Conditional Normalizing Flows for Low-Dose Computed Tomography Image Reconstruction

Alexander Denker [1] **Maximilian Schmidt** [1] **Johannes Leuschner** [1] **Peter Maass** [1] **Jens Behrmann** [1]

Abstract

Image reconstruction from computed tomography (CT) measurement is a challenging statistical inverse problem since a high-dimensional conditional distribution needs to be estimated. Based on training data obtained from high-quality reconstructions, we aim to learn a conditional density of images from noisy low-dose CT measurements. To tackle this problem, we propose a hybrid conditional normalizing flow, which integrates the physical model by using the filtered back-projection as conditioner. We evaluate our approach on a low-dose CT benchmark and demonstrate superior performance in terms of structural similarity of our flow-based method compared to other deep learning based approaches.

1. Introduction

Many important applications in medical imaging, such as computed tomography (CT) or magnetic resonance imaging (MRI), can be formulated as an inverse problem. The inverse problem consists in the reconstruction of an internal image of a patient based on radiological data. Many of these applications are ill-posed inverse problems, as small measurement errors can result in large errors in the reconstruction. In a classical way, an inverse problem is often formulated as follows: A forward operator $A : X \rightarrow Y$ maps the image $x^\dagger$ to (noisy) measurements

$$y^\delta = Ax^\dagger + \mu, \tag{1}$$

where μ describes the noise. The research in inverse problems is focused in particular on developing algorithms for obtaining stable approximations of the true solution $x^\dagger$. In

[1]University of Bremen, Center for Industrial Mathematics. Correspondence to: Alexander Denker <adenker@uni-bremen.de>, Maximilian Schmidt <maximilian.schmidt@uni-bremen.de>, Johannes Leuschner <jleuschn@uni-bremen.de>.
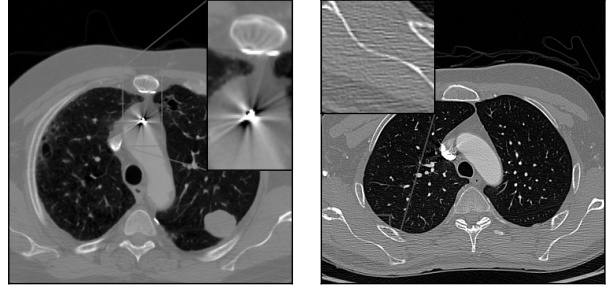
Figure 1. Ground truth samples from the LoDoPaB-CT dataset containing artifacts. These errors stem from the reconstruction technique that was used on the normal-dose measurements.

order to cover the uncertainties that occur especially with ill-posed problems, the theory of Bayesian inversion considers the posterior distribution $p(x|y^\delta)$ (Dashti & Stuart, 2017). This posterior is the conditional density of the image x conditioned on the measurements y^δ.

The main task in statistical inverse problems is to approximate this high-dimensional conditional distribution. For high-dimensional, structured images, like they arise in CT, this is a challenging process. In the field of density estimation, conditional normalizing flows (NF) (Winkler et al., 2019; Ardizzone et al., 2019b) allow to learn expressive conditional densities by maximum likelihood training. Since the physical model is known in CT (Eq. 2), we propose a hybrid approach which integrates model-based reconstruction with conditional NFs.

In many CT image reconstruction tasks the mean squared error (MSE) is used (Chen et al., 2017; He et al., 2020), which, however, has many known limitations (Zhao et al., 2017). In the context of maximum likelihood estimation (MLE), the MSE loss arises from the assumption of i.i.d. standard Gaussian noise. However, this assumption is violated in CT training data since they are often obtained from reconstructions of high-dose or normal-dose measurements. E.g. the choice of the reconstruction algorithm can lead to artifacts, as shown in Figure 1. This implies that the reconstruction error for individual pixels is no longer independent. We argue that these dependencies can be better captured by a flow-based model.

Our contributions are twofold: 1) We apply conditional normalizing flows to CT image reconstruction. 2) We propose a hybrid approach, which integrates the physical model by using the filtered back-projection as conditioner.

1.1. Related Work

Deep learning methods have been successfully applied to many ill-posed inverse problems such as CT (Arridge et al., 2019). In particular, end-to-end learned methods have been used. Those methods can be classified in three main groups: post-processing (Chen et al., 2017), fully-learned (He et al., 2020) and learned iterative algorithms (Adler & Öktem, 2018). These end-to-end methods have in common that they learn a parameterized operator $T_\theta : Y \to X$ by optimizing the parameters θ using training data $\{(y_i^\delta, x_i^\dagger)\}_{i=1}^N$. Usually, this is done by minimizing the MSE between the ground truth data $x_i^\dagger$ and the reconstruction $T_\theta(y_i^\delta)$ as

$$\hat{\theta} \in \arg \min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^N \|T_\theta(y_i^\delta) - x_i^\dagger\|^2.$$

Recently, Deep Image Priors were used for CT, achieving promising results in the low-data regime (Baguer et al., 2020). Similar to our approach is the work of (Adler & Öktem, 2018), who employed a Wasserstein GAN to draw samples from the conditional distribution. However, in this approach it is not possible to evaluate the likelihood of the generated samples. (Ardizzone et al., 2019a) have used invertible neural networks to approximate the conditional distribution and to analyze inverse problems. In a subsequent paper this concept was extended to conditional invertible neural networks (cINNs) which yielded good performance in the field of conditional image generation (Ardizzone et al., 2019b).

2. Background on Computed Tomography

Computed tomography allows for a non-invasive acquisition of the inside of the human body, which makes it one of the most important tools in modern medical imaging (Buzug, 2008). CT is a primary example of an inverse problem. The determination of the interior distribution cannot be achieved directly. It has to be inferred from the measured attenuation of X-rays sent through the body.

Current research focuses on reconstruction methods for low-dose CT measurements to reduce the health risk from radiation (Shan et al., 2019; Baguer et al., 2020). One strategy to reduce the dose is to measure at fewer angles. This can result in undersampled measurements and therefore in the existence of ambiguous solutions to the inverse problem. Another option is to lower the intensity of the X-ray. This leads to increased Poisson noise on the measurements
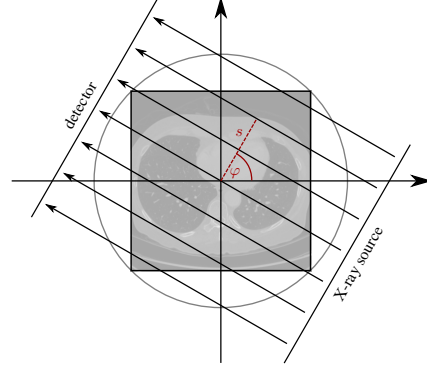

Figure 2. Schematic illustration of a CT scanner with a parallel beam geometry (Baguer et al., 2020). The scanner is rotated around the patient during the measurement.

and adds to the instability of the inversion. In this paper, we test our reconstruction model for the lower intensity case.

The basic principle of a CT machine with a parallel beam geometry can be described by the 2D Radon transform $A : X \to Y$ (Radon, 1986):

$$Ax(s, \varphi) = \int_{\mathbb{R}} x \left(s \begin{bmatrix} \cos(\varphi) \\ \sin(\varphi) \end{bmatrix} + t \begin{bmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{bmatrix} \right) \mathrm{d}t. \quad (2)$$

It is an integration along a line, which is parameterized by the distance $s \in \mathbb{R}$ and angle $\varphi \in [0, \pi]$ (cf. Figure 2). For a fixed pair (s, φ) this results in the log ratio of initial and final intensity at the detector for a single X-ray beam (Beer-Lambert's law). The whole measurement, called *sinogram*, is the collection of the transforms for all pairs (s, φ). The task in CT is to invert this process to get a reconstruction of the body. The inversion of the Radon transform is an ill-posed problem since the operator is linear and compact (Natterer, 2001). The consequences is an instable inverse mapping, which amplifies even small measurement noise.

A common inversion model is the filtered back-projection (FBP) (Shepp & Logan, 1974). The reconstruction for position (i, j) is calculated by a convolution over s and an integration along φ as

$$x(i, j) = \int_0^\pi y(s, \varphi) \star h(s)|_{s=i\cos(\varphi)+j\sin(\varphi)} \, \mathrm{d}\varphi.$$

In general, h is chosen as a high-pass filter such as the Ram-Lak filter (Ramachandran & Lakshminarayanan, 1971). In reality, we can only measure for a finite number of angles and distances. In this discrete setting the FBP only works well for a high number of measurement angles. Otherwise severe streak artifacts can appear in the reconstruction.

3. Methods

3.1. Problem Setting

To estimate conditional densities, data pairs from measurements y^δ and ground truth images $x^\dagger$ are required. In computed tomography (CT) it is not possible to obtain actual ground truth data, because no picture can be taken of the interior of the human body. Instead of using ground truth images we use reconstructions based on high-dose measurements $y^{\delta_1} = Ax^\dagger + \mu_1$, i.e. $x^{\delta_1} = T_{\text{FBP}}(y^{\delta_1})$. Because x^{δ_1} is the output of an reconstruction algorithms, it can contain artifacts and differ from the actual image $x^\dagger$, see Figure 1 for an example. In the next step we simulate low-dose CT measurements using this reconstruction as $y^{\delta_2} = Ax^{\delta_1} + \mu_2$, where $\text{Var}[\mu_2] \geq \text{Var}[\mu_1]$, since low-dose measurements are more prone to measurement noise. The training set then consists of data pairs $\{y^{\delta_2}, x^{\delta_1}\}$. An example of such a dataset is LoDoPaB-CT (Leuschner et al., 2019), which we use to benchmark our proposed conditional flow.

3.2. Normalizing Flow with FBP Conditioning

From a statistical point-of-view, an inverse problem can also be seen as a generating process $x \sim p(x|y)$ (Dashti & Stuart, 2017; Arridge et al., 2019). The task in such a statistical inverse problem is to estimate this conditional distribution. We are using conditional normalizing flows (NF) (Winkler et al., 2019) to approximate the target density $p(x|y)$. The conditional NF is composed of a series of invertible transformations $F = f_K \circ \cdots \circ f_1$. Here, every individual transformation is parameterized by θ and receives a conditional input y: $f_i = f_{\theta_i}(\cdot, y)$. This transformation defines a transport map, which converts the initial density into a simple, easy-to-sample density p_Z. This model defines a conditional density $q(x|y, \theta)$ and using the change-of-variables formula the conditional density can be calculated:

$$q(x|y; \theta) = p_Z(F_\theta(x; y)) \left| \det\left(\frac{\partial F_\theta(x; y)}{\partial x} \right) \right|.$$

We denote the Jacobian for one data point x_i, y_i with $J_i = \frac{\partial F_\theta(x_i; y_i)}{\partial x}$. Instead of directly using the measurements y_i as conditional inputs, we propose to employ a reconstruction, e.g. the filtered back-projection $\hat{x}_i = T_{\text{FBP}}(y)$.

Assume a dataset $\{(x_i, y_i)\}_{i=1}^N$ of measurements y_i and reconstructions x_i. To approximate the target density $p(x|y)$ a conditional NF $q(x|y, \theta)$ can be trained using the negative log-likelihood as a loss function. Using a standard normal distribution, i.e. $p_Z \sim \mathcal{N}(0, I)$, this amounts to minimizing

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^N \frac{\|F_\theta(x_i; T_{\text{FBP}}(y_i))\|_2^2}{2} - \log|\det J_i|.$$

3.3. Conditional coupling layers

We are using the conditional coupling layer from (Ardizzone et al., 2019b) to construct a conditional invertible neural network (cINN), which is an extension of the affine coupling layer from (Dinh et al., 2017). We propose to integrate the model-driven approach of inverse problems by using the filtered back-projection $\hat{x} = T_{\text{FBP}}(y)$ as conditional input instead of the raw sinogramm measurements y. The input $u = [u_1, u_2]$ to an affine coupling layer is split into two parts and both parts are transformed individually:

$$v_1 = u_1 \odot \exp(s_1(u_2, \hat{x})) + t_1(u_2, \hat{x})$$
$$v_2 = u_2 \odot \exp(s_2(v_1, \hat{x})) + t_2(v_1, \hat{x}).$$

The transformations s_1, s_2, t_1, t_2 do not need to be invertible and are modelled as convolutional neural networks (CNNs). The inverse of an affine coupling layer is:

$$u_1 = (v_1 - t_1(u_2, \hat{x})) \odot \exp(-s_1(u_2, \hat{x}))$$
$$u_2 = (v_2 - t_2(v_1, \hat{x})) \odot \exp(-s_2(v_1, \hat{x})).$$

The log-determinant of the Jacobian for one affine coupling layer can be calculated as the sum over s_i, i.e. $\sum_j s_1(u_2, \hat{x})_j + \sum_j s_2(v_1, \hat{x})_j$. A deep invertible network can be built as a sequence of multiple such layers, with a permutation of the dimensions after each layer.

The conditional input $\hat{x}$ is added as an extra input to each transformation in the coupling layer. In practice, an additional conditioning network H is added, so instead of $\hat{x}$ the output $H(\hat{x})$ is used. This conditioning network H is under no architectural constraints and can contain all usual elements (i.e. BatchNorm, pooling layer, etc.) of a CNN.

4. Results

Sampling from the model is a two-step process: First, a sample z is drawn from the base density p_Z. Second, this sample is transformed by the inverse to obtain an image. By repeatedly sampling z_j for a fixed input y^δ we thus estimate the conditional mean as

$$\hat{x} = \mathbb{E}_z[F^{-1}(z, T_{\text{FBP}}(y^\delta))] \approx \frac{1}{n} \sum_{j=1}^n F^{-1}(z_j, T_{\text{FBP}}(y^\delta)).$$

We evaluate our model on the LoDoPaB-CT dataset (Leuschner et al., 2019). For this dataset we are in the case of oversampling, so we expect a uni-modal distribution. This enables the choice of the conditional mean as the reconstruction method. If a highly multi-modal distribution is expected, the conditional mean is not the optimal choice. To measure the error between reconstruction and ground truth, the PSNR and the SSIM (Wang et al., 2004) are evaluated. Both are common quality metrics for the evaluation of CT and MRI reconstructions (Joemai & Geleijns, 2017;
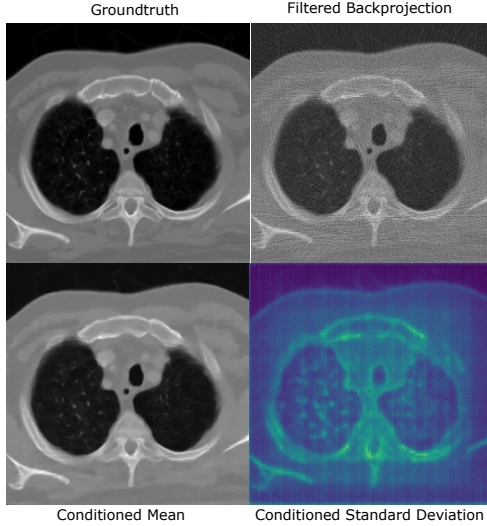
Groundtruth | Filtered Backprojection

Conditioned Mean | Conditioned Standard Deviation

Figure 3. Reconstruction and standard deviation of cINN. 1000 Samples were used for the reconstruction. The top row shows the ground truth image and the corresponding FBP.

Adler & Öktem, 2018; Zbontar et al., 2018). The PSNR is a pixel-wise metric which is defined via the MSE. The SSIM is a structural metric, which compares local patterns of pixels and is not calculated on a per pixel basis.

4.1. Implementation

We follow the multi-scale architecture design of RealNVP (Dinh et al., 2017). After each block, consisting of 6 coupling layers, downsampling is performed. The downsampling is done using the Haar downsampling from (Ardizzone et al., 2019b) and the variant used in (Jacobsen et al., 2018). The dimensions have to be permuted after each coupling layer. This is done using the invertible 1x1 convolutions from (Kingma & Dhariwal, 2018). The model is implemented using the library FrEIA[1]. A conditioning network was used to further extract features from the filtered back-projection. This conditioning network was trained together with the full flow-based model. Details on the implementation can be found in the supplementary material.

4.2. LoDoPaB-CT Dataset

We evaluate our method on the low-dose parallel beam (LoDoPaB) CT dataset (Leuschner et al., 2019), which contains over $40\,000$ two-dimensional CT images and corresponding simulated low photon count measurements. The ground truth images x^{δ_1} are human chest CT reconstructions from the LIDC/IDRI database (Armato III et al., 2011), cropped to 362×362 pixels. Projections are computed using parallel beam geometry with 1000 angles and

---

[1]https://github.com/VLL-HD/FrEIA

513 beams. Poisson noise is applied to model a low photon count (μ_2 in Section 3.1).

4.3. Evaluation on LoDoPaB-CT

We have evaluated our model on the LoDoPaB-CT dataset. First we examined the dependence of PSNR and SSIM on the number of samples for the conditional mean. The results are shown in Figure 4 (appendix). Both PSNR and SSIM increase with a higher number of samples. This allows for a trade-off between quality of reconstruction and time. For our evaluation we have chosen a conditional mean with 1000 samples. Table 1 shows the scores on the test dataset. The comparison includes several classical and deep learning approaches.

In terms of PSNR our model is comparable to other state-of-the-art deep learning approaches, despite not explicitly trained to minimize the MSE between the prediction and the ground truth images. Regarding the SSIM our model outperforms all other approaches. This further underlines our hypothesis that using the more flexible flow objective enable to incorporate structural properties.

Model	PSNR	SSIM
FBP (Shepp & Logan, 1974)	30.52	0.74
FBP + U-Net (Jin et al., 2017)	35.90	0.85
TV Regularization	32.27	0.78
DIP + TV (Baguer et al., 2020)	34.79	0.82
iRadonMap (He et al., 2020)	31.23	0.76
Learned GD (Adler & Öktem, 2017)	34.67	0.82
Learned PD (Adler & Öktem, 2018)	**36.12**	0.85
cINN (Ours)	35.68	**0.88**

Table 1. Results on the LoDoPaB-CT test set. The baseline methods were evaluated in (Baguer et al., 2020).

5. Conclusion

We have investigated how flow-based models can be applied as a conditional density estimator for the reconstruction of low-dose CT images. Using this generative approach, we were able to obtain high-quality reconstructions that outperformed all other deep learning approaches in terms of structural similarity. So far only coupling-based INNs were used, but future work should explore other architectures such as i-ResNets (Behrmann et al., 2019) for this conditional density estimation task. Furthermore, our hybrid approach that integrates the physical model into the conditioning could enable the use of more advanced reconstruction algorithms. Thus, conditional flows are a promising avenue for statistical model-based inverse problems such as CT reconstruction.

References

Adler, J. and Öktem, O. Deep bayesian inversion. *arXiv preprint, arXiv:1811.05910*, 2018.

Adler, J. and Öktem, O. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6): 1322–1332, 06 2018. doi: 10.1109/TMI.2018.2799231.

Adler, J. and Öktem, O. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, 33(12):124007, 11 2017. doi: 10.1088/1361-6420/aa9581.

Ardizzone, L., Kruse, J., Rother, C., and Köthe, U. Analyzing inverse problems with invertible neural networks. In *7th International Conference on Learning Representations, ICLR*, 2019a.

Ardizzone, L., Lüth, C., Kruse, J., Rother, C., and Köthe, U. Guided image generation with conditional invertible neural networks. *arXiv preprint, arXiv:1907.02392*, 2019b.

Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., van Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., Max, D., Pais, R. C., Qing, D. P.-Y., Roberts, R. Y., Smith, A. R., Starkey, A., Batra, P., Caligiuri, P., Farooqi, A., Gladish, G. W., Jude, C. M., Munden, R. F., Petkovska, I., Quint, L. E., Schwartz, L. H., Sundaram, B., Dodd, L. E., Fenimore, C., Gur, D., Petrick, N., Freymann, J., Kirby, J., Hughes, B., Vande Casteele, A., Gupte, S., Sallam, M., Heath, M. D., Kuhn, M. H., Dharaiya, E., Burns, R., Fryd, D. S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B. Y., and Clarke, L. P. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. *Med. Phys.*, 38(2):915–931, 02 2011. ISSN 0094-2405. doi: 10.1118/1.3528204.

Arridge, S., Maass, P., Öktem, O., and Schönlieb, C.-B. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019. ISSN 0962-4929. doi: 10.1017/S0962492919000059.

Baguer, D. O., Leuschner, J., and Schmidt, M. Computed tomography reconstruction using deep image prior and learned reconstruction methods. *arXiv preprint, arXiv:2003.04989*, 2020.

Behrmann, J., Grathwohl, W., Chen, R. T. Q., Duvenaud, D., and Jacobsen, J.-H. Invertible residual networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 573–582, 2019.

Buzug, T. *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*. Springer Berlin Heidelberg, 2008. ISBN 9783540394082. doi: 10.1007/978-3-540-39408-2.

Chen, H., Zhang, Y., Kalra, M. K., Lin, F., Chen, Y., Liao, P., Zhou, J., and Wang, G. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging*, 36 (12):2524–2535, 12 2017. ISSN 0278-0062. doi: 10.1109/TMI.2017.2715284.

Dashti, M. and Stuart, A. M. *The Bayesian Approach to Inverse Problems*, pp. 311–428. Springer International Publishing, Cham, 2017. ISBN 978-3-319-12385-1. doi: 10.1007/978-3-319-12385-1_7.

Dinh, L., Sohl-Dickstein, J., and Bengio, S. Density estimation using real NVP. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

He, J., Wang, Y., and Ma, J. Radon inversion via deep learning. *IEEE Transactions on Medical Imaging*, 39 (6):2076–2087, 2020.

Jacobsen, J.-H., Smeulders, A., and Oyallon, E. i-RevNet: Deep invertible networks. *ICLR 2018 - International Conference on Learning Representations*, 2018.

Jin, K. H., McCann, M. T., Froustey, E., and Unser, M. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 09 2017. ISSN 1057-7149. doi: 10.1109/TIP.2017.2713099.

Joemai, R. M. S. and Geleijns, J. Assessment of structural similarity in CT using filtered backprojection and iterative reconstruction: a phantom study with 3D printed lung vessels. *The British Journal of Radiology*, 90 (1079):20160519, 2017. doi: 10.1259/bjr.20160519. PMID: 28830200.

Kingma, D. P. and Ba, J. Adam: A Method for Stochastic Optimization. *arXiv preprint, arXiv:1412.6980*, 12 2014.

Kingma, D. P. and Dhariwal, P. Glow: Generative flow with invertible 1x1 convolutions. In Bengio, S., Wallach, H. M., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 31*, pp. 10236–10245, 2018.

Leuschner, J., Schmidt, M., Baguer, D. O., and Maaß, P. The LoDoPaB-CT Dataset: A benchmark dataset for low-dose CT reconstruction methods. *arXiv preprint, arXiv:1910.01113*, 2019.

Natterer, F. *The mathematics of computerized tomography*. Classics in applied mathematics ; 32. Society for Industrial and Applied Mathematics, Philadelphia, 2001. ISBN 9780898714937. doi: 10.1137/1.9780898719284.

Radon, J. On the determination of functions from their integral values along certain manifolds. *IEEE Transactions on Medical Imaging*, 5(4):170–176, 12 1986. ISSN 0278-0062. doi: 10.1109/TMI.1986.4307775.

Ramachandran, G. N. and Lakshminarayanan, A. V. Three-dimensional reconstruction from radiographs and electron micrographs: Application of convolutions instead of fourier transforms. *Proceedings of the National Academy of Sciences*, 68(9):2236–2240, 1971. ISSN 0027-8424. doi: 10.1073/pnas.68.9.2236.

Shan, H., Padole, A., Homayounieh, F., Kruger, U., Khera, R. D., Nitiwarangkul, C., Kalra, M. K., and Wang, G. Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction. *Nature Machine Intelligence*, 1(6):269–276, 2019. ISSN 2522-5839. doi: 10.1038/s42256-019-0057-9.

Shepp, L. A. and Logan, B. F. The fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, 21(3):21–43, 1974.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 04 2004. ISSN 1057-7149. doi: 10.1109/TIP.2003.819861.

Winkler, C., Worrall, D., Hoogeboom, E., and Welling, M. Learning likelihoods with conditional normalizing flows. *arXiv preprint, arXiv:1912.00042*, 2019.

Zbontar, J., Knoll, F., Sriram, A., Muckley, M. J., Bruno, M., Defazio, A., Parente, M., Geras, K. J., Katsnelson, J., Chandarana, H., Zhang, Z., Drozdzal, M., Romero, A., Rabbat, M., Vincent, P., Pinkerton, J., Wang, D., Yakubova, N., Owens, E., Zitnick, C. L., Recht, M. P., Sodickson, D. K., and Lui, Y. W. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint, arXiv:1811.08839*, 2018.

Zhao, H., Gallo, O., Frosio, I., and Kautz, J. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 2017. doi: 10.1109/TCI.2016.2644865.

6. Appendix

6.1. Model Architecture

The model was trained for 15.000 stochastic gradient steps of batchsize 10 with the Adam-Optimizer (Kingma & Ba, 2014) using a weight decay of 10^{-5}. The last layer in the subnetworks of each coupling layer is initialized with zero. This initializes the model as a whole with the identity.

A multiscale architecture was used for implementation of the cINN. The model includes 6 resolution levels. After each level a part of the channels is split off and passed on to the output. After each resolution level downsampling is performed. Downsampling was performed using the iRevNet variant (Jacobsen et al., 2018) as well as the Haar Downsampling by (Ardizzone et al., 2019b). The input size of the CT-images is $1 \times 352 \times 352$. The full cINN was build as follows.

cINN	Output size
iRevNet-Downsampling	$4 \times 176 \times 176$
level 1 conditional section	$4 \times 176 \times 176$
iRevNet-Downsampling	$16 \times 88 \times 88$
Split: $8 \times 88 \times 88$ to output	$8 \times 88 \times 88$
level 2 conditional section	$8 \times 88 \times 88$
iRevNet-Downsampling	$32 \times 44 \times 44$
Split: $16 \times 44 \times 44$ to output	$16 \times 44 \times 44$
level 3 conditional section	$16 \times 44 \times 44$
iRevNet-Downsampling	$64 \times 22 \times 22$
Split: $32 \times 22 \times 22$ to output	$32 \times 22 \times 22$
level 4 conditional section	$32 \times 22 \times 22$
Haar-Downsampling	$128 \times 11 \times 11$
Split: $96 \times 11 \times 11$ to output	$32 \times 11 \times 11$
level 5 conditional section	$32 \times 11 \times 11$
Split: $28 \times 11 \times 11$ to output	$4 \times 11 \times 11$
level 6 Dense-conditional section	484

A conditioning network was used to extract features from the conditional input. Similar to the cINN, this network consists of 6 resolution levels. The output from the resolution level of the conditioning network is used as the conditioning input for the respective resolution level in the cINN. If not specified otherwise, a kernel size of $k = 3$ is used. In addition, batch normalization (BN) is applied.

Conv2d: $1 \rightarrow 3$, stride=2 + LeakyReLU
Conv2d: $32 \rightarrow 64$ + LeakyReLU
Conv2d: $64 \rightarrow 128$ + BN + LeakyReLU
Conv2d: $128 \rightarrow 64$ + BN + LeakyReLU
Conv2d: $64 \rightarrow 32$ + BN + LeakyReLU
Conv2d: $32 \rightarrow 4$ + BN ($\rightarrow$ level 1)

LeakyRelu
Conv2d: $4 \rightarrow 32$, stride=2 + BN + LeakyReLU
Conv2d: $32 \rightarrow 32$ (k=1) + LeakyReLU
Conv2d: $32 \rightarrow 32$ + BN + LeakyReLU
Conv2d: $32 \rightarrow 8$ + BN ($\rightarrow$ level 2)

LeakyRelu
Conv2d: $8 \rightarrow 32$ (k=1) + LeakyReLU
Conv2d: $32 \rightarrow 64$, stride = 2+ LeakyReLU
Conv2d: $64 \rightarrow 16$ + BN ($\rightarrow$ level 3)

LeakyRelu
Conv2d: $16 \rightarrow 64$ (k=1) + LeakyReLU
Conv2d: $64 \rightarrow 64$, stride = 2+ LeakyReLU
Conv2d: $64 \rightarrow 32$ + BN ($\rightarrow$ level 4)

LeakyRelu
Conv2d: $32 \rightarrow 96$ (k=1) + LeakyReLU
Conv2d: $96 \rightarrow 128$, stride = 2+ LeakyReLU
Conv2d: $128 \rightarrow 32$ (k=1) + BN ($\rightarrow$ level 5)

LeakyRelu
Conv2d: $32 \rightarrow 64$, stride=2 + LeakyReLU
Conv2d: $64 \rightarrow 256$, stride = 2+ LeakyReLU
Average Pooling + Flatten + BN ($\rightarrow$ level 6)

To implement the subnetworks in the coupling layers a CNN variant and a fully connected variant were used. The input channels are denoted by c_{in} and the output channels by c_{out}.

CNN-subnetwork (k=1) or (k=3)
Conv2d: $c_{in} \rightarrow 64$, + LeakyReLU
Conv2d: $64 \rightarrow 92$ + LeakyReLU
Conv2d: $92 \rightarrow c_{out}$

Dense-subnetwork
Dense: $c_{in} \rightarrow 512$, + LeakyReLU
Dense: $512 \rightarrow 512$ + LeakyReLU
Dense: $512 \rightarrow c_{out}$

Using this two variants of subnetworks the conditional sections are implemented as follows.

conditional section	
Coupling (CNN-subnet k=1)	
Glow 1×1 convolution	
Coupling (CNN-subnet k=3)	3x
Glow 1×1 convolution	

dense conditional section	
Random permutation	
Dense-subnetwork	4x

After each downsampling a small unconditioned subnetwork is used:

CNN-subnetwork (without conditional input)
Conv2d: $c_{in} \rightarrow 64$ (k=1), + LeakyReLU
Conv2d: $64 \rightarrow 64$ (k=1) + LeakyReLU
Conv2d: $64 \rightarrow c_{out}$ (k=1)

The downsampling section is built as follows:

Downsample section (Haar or iRevNet)
Haar or iRevNet downsampling
Glow 1×1 convolution
Coupling (unconditional CNN-subnetwork)
Glow 1×1 convolution
Coupling (unconditional CNN-subnetwork)

6.2. Evaluation of the Conditional Mean

We have used the conditional mean as a reconstruction for the CT image. Figure 4 shows the performance in relation to the number of samples used.
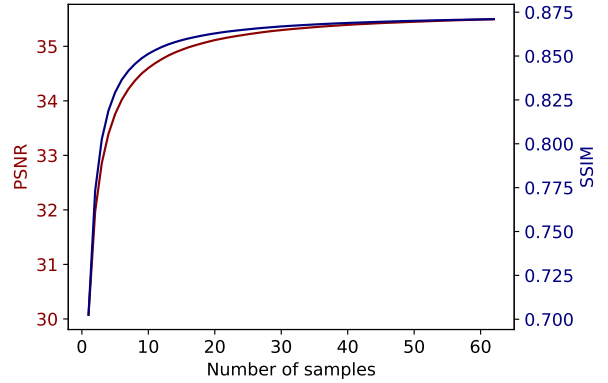

Figure 4. PSNR and SSIM for the validation set of the LoDoPaB-CT dataset. The PSNR is colored in red and the SSIM is colored in blue.

6.3. Additional Examples

Groundtruth   Filtered Backprojection   Conditioned Mean