function

$$h^* = \arg\min_h \mathbb{E}_{\mathsf{u},\mathsf{v}} \left[\|h(\mathsf{u}) - \mathsf{v}\|_2^2 \right] \approx \arg\min_h \sum_{i=1}^{N} \|h(\mathrm{u}_i) - \mathrm{v}_i\|_2^2, \tag{1.14}$$

that minimizes the expected prediction error among a class of functions [66]. In the case of the squared L_2-norm, the regression function is known [4] and equals the conditional expectation

$$h^*(u) = \mathbb{E}\left[\mathsf{v} \mid \mathsf{u} = u\right]. \tag{1.15}$$

In practice, the trained neural network does not equal h^* and an approximation is obtained. However, since equation (1.15) describes the solution in the ideal case, it can be used to predict properties of the output of trained neural networks in practice, as we show in Chapter 3.

1.1.8 Deep learning challenges in tomography applications

Tomographic applications of deep learning face a challenge in obtaining a dataset that is suitable for supervised training. We first identify three issues that underly

this challenge, followed by common strategies to work around them. First of all, as outlined in Section 1.1.2, obtaining high-quality images may be complicated, expensive, or impossible.

Second, obtaining a large quantity of similar data can be difficult. That is, data that is similar enough to the object under investigation that it can provide a useful inductive bias during neural network training. At synchrotron facilities, beam time may be limited or too expensive. Using micro-CT, scanning multiple objects may take prohibitively long (batteries, for instance). In addition, the object under investigation may be unique or expensive, for instance in the case of a historical art object. In medical applications, large quantities of similar data are available, as many scans are performed every day. Nonetheless, the collection and dissemination of large datasets may encounter ethical concerns and the normalization of data acquired using different scanners is challenging in both CT and magnetic resonance imaging (MRI) [35].

Finally, the exact pairing of the low- and high-quality datasets can be time consuming, requiring manual tuning to register the images. This is important because the accuracy of registration directly impacts the quality of the trained network [166]. When the object moves or degrades during a scan, accurate pairing is even more complicated and may require advanced deformable registration methods [49]. Therefore, deficiencies in data quality, sample quantity, and image pairing are important to address.

In practice, these issues are circumvented using several strategies, each with their own downsides. The most prevalent is the use of synthetic degradation of high-quality data. For instance, noise can be added to high-quality projection images to simulate a low dose acquisition [123]. Another tactic that is popular in MRI is the use of undersampling, where the low-quality dataset is obtained by removing a fraction of the projection images [96, 140]. Synthetic degradation solves the problem of pairing, but obtaining high-quality data remains an issue.

Another approach is to use a computationally expensive reconstruction algorithm to compute a "gold standard" reconstruction. An example of this approach is found in neutron imaging [181], where cheap FBP-type reconstructions are paired with expensive model-based reconstructions. Here, the trained networks do not exceed the reconstruction quality of the expensive reconstruction algorithm and the aim is mainly to speed up future computations of the reconstruction.

When accurate pairing is a problem, but low-quality data and high-quality data can be obtained separately, then training using cycle-consistency may be used. In this approach, two translation networks are trained. One is trained to convert the low-quality to high-quality data and the other is trained to perform the reverse transformation. Meanwhile, two discriminator networks are trained to distinguish between the outputs of the networks and real collected data. The translation networks are optimized to minimize their chance of detection by the discriminator networks and to maximize their cycle consistency: ideally, the application of the first translation network followed by the second should yield the original image. This strategy is applied to multiphase coronary CT angiography, where multiple CT measurements of the heart are taken using different levels of radiation dose [84].

In practice, however, the training process is extremely sensitive to hyperparameters and may require repetitive manual tuning to prevent optimization instability [7, 62] and mode collapse [125, 151].

Finally, synthetic data may be used to obtain a training dataset. In practice, however, it is questionable whether the data is similar and representative enough to trust the trained neural network on real-world data. This vulnerability has raised concerns about the proliferation of synthetic data in medical applications, for instance [36]. The use of synthetic data is therefore advised mainly to complement real-world data.

1.2 Research questions

In this thesis, strategies are explored to employ deep learning in situations where a large set of similar data is not available, or high-quality measurements are absent. An outline of the thesis is provided first, followed by a description of the research questions on the next pages.

In Chapter 2, we investigate how the resolution of reconstructed images can be improved using deep learning when only a single object is available. This involves an acquisition strategy where the object is scanned twice without scanning the full object at high-resolution. Using this approach, it is possible to construct a supervised training dataset containing paired low-resolution and high-resolution images, from which a neural network can be trained.

In Chapter 3, we study the properties of noise in tomography and determine whether techniques from photographic image denoising carry over into the domain tomographic of reconstructed images. In addition, we develop a deep learning technique for denoising (Noise2Inverse) that can be applied to tomography and related inverse problems. This approach does not require a supervised training dataset containing high-quality images: in fact, the method can be applied to any existing tomographic dataset.

In Chapter 4, the applicability of Noise2Inverse to real-world synchrotron data is explored. Specifically, the use of additional information from the space, time, and spectrum-like domain is investigated as a means to improve reconstruction quality.

While validating the aforementioned approaches using real-world data, a recurring problem was determining the exact geometry of the tomographic acquisition. In Chapter 2 for instance, a low-resolution and high-resolution reconstruction had to be perfectly registered (made to overlap) in order to train a neural network. This was complicated by the fact that the positions of the source and detector were not known with sufficient precision, requiring sensitive tuning after the acquisition was complete. In Chapter 4, a substantial improvement in image quality depended on a tiny deviation of the investigated object's rotation speed from its reported value that was only discovered after careful investigation.

It has therefore become apparent that software is needed that can provide support to perform small and common adjustments to standard geometries, as

well as support to devise new acquisition geometries that are increasingly used in advanced experiments. This is the topic of Chapter 5.

The chapters deal with the following research questions, which are described on a separate page each.

Research question 1. *Can deep learning be used to improve the resolution of reconstructed images using a tomographic acquisition of a single object?*

In Chapter 2, we propose a deep learning method for improving the resolution of reconstructed images of a single object. It involves a custom acquisition protocol for cone beam micro-CT that enables the reconstruction of a region of interest at low-resolution and at high-resolution. The reconstructions of the region of interest serves as a training dataset for a deep learning model. The trained model is applied to the full low-resolution reconstruction to obtain images of the whole object at high resolution.

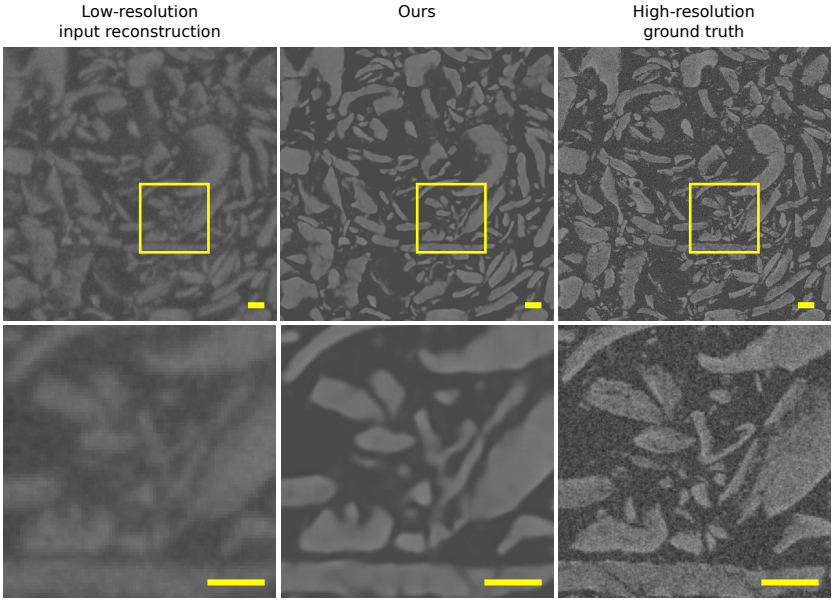


Figure 1.5: The result of applying our proposed method for resolution improvement to a dataset containing oatmeal grains. The images display a slice through a region of interest that was acquired for validation. The top row shows the original low-resolution reconstruction, the results of the proposed method, and a ground truth reconstruction that was used for validation. The bottom row shows a 4× magnification of the region indicated by the yellow square.

Research question 2A. *Many image denoising methods depend on the assumption that noise in one pixel is uncorrelated to noise in any other pixel. Can such image denoising techniques achieve similar performance on tomographic reconstructed images as they do on noisy images that satisfy the no-correlation assumption?*

In Chapter 3, the properties of noise in photographic images and in reconstructed images is compared. Many photographic image denoising methods depend on the assumption that noise in one pixel is not correlated to noise in another pixel. In tomography, however, backprojection smears out the noise in a detector pixel across a line through the reconstructed images, which may cause the noise in one pixel to be correlated to noise in other pixels of the reconstructed image. The effect of applying methods with the no-correlation assumption to tomographic images is investigated in Section 3.3.4.

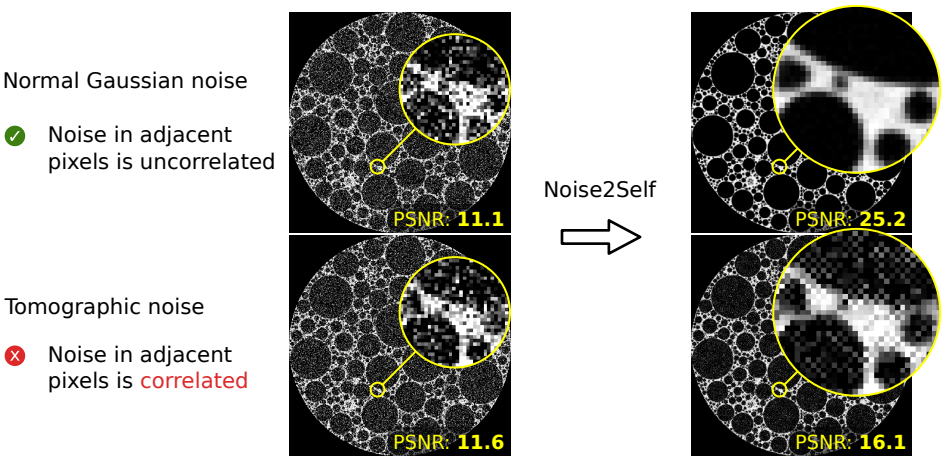


Figure 1.6: The performance of Noise2Self, a photographic image denoising method introduced in Chapter 3, is compared as it is applied to noise common to photography (Gaussian noise, top row) and to noise common to tomography (bottom row). The noise levels (PSNR with respect to ground truth) of the noisy images are similar (left column). After application of Noise2Self both images are less noisy (right column), but the PSNR improvement is substantially smaller in the case of tomographic noise.

Research question 2B. *Can deep learning be used to denoise a single 3D tomographic acquisition without any additional training data?*

In Chapter 3, we propose Noise2Inverse, a method for training deep neural networks to denoise reconstructed images using only noisy training images, i.e., no additional noise-free measurements are required. By considering the training process of neural networks in an idealized mathematical framework, the solution to the training process can be written down in simple terms, using Bayesian probabability theory. Making use of standard results on statistical independence and linearity in probability theory, an argument is made that any solution to the proposed training scheme is forced to remove noise from its input. It is demonstrated that neural networks trained using the proposed scheme indeed remove noise. In fact, they obtain better image metrics than conventional reconstruction methods on simulated data.

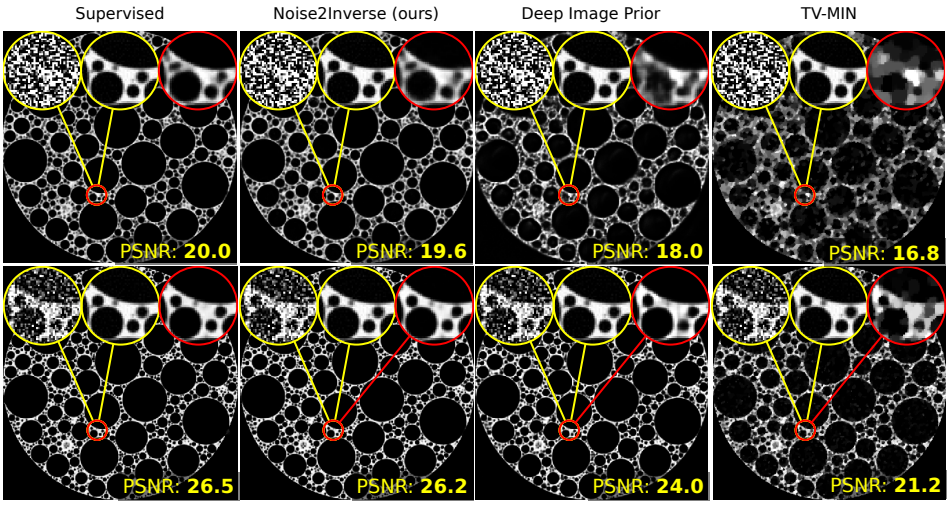


Figure 1.7: A comparison of the proposed Noise2Inverse method to other reconstruction methods on a simulated noisy foam phantom. Results are displayed of supervised training (with access to noise-free data), the proposed Noise2Inverse method (with access to only noisy data), the unsupervised Deep Image Prior, and total-variation minimization (TV-MIN). In each panel, the insets show a magnification of the noisy FBP reconstruction, the ground truth, and the result of the reconstruction method (in red). The noise level in the top row is more challenging than in the bottom row.

Research question 3. *Can self-supervised denoising be applied to real-world synchrotron-based tomographic experiments?*

In Chapter 4, we demonstrate that the Noise2Inverse method can be used to denoise challenging real-world tomographic datasets obtained at synchrotron X-ray facilities. In Chapter 3, only 2D spatial information was taken into account in applications of Noise2Inverse. In this chapter, we expand the application of Noise2Inverse in space, time, and spectrum-like domains. This development enhances applications to static and dynamic micro-tomography as well as X-ray diffraction computed tomography (XRD-CT). We show results on a micro-tomography dataset, a dynamic micro-tomography dataset, and on an X-ray diffraction computed tomography dataset. These results demonstrate the ability of Noise2Inverse to perform accurate denoising and its potential to enable a substantial reduction in acquisition time while maintaining image quality.

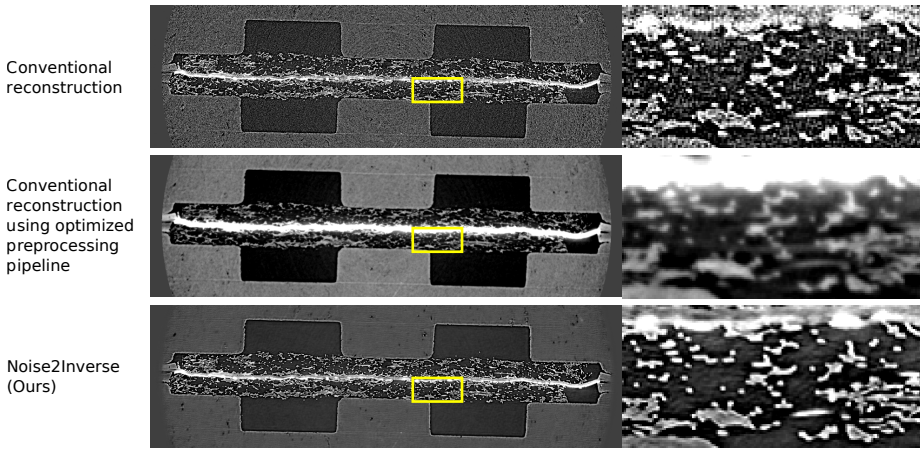


Figure 1.8: Reconstructions of a noisy synchrotron micro-tomography dataset of a fuel cell. From top to bottom, reconstructions of a horizontal slice using a conventional reconstruction algorithm (GridRec [118]), using a manually optimized preprocessing pipeline [29], and a reconstruction using Noise2Inverse. Maginifications of the region indicated by the yellow square are shown on the right.

Research question 4. *Can the implementation of reconstruction algorithms for advanced synchrotron tomography techniques be aided by a concise and efficient way to express acquisition geometries in terms of basic building blocks, primitive geometric transformations, and their compositions?*

In Chapter 5, we present the tomosipo software package. Its application programming interface (API) provides primitives to create common tomographic acquisition geometries and several composable tools to manipulate them, such as geometric transformations. In addition, the package allows making full use of the graphics processing unit (GPU) to efficiently compute reconstructions using advanced algorithms.

We demonstrate the ease of making common adjustments to an acquisition geometry, such as changing the center of rotation. In addition, the design and implementation of recently developed synchrotron-based tomography techniques is demonstrated, specifically diffraction contrast tomography (DCT) [184] and X-ray scattering tensor tomography (XSTT) [93]. Reconstructions of real-world data from synchrotron and laboratory micro-CT sources are shown, computed using several common reconstruction algorithms. Finally, bechmarks demonstrate the utility of being able to take full advantage of the GPU.

```
t = np.linspace(-1, 1, 100)   # Time t = -1.0, -.98, ..., 1
s = 2 * np.pi * t             # Angle
radius = 2                    # Radius of helix
h = 1.0                       # Vertical "speed"

vg = ts.volume()
pg = ts.cone(src_orig_dist=radius, src_det_dist=2 * radius)

R = ts.rotate(pos=0, axis=(1, 0, 0), angles=s)
T = ts.translate(axis=(1, 0, 0), alpha = h * s / (2 * np.pi))
H = T * R

ts.svg(vg, H * pg.to_vec())
```

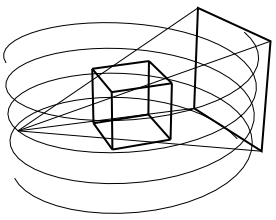


Figure 1.9: A demonstration of composing primitive geometric transforms in tomosipo. A helical acquisition geometry is defined using the code on the left and displayed on the right. First, a non-moving reconstruction volume and source-detector pair are defined. Next, a helical movement is defined by composing the primitive translation and rotation transforms. Finally, the helical transform is applied to the source-detector pair to set it in motion.