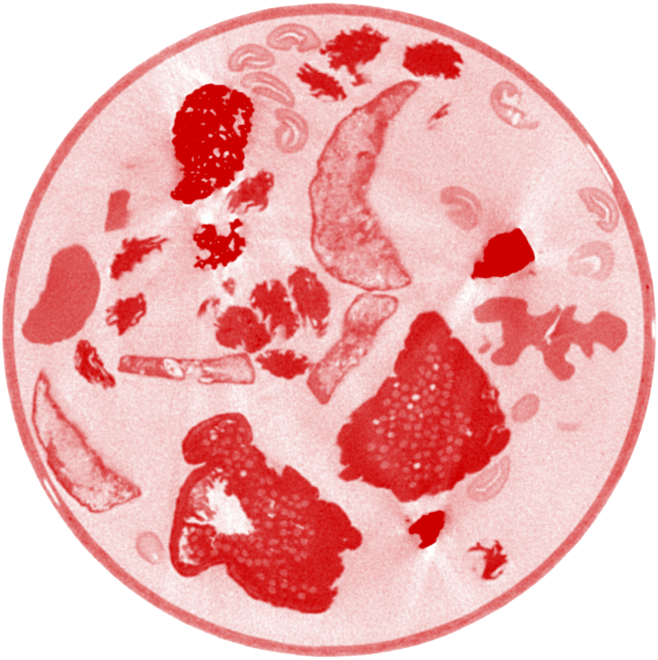
Advancing learned algorithms for 2D X-ray computed tomography

Kiss, M.B.



7
Conclusion and outlook

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

7.1 Contributions and limitations

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

In **chapter 3**, we showed the importance of tailoring the CT acquisition to the scanned samples in the context of cultural heritage objects. To make the results comprehensible not only to X-ray imaging specialists but a broad audience including museum professionals we introduced also several of the key factors that influence CT image formation. In particular, we discussed the use of beam filtration and the underlying physics of manipulating the emitted X-ray beam spectrum of a CT source to improve the image quality of CT scans of multi-material and multi-scale cultural heritage objects. We gave instructions on how this can be done with limited resources in a low-cost DIY fashion. We not only presented the improved final result but demonstrated the influence of the CT acquisition parameters on the radiographs and CT reconstructions. Illustrated by case study objects from the textile collection of the Rijksmuseum, Amsterdam, The Netherlands, we provided insights and intuitions on choosing suitable acquisition parameters and on how to design an informative CT scan. To enable future work of museum professionals building on these insights we extracted a general concept of steps to design an object-tailored CT scan for individual cases.

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

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

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

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

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

In **chapter 6**, we proposed a benchmarking framework of various machine learning algorithms for different image reconstruction tasks in X-ray computed tomography. We categorized these methods into post-processing networks, learned/unrolled iterative methods, learned regularizer methods, and plug-and-play methods. We created a pipeline for easy implementation and evaluation using key performance metrics, including SSIM and PSNR, to showcase the effectiveness of various algorithms on tasks such as full data reconstruction, limited-angle reconstruction, sparse-angle reconstruction, low-dose reconstruction, and beam-hardening corrected reconstruction. To enable reproducibility and future extensions in machine learning based CT image reconstruction research, both the dataset and toolbox are published open source. This benchmarking framework also helps to develop new methods significantly faster and to compare against different state-of-the art methods easier as time-consuming implementations of data loaders, reconstruction tasks, comparison methods and evaluation protocols do not have to be redone.

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

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

7.2 Outlook

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

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

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

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

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