

Unrolled three-operator splitting for parameter-map learning in Low Dose X-ray CT reconstruction

Andreas Kofler¹, Fabian Altekrüger^{2,3}, Fatima Antarou Ba³, Christoph Kolbitsch¹, Evangelos Papoutsellis^{4,*}, David Schote¹, Clemens Sirotenko⁵, Felix Frederik Zimmermann¹, and Kostas Papafitsoros⁶

¹Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany

²Humboldt-Universität zu Berlin, Department of Mathematics, Berlin, Germany

³Technische Universität Berlin, Institute of Mathematics, Berlin, Germany

⁴Finden Ltd, Rutherford Appleton Laboratory, Harwell Campus, Didcot, United Kingdom

⁵Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

⁶School of Mathematical Sciences, Queen Mary University of London, United Kingdom

*Corresponding author: epapoutsellis@gmail.com

Abstract We propose a method for fast and automatic estimation of spatially dependent regularization maps for total variation-based (TV) tomography reconstruction. The estimation is based on two distinct sub-networks, with the first sub-network estimating the regularization parameter-map from the input data while the second one unrolling T iterations of the Primal-Dual Three-Operator Splitting (PD3O) algorithm. The latter approximately solves the corresponding TV-minimization problem incorporating the previously estimated regularization parameter-map. The overall network is then trained end-to-end in a supervised learning fashion using pairs of clean-corrupted data but crucially without the need of having access to labels for the optimal regularization parameter-maps.

1 Introduction

Over recent years, Low Dose X-ray Computed Tomography (LDCT) has received a growing interest in the medical imaging field due to its ability to reduce the radiation dose. Patients are exposed to low levels of radiation by reducing the energy of the photons emitted from the X-ray source. Using traditional and analytic reconstruction methods such as filtered back projection (FBP), several imaging artifacts are introduced, compromising the quality of the reconstructed image and clinical diagnosis.

To overcome this problem, iterative reconstruction methods have been proposed such as algebraic reconstruction technique (ART), simultaneous algebraic reconstruction technique (SART) and projection onto convex sets (POCS). In addition, such reconstruction procedures often require the use of regularization methods in order to eliminate noise and artifacts, such as for instance, the well-known Tikhonov and Total Variation (TV) regularization [1].

The acquired measured tomography data can be described by the equation $\mathbf{z} = \mathbf{A}\mathbf{x}_{\text{true}} + \mathbf{e}$, where $\mathbf{x}_{\text{true}} \in \mathbb{R}^n$ is the ground truth image, $\mathbf{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear operator which models the data-acquisition process, i.e. the discretized Radon transform, and $\mathbf{e} \in \mathbb{R}^m$ denotes some random noise component. Regularized iterative methods solve minimization problems of the form

$$\min_{\mathbf{x}} \mathcal{D}(\mathbf{A}\mathbf{x}, \mathbf{z}) + \mathcal{R}(\mathbf{x}), \quad (1)$$

where $\mathcal{D}(\cdot, \cdot)$ denotes a data-discrepancy measure and $\mathcal{R}(\cdot)$ a regularization term. A classical example is the TV tomography reconstruction problem under Gaussian noise which can be written as

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{z}\|_2^2 + \lambda \|\nabla \mathbf{x}\|_1 + \mathbb{I}_{\{\mathbf{x} > 0\}}(\mathbf{x}). \quad (2)$$

A key factor which impacts the quality of the reconstructed image is the careful choice of the regularization parameter λ which balances the strength between the regularization and the data fidelity term. Underestimating λ yields poor regularization, while overestimating it results in smooth images with an artificial “cartoon-like” appearance. Particularly in medical imaging applications, where images are at the basis of diagnostic decisions and therapy planning, a proper choice of any regularization parameter is crucial.

Employing a single scalar parameter λ implies that the regularization is enforced with equal strength for each pixel/voxel. Depending on the application, this might be undesirable due to different features contained in the image. In this case, one can replace the scalar parameter with a spatially varying, i.e. a pixel/voxel dependent one, denoted now by $\Lambda \in \mathbb{R}_+^q$. Here, q denotes the number of directions for which the partial derivatives are computed. Implementation-wise, Λ corresponds to a stack of diagonal operators which contain a different regularization parameter for every single pixel/voxel in the respective gradient domain of the image. Then, the resulting problem has the form

$$\min_{\mathbf{x}} \mathcal{D}(\mathbf{A}\mathbf{x}, \mathbf{z}) + \|\Lambda \nabla \mathbf{x}\|_1 + \mathbb{I}_{\{\mathbf{x} > 0\}}(\mathbf{x}). \quad (3)$$

However, this problem requires a precise data-adaptive estimation of the spatially varying parameter-map Λ which is a highly non-trivial task.

2 Methods

One approach for the automatic estimation of the spatially varying regularization parameter Λ is employing bilevel

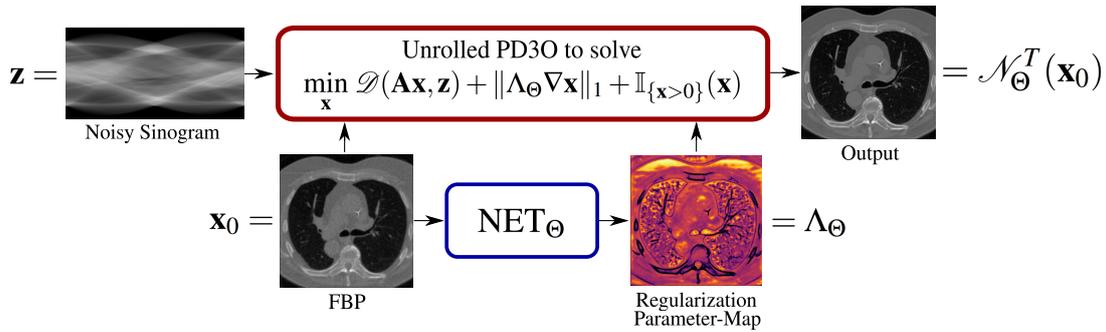


Figure 1: Network architecture for the LDCT reconstruction problem. It consists of a sub-network NET_Θ that estimates the regularization parameter-map (blue), and a sub-network that unrolls T iterations of the PD3O algorithm (red).

optimization techniques. Given M pairs of measured data and the corresponding ground truth $(\mathbf{z}_i, \mathbf{x}_{\text{true}}^i)_{i=1}^M$, the general bilevel formulation is

$$\begin{cases} \min_{\Lambda} \sum_{i=1}^M l(\mathbf{x}^i(\Lambda), \mathbf{x}_{\text{true}}^i) \text{ subject to} \\ \mathbf{x}^i(\Lambda) = \underset{\mathbf{x}}{\text{argmin}} \mathcal{D}(\mathbf{A}\mathbf{x}, \mathbf{z}) + \|\Lambda \nabla \mathbf{x}\|_1 + \mathbb{I}_{\{\mathbf{x}>0\}}(\mathbf{x}), \end{cases} \quad (4)$$

where l is a suitable upper level objective. For instance, if $l(x_1, x_2) = \|x_1 - x_2\|_2^2$, the bilevel problem (4) seeks to compute the parameters Λ which are “the best on average”, i.e. PSNR-maximizing, for the given M data pairs. Hence, given some new data \mathbf{z}_{test} which has been measured in a similar way as $(\mathbf{z}_i)_{i=1}^M$, solving (3) with the precomputed Λ will yield a good reconstruction.

Although this scheme has been extensively studied both for scalar and spatially varying regularization parameters, it has been mainly applied in image denoising applications, i.e. $\mathbf{A} = \mathbf{I}_n$ with Gaussian noise, see [2–4]. Further, unsupervised approaches employing upper level energies that do not depend on the ground truth \mathbf{x}_{true} , i.e., $l := l(\mathbf{x}(\Lambda))$, have also been considered in a series of works [5–8]. Even though these bilevel optimization methods are typically supported by rigorous mathematical theories, they are computationally demanding which has limited their use on tomographic problems.

2.1 An Unrolled Neural Network Framework

Here, inspired by the recent success of unrolled neural networks (NNs) [9], we consider an unrolled neural network approach in order to learn the regularization parameter Λ . The proposed framework is summarized in Figure 1 and it is outlined next.

An unrolled NN which corresponds to an implementation of an iterative scheme of finite length is constructed to approach the solution of problem (2) assuming a *fixed* regularization parameter-map. Within the unrolled NN, the regularization parameter-map is estimated from the input data via a sub-network NET_Θ and is used throughout the whole reconstruction scheme. To be more precise, given some initial estimate \mathbf{x}_0 we work with an iterative

scheme (specified in the next section)

$$\mathbf{x}_T = S^T(\mathbf{x}_0, \mathbf{z}, \Lambda, \mathbf{A}), \quad T = 0, 1, 2, \dots, \quad (5)$$

that solves (3) in the limit $T \rightarrow \infty$. Then, for some fixed number of iterations $T \in \mathbb{N}$, our unrolled NN reads as follows:

$$\begin{cases} \Lambda_\Theta = \text{NET}_\Theta(\mathbf{x}_0), \\ \mathbf{x}_1 = S^1(\mathbf{x}_0, \mathbf{z}, \Lambda_\Theta, \mathbf{A}), \\ \vdots \\ \mathbf{x}_T = S^T(\mathbf{x}_0, \mathbf{z}, \Lambda_\Theta, \mathbf{A}). \end{cases} \quad (6)$$

Here, NET_Θ denotes a U-Net [10] with learnable parameters Θ . We denote by \mathcal{N}_Θ^T the overall resulting network, i.e.

$$\mathcal{N}_\Theta^T(\mathbf{x}_0) = S^T(\mathbf{x}_0, \mathbf{z}, \Lambda_\Theta, \mathbf{A}) = S^T(\mathbf{x}_0, \mathbf{z}, \text{NET}_\Theta(\mathbf{x}_0), \mathbf{A}).$$

The unrolled NN can then be end-to-end trained in a supervised manner on a set of input-target image-pairs. This resulting network can be identified as a pipeline that combines in a sequential way 1) the estimation of the regularization parameter-map which is adapted to the data \mathbf{z} (and hence in medical imaging to the new patient) and 2) the iterative scheme that solves the image reconstruction problem.

2.2 Primal-Dual Three-Operator Splitting

The iterative scheme selected here for the LDCT reconstruction problem is the Primal-Dual Three-Operator (PD3O) splitting algorithm. The PD3O was introduced in [11] and it is a generalized version of the Primal-Dual Hybrid Gradient (PDHG) algorithm [12]. It is used to minimize objectives that consist of a proximable function g , a composite function f with the linear operator \mathbf{K} and a differentiable function h with a Lipschitz constant L :

$$\min_{\mathbf{x}} f(\mathbf{K}\mathbf{x}) + g(\mathbf{x}) + h(\mathbf{x}).$$

The algorithm is summarized in Algorithm 1 and explained next. Unlike the standard L^2 -squared fidelity

Algorithm 1 Unrolled PD3O algorithm

Input: $L = \text{Lip}(\nabla h)$, $\tau = 2/L$, $\sigma = 1/(\tau\|\mathbf{K}\mathbf{K}^\top\|)$,
initial guess $\bar{\mathbf{x}}_0$

Output: reconstructed image \mathbf{x}_{TV}

- 1: $\mathbf{p}_0 = \bar{\mathbf{x}}_0$
- 2: $\mathbf{q}_0 = \mathbf{0}$
- 3: **for** $k < T$ **do**
- 4: $\mathbf{q}_{k+1} = \text{prox}_{\sigma f^*}(\mathbf{q}_k + \sigma\mathbf{K}\bar{\mathbf{x}}_k)$
- 5: $\mathbf{p}_{k+1} = \text{prox}_{\tau g}(\mathbf{p}_k - \tau\nabla h(\mathbf{p}_k) - \tau\mathbf{K}^\top\mathbf{q}_{k+1})$
- 6: $\bar{\mathbf{x}}_{k+1} = 2\mathbf{p}_{k+1} - \mathbf{p}_k + \tau\nabla h(\mathbf{p}_k) - \tau\nabla h(\mathbf{p}_{k+1})$
- 7: **end for**
- 8: $\mathbf{x}_{\text{TV}} = \bar{\mathbf{x}}_T$

term that is commonly used in tomography reconstruction problems with Gaussian noise, here we employ the Kullback-Leibler divergence which is more suitable to describe the noise distribution of the measured tomographic data \mathbf{z} . We have that $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{e}$, where

$$\mathbf{e} = -\mathbf{A}\mathbf{x} - \log(\tilde{\mathbf{N}}_1/N_0), \quad \tilde{\mathbf{N}}_1 \sim \text{Pois}(N_0 \exp(-\mathbf{A}\mathbf{x}\boldsymbol{\mu})).$$

We denote with $\boldsymbol{\mu}$ and N_0 the normalization constant and the mean photon count per detector bin without attenuation, respectively. The data-discrepancy in (3) can be derived from a Bayesian viewpoint and is

$$\mathcal{D}(\mathbf{A}\mathbf{x}, \mathbf{z}) = \sum_{i=1}^m e^{-(\mathbf{A}\mathbf{x})_i\boldsymbol{\mu}} N_0 - e^{-\mathbf{z}_i\boldsymbol{\mu}} N_0 (- (\mathbf{A}\mathbf{x})_i\boldsymbol{\mu} + \log(N_0)), \quad (7)$$

see [13] for more details. To configure PD3O for (3) we define the following

$$f(\mathbf{q}) = \|\Lambda\mathbf{q}\|_1, \quad g(\mathbf{p}) = \mathbb{I}_{\{\mathbf{p}>0\}}(\mathbf{p}), \quad \mathbf{K} = \nabla, \\ h(\mathbf{p}) = \sum_{i=1}^m e^{-\mathbf{p}_i\boldsymbol{\mu}} N_0 - e^{-\mathbf{z}_i\boldsymbol{\mu}} N_0 (-\mathbf{p}_i\boldsymbol{\mu} + \log(N_0)).$$

Notice that with the standard L^2 -squared fidelity term, it is sufficient to use the PDHG algorithm since its convex conjugate has a closed-form proximal operator, which is not the case with (7). However, the additional function in the PD3O algorithm allows to express the data discrepancy in the differentiable term h . Note that ∇h is not globally Lipschitz continuous but due to the non-negativity constraint, we only have to consider $\nabla h(\mathbf{p})$ for \mathbf{p} with non-negative entries. Consequently, we can find an upper bound of the Lipschitz constant of ∇h by $\text{Lip}(\nabla h) \leq \|\mathbf{A}\|^2\boldsymbol{\mu}^2N_0$.

3 Results

To evaluate our proposed unrolled NN, we use the LoDoPaB dataset [14] for low-dose CT imaging. It is based on scans of the Lung Image Database Consortium and Image Database Resource Initiative which serve as

ground truth images, while the measurements are simulated. The dataset contains 35820 training images, 3522 validation images and 3553 test images. Here the ground truth images have a resolution of 362×362 on a domain of $26\text{cm} \times 26\text{cm}$. We only use the first 300 training images and the first 10 validation images. For the forward operator we consider a normalization constant $\boldsymbol{\mu} = 81.35858$, the mean photon count per detector bin $N_0 = 4096$ as well as 513 equidistant detector bins and 1000 equidistant angles between 0 and π .

In Figure 2 we compare the FBP reconstruction with the PD3O reconstructions where we use (i) a scalar parameter (λ), chosen to maximize the PSNR ‘‘on average’’, and (ii) our computed spatially dependent parameter map (Λ_Θ). Using the latter, we obtain a significant improvement in the reconstruction both visually and in terms of quality measures, e.g., PSNR and SSIM. In particular, sharp edges are retained, while the constant regularizing parameter results in blurry and blocky-like reconstructions. One can observe that the network attributes higher regularization parameters to image content with smooth structures while it yields lower regularization parameters at the edges to prevent smoothing.

4 Discussion

We have presented a data-driven approach to automatically estimate spatially dependent parameter-maps for TV regularization for the low dose X-ray CT tomography reconstruction problem. Although only the TV regularization is considered in this paper, higher order or combinations of regularizers can be used for different CT applications, see [15, 16]. Moreover, our unrolled framework is quite flexible and can be easily used for other modalities such as qualitative and quantitative MRI reconstruction, image denoising as well as their dynamic versions, see [17]. Finally, more sophisticated network architectures than the U-Net have been proposed recently, e.g. [18, 19], which could be potentially adopted for the estimation of the regularization parameter-maps as well.

5 Acknowledgments

The authors acknowledge the support of the German Research Foundation (DFG) under Germany’s Excellence Strategy – The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689) as this work was initiated during the Hackathon event ‘‘Maths Meets Image’’, Berlin, March 2022, which was part of the MATH+ Thematic Einstein Semester on ‘‘Mathematics of Imaging in Real-World Challenges’’. This work is further supported via the MATH+ project EF3-7.

This work was funded by the UK EPSRC grants the ‘‘Computational Collaborative Project in Synergistic Re-

construction for Biomedical Imaging" (CCP SyneRBI) EP/T026693/1; "A Reconstruction Toolkit for Multi-channel CT" EP/P02226X/1 and "Collaborative Computational Project in tomographic imaging" (CCPi) EP/M022498/1 and EP/T026677/1. This work made use of computational support by CoSeC, the Computational Science Centre for Research Communities, through CCP SyneRBI and CCPi.

This work is part of the Metrology for Artificial Intelligence for Medicine (M4AIM) project that is funded by the German Federal Ministry for Economic Affairs and Energy (BMWi) in the framework of the QI-Digital initiative.

References

- [1] E. Y. Sidky, J. H. Jørgensen, and X. Pan. "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm". *Physics in Medicine & Biology* 57.10 (2012). <https://doi.org/10.1088/2F0031-9155/2F57/2F10/2F3065>, p. 3065.
- [2] K. Kunisch and T. Pock. "A Bilevel Optimization Approach for Parameter Learning in Variational Models". *SIAM Journal on Imaging Sciences* 6.2 (2013). <http://dx.doi.org/10.1137/120882706>, pp. 938–983.
- [3] L. Calatroni, C. Chung, J. C. De Los Reyes, et al. "Bilevel approaches for learning of variational imaging models". *RADON book Series on Computational and Applied Mathematics, vol. 18*. <https://www.degruyter.com/view/product/458544>. Berlin, Boston: De Gruyter, 2017.
- [4] L. Chung, J. C. De los Reyes, and C. B. Schönlieb. "Learning optimal spatially-dependent regularization parameters in total variation image denoising". *Inverse Problems* 33 (2017). <https://doi.org/10.1088/1361-6420/33/7/074005>, p. 074005.
- [5] M. Hintermüller, C. N. Rautenberg, T. Wu, et al. "Optimal Selection of the Regularization Function in a Weighted Total Variation Model. Part II: Algorithm, Its Analysis and Numerical Tests". *Journal of Mathematical Imaging and Vision* 59.3 (2017). <https://doi.org/10.1007/s10851-017-0736-2>, pp. 515–533.
- [6] M. Hintermüller and K. Papafitsoros. "Generating structured nonsmooth priors and associated primal-dual methods". *Processing, Analyzing and Learning of Images, Shapes, and Forms: Part 2*. Ed. by R. Kimmel and X.-C. Tai. Vol. 20. Handbook of Numerical Analysis. <https://doi.org/10.1016/bs.hna.2019.08.001>. 2019, pp. 437–502.
- [7] M. Hintermüller, K. Papafitsoros, C. N. Rautenberg, et al. "Dualization and Automatic Distributed Parameter Selection of Total Generalized Variation via Bilevel Optimization". *Numerical Functional Analysis and Optimization* 43.8 (2022). <https://doi.org/10.1080/01630563.2022.2069812>, pp. 887–932.
- [8] V. Pagliari, K. Papafitsoros, B. Raita, et al. "Bilevel training schemes in imaging for total-variation-type functionals with convex integrands". *SIAM Journal on Imaging Sciences* 15.4 (2022). <https://doi.org/10.1137/21M1467328>, pp. 1690–1728.

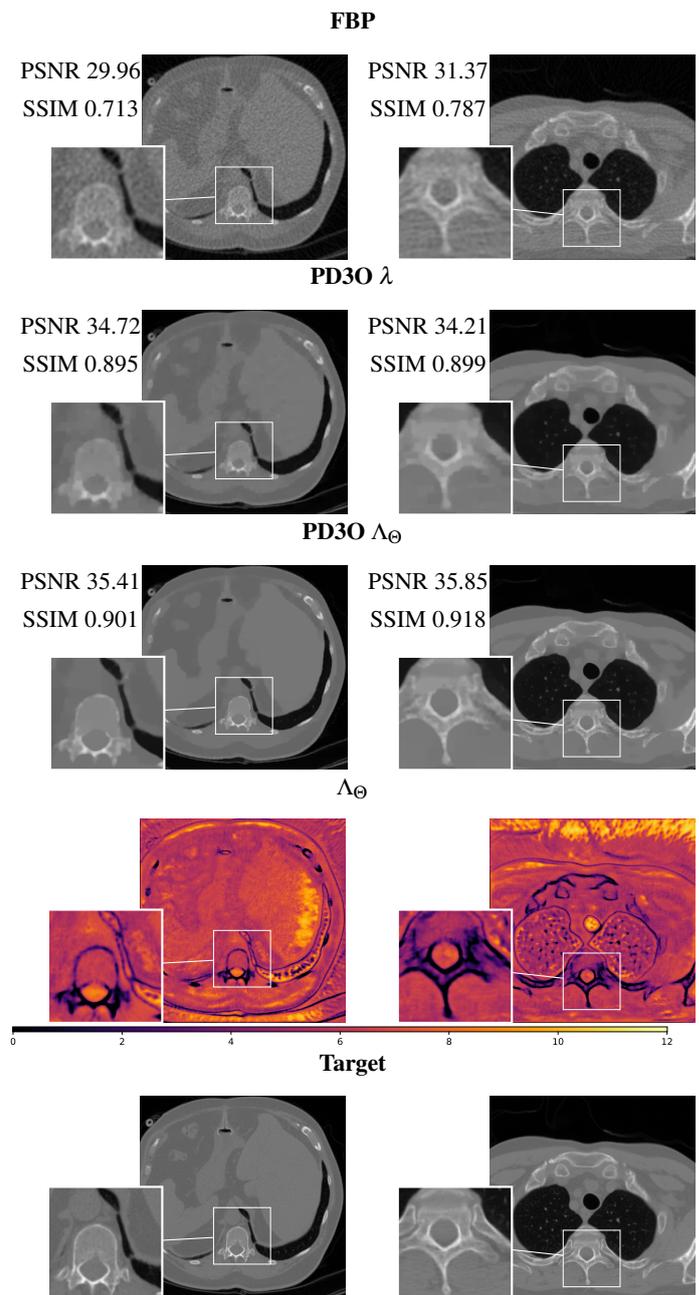


Figure 2: Different reconstructions obtained with PD3O employing a scalar regularization and the regularization parameter-maps obtained with the proposed CNN. From top to bottom: initial FBP-reconstruction, PD3O λ , PD3O Λ_{Θ} , spatial regularization parameter-map and ground truth image. The corresponding PSNR and SSIM values are given in the top left corner of the image.

- [9] V. Monga, Y. Li, and Y. C. Eldar. "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing". *IEEE Signal Processing Magazine* 38.2 (2021). <https://doi.org/10.1109/MSP.2020.3016905>, pp. 18–44. DOI: <https://doi.org/10.1109/MSP.2020.3016905>.
- [10] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation". *International Conference on Medical image computing and computer-assisted intervention*. https://doi.org/10.1007/2F978-3-662-54345-0_3. Springer, 2015, pp. 234–241.
- [11] M. Yan. "A New Primal–Dual Algorithm for Minimizing the Sum of Three Functions with a Linear Operator". *Journal*

- of Scientific Computing* 76 (2018). <https://doi.org/10.1007%2Fs10915-018-0680-3>, pp. 1698–1717.
- [12] A. Chambolle and T. Pock. “A first-order primal-dual algorithm for convex problems with applications to imaging”. *Journal of Mathematical Imaging and Vision* 40.1 (2011). <https://doi.org/10.1007%2Fs10851-010-0251-1>, pp. 120–145.
- [13] F. Altekürger, A. Denker, P. Hagemann, et al. “PatchNR: Learning from Small Data by Patch Normalizing Flow Regularization”. *arXiv preprint arXiv:2205.12021* (2022). <https://doi.org/10.48550/arXiv.2205.12021>.
- [14] J. Leuschner, M. Schmidt, D. O. Bager, et al. “LoDoPaB-CT, a benchmark dataset for low-dose computed tomography reconstruction”. *Scientific Data* 8.109 (2021). <https://doi.org/10.1038%2Fs41597-021-00893-z>.
- [15] E. Papoutsellis, E. Ametova, C. Delplancke, et al. “Core Imaging Library - Part II: multichannel reconstruction for dynamic and spectral tomography”. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379.2204 (2021). <https://doi.org/10.1098/rsta.2020.0193>, p. 20200193.
- [16] R. Warr, E. Ametova, R. J. Cernik, et al. “Enhanced hyperspectral tomography for bioimaging by spatio-spectral reconstruction”. *Scientific Reports* 11.1 (2021). <https://doi.org/10.1038/s41598-021-00146-4>.
- [17] A. Kofler, F. Altekürger, F. A. Ba, et al. “Learning Regularization Parameter-Maps for Variational Image Reconstruction using Deep Neural Networks and Algorithm Unrolling”. *arXiv preprint arXiv:2301.05888* (2023). <https://arxiv.org/abs/2301.05888>.
- [18] Z. Lu, J. Li, H. Liu, et al. “Transformer for Single Image Super-Resolution”. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. <https://doi.org/10.1109%2Fcvprw56347.2022.00061>. IEEE, 2022.
- [19] J. Liang, J. Cao, G. Sun, et al. “SwinIR: Image Restoration Using Swin Transformer”. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. <https://doi.org/10.1109%2Ficcvw54120.2021.00210>. IEEE, 2021.