6 Conclusion and outlook

"When the press wants to burn me, I can have peace with that. But then I want to be burned for my own vision."

"Wanneer de pers me wil afbranden heb ik daar vrede mee, maar dan wil ik op mijn eigen visie afgebrand worden."

Johan Cruijff,
Voetbal International, 14 Oct 1989

The main goal of the research presented in this thesis was to develop practical deep learning techniques to improve tomographic reconstruction when acquiring a dataset of additional measurements is not feasible. To support this aim, a subordinate goal was to develop software that supports both the concise expression of complex tomographic acquisition geometries and the development of computationally efficient reconstruction algorithms. In this chapter, we summarize the contributions and limitations of this thesis and suggest directions for future research.

6.1 Contributions and limitations

The contributions of the first three chapters of this thesis can be categorized along two axes. The first is *what* factor that determines image quality is improved, i.e., resolution or noise. The second is *how* the problem of limited training data is circumvented. We propose two ways: the first is by changing the scanning protocol and the second is by changing the neural network training scheme. In Chapter 2, we propose to improve the resolution of reconstructed images by using a custom scanning protocol to create a dataset of low- and high-resolution images. In Chapters 3 and 4, we propose to remove noise from reconstructed images by training a neural network using a custom neural network training scheme.

The contributions of the final chapter cannot be categorized in this way. Instead, its contribution is a software package that makes it substantially easier to (1) compute reconstructions with a complex scanning protocol, and (2) incorporate tomographic operators in neural networks. In this way, it supports the aims of the previous chapters by enabling the custom scanning protocol and custom training scheme. This section is organized as follows. First, we describe the contributions in each chapter. Next, we describe remaining practical limitations.

In **Chapter 2**, we proposed a novel technique for improving the resolution of tomographic reconstructions. To enable application on single objects, a custom scanning procedure is developed that enables obtaining a low-resolution and a high-resolution reconstruction of a small region of the object. On this region, a data-efficient neural network is trained to transform low-resolution images to high-resolution images. The trained network can then be applied to the full low-resolution reconstruction. On both simulated and experimental data, the results demonstrate that the proposed technique is able to significantly improve the resolution of tomographic reconstructions of a single object.

The topic of **Chapter 3** was the removal of noise from reconstructed images. It consisted of an investigation of existing methods and the development of a custom training strategy. First, self-supervised deep learning methods for photographic image denoising were investigated. Such methods depend on the assumption that noise in one pixel is not correlated to noise in another pixel. An experiment was conducted in which a photographic image denoising method was applied to images with Gaussian noise and images with comparable tomographic noise. The results show that the method performs substantially worse on tomographic noise than on Gaussian noise. This failure can be explained by the fact that tomographic noise violates the no-correlation assumption.

Therefore, Noise2Inverse was developed, a method for training deep neural networks to denoise reconstructed images using only noisy training images. Starting from a single noisy tomographic acquisition, multiple reconstructions are computed from mututally non-overlapping subsamples of the projection images. Images from these separate noisy reconstruction volumes serve as input and target to train a neural network. The noise in each reconstructed volume is uncorrelated to noise in the other volumes, thereby enforcing non-correlation of the noise in the input and target images. We developed a Bayesian argument that shows that therefore such a training regime should result in a denoising neural network. Results on a diverse set of simulated and real-world data confirm that this is indeed the case: the denoising accuracy of the proposed method approaches that of a network trained using ground truth images. In addition, a comparison to state-of-the-art image denoising methods and conventional reconstruction methods, such as total-variation minimization was performed. Here, Noise2Inverse demonstrated an improvement in key image metrics, such as peak signal-to-noise ratio and structural similarity index.

In **Chapter 4**, we investigated the applicability of Noise2Inverse to tomographic imaging techniques in common use at synchrotron X-ray facilities. Specifically, we investigated applications to static and dynamic micro-tomography and X-ray diffraction tomography (XRD-CT). To obtain optimal results, we extended the training strategy in space (micro-tomography), time (dynamic micro-tomography), and spectrum-like domains (XRD-CT). In the case of static and dynamic micro-tomography, results show that Noise2Inverse offers significant improvements in the quality of denoising over alternative reconstruction methods. In the case of XRD-CT, results suggest that the acquisition time can be substantially reduced while maintaining image quality.

In **Chapter 5**, we presented the tomosipo software package. It supports the development of reconstruction algorithms for data from advanced experiments. Developing these algorithms can be slowed down by (1) the difficulty of defining the acquisition geometry, i.e., the trajectory of the source and detector, and (2) the inefficient interoperability of GPU-accelerated tomographic primitives with GPU-accelerated array libraries. This is exactly where tomosipo excels, as it allows the user to specify complex geometries in an intuitive and convenient framework. We have demonstrated the ease of making common adjustments to an acquisition geometry, e.g., changing the center of rotation, as well the ease of designing and implementing more complex geometries, e.g., X-ray diffraction and scattering setups. In addition, bechmarks demonstrate that tomosipo enables the user to overcome existing performance challenges in reconstruction algorithm development. The package enables the user to take full advantage of the GPU, achieving parity with existing optimized implementations and providing a 2–9$\times$ speed up compared to CPU-based implementations. Several common reconstruction algorithms are already implemented and demonstrated on real-world data from synchrotron and laboratory micro-CT sources.

The results in this thesis demonstrate that deep learning can be practically applied to improve resolution and noise in reconstructed images. Specifically, the developed techniques enable deep learning to be applied in situations where limited training data would previously have hampered its adoption. Some practical limitations still bear consideration though.

One challenge that becomes more prominent as a result of our work is the long training times of neural networks. As we show in Chapter 4 for instance, applying Noise2Inverse takes almost twice as long ($\sim$ 43 hours) as using a variational reconstruction algorithm ($\sim$ 30 hours) in the best case. The majority of this time is spent training the network and only a minor fraction of time is spent on applying the trained network to obtain the final result. Compared to other works that propose to train a network once and incorporate the trained network in a reconstruction algorithm, our methods are in the atypical position that network training can be part of the reconstruction algorithm itself. Therefore, a reduction in training times would be a major improvement in practice.

An issue that is not analyzed in detail is *generalization*, i.e., to which extent networks must be retrained when reconstructions of new structures are encountered. This issue is discussed in Chapter 4 in relation to scanning of additional objects under possibly different acquisition conditions. A more subtle issue arises in the super-resolution techniques of Chapter 2, where the local structure may vary throughout the object, contrary to the assumption that the structure in the region of interest is similar to the structure of the rest of the object. This may require attentiveness on the part of the user to identify changing structures and acquire a training dataset containing reconstructions of multiple regions of interest that more accurately reflect the heterogeneous structure of the object.

Another challenge is the calibration of the acquisition geometry, e.g., the center of rotation. When reconstructed images are extremely noisy, or blurred as a result of denoising, then artifacts resulting from improper calibration of the acquisition geometry are not immediately visible. In the results of Noise2Inverse, however, these artifacts are often noticeable. This could be too late, however: it may not feasible to repeat the computationally expensive neural network training with different calibration values for the acquisition geometry. Therefore, Noise2Inverse would benefit greatly from accurate and computationally efficient methods to detect subtle misconfigurations of the acquisition geometry.

Finally, there is the issue of completeness. We have proposed techniques to deal with insufficient detector resolution and measurement noise, but not angular undersampling. Measuring fewer projection images is one of the easiest ways to speed up tomographic acquisition and lower the total radiation dose. In addition, it requires fewer bits and bytes to store measured data. The data size may be a limitation in extremely fast experiments where the transmission of measurement data is bandwidth-limited [131].

6.2 Outlook

Several promising research directions are unfolding to improve image quality and reduce training time. In this section, we discuss these in turn. First, image quality can be improved by forcing the neural network output a more “likely” image, or output a distribution of several likely images. In addition, image quality can be substantially improved by proper calibration of the acquisition geometry. Finally, we discuss several approaches to reduce the training time of neural networks.

The use of the mean-square error objective during training can lead to blurred results. This is a consequence of the fact that the trained neural network approximates the conditional mean (Equation (1.15)). For instance, if an observed low-quality image is equally likely to have resulted from two different high-quality images, then traing will force a neural network to output the average of the two images [106]. In super-resolution of photographic images, methods have been developed to work around this issue using generative adversarial networks (GANs) [7, 58]. GAN-based approaches force the trained network to output an image that is likely to occur, rather than the average of likely images [84, 106]. In practice, however, this technique should be approached with some caution, as it can output a single very realistic looking result when the measurement supports multiple equally likely reconstructions. It could therefore lead a user to become overconfident in the results of a measurement.

To prevent overconfidence in a single result, multiple results can be sampled instead. Sampling approaches based on diffusion models have been demonstrated for improving resolution [160], denoising [91], and linear inverse problems such as deblurring [90], for instance. In these cases, output generation starts with Gaussian noise and repeatedly refines the noisy image using a neural network while noise of diminishing intensity is injected in the image. This process continues for up to a 100 steps [160], and under some regularity conditions the resulting image can be shown to be sampled from the posterior distribution [171]. In practice, however, applying a neural network a thousand times might be too computationally demanding.

The importance of accurate calibration of the acquisition geometry will increase as a result of faster and finer-scale acquisition and the demands of dynamic quantitative analysis. As discussed in Chapter 4, fast acquisition can cause vibrations in the measured object as a result of the high rotation speed. When tracking the internal structural dynamics of the object, it is essential to accurately overlay the 3D volumes at multiple time steps [115]. When vibration is an issue, the correct alignment of two time steps requires not just the correction of some rigid movement between the time steps, but also the correction of orientation changes between the projection images within a single time step. Iterative approaches that incorporate such calibration in the reconstruction, such as TIMBIR [3], have already been developed, but in terms of computational demands it may be desirable to develop preprocessing methods that do not require the simultaneous reconstruction of the object. Preprocessing-type methods may also be easier to combine with techniques such as Noise2Inverse and other deep learning-based methods.

Finally, to reduce the problem of long network training times, we discuss three approaches. These are based on (1) network modifications to speed up training (2) dataset modifications to improve generalization, and (3) transfer learning to jump-start network training. An instance of network modification to speed up training, is Noise2Filter [103], which uses a very small network that is adapted to tomography [144] to reduce training times to a minute or less. A drawback of these methods so far has been that the resulting image quality lags behind what can be achieved using convolutional neural networks. In Cryo-electron microscopy, a recent approach has been to train a single "super network" on a very large and diverse set of data that can ideally be applied to a variety of new datasets [18]. As this method gains more popularity, a more accurate assessment can be made whether it generalizes to unexpected datasets in practice. Finally, a hybrid between training a network from scratch and using a super network is transfer learning. Here, a network is first trained on a large and diverse dataset and the resulting trained network is retrained on new datasets for specific tasks. These techniques have been used in medical image classification [51, 63].