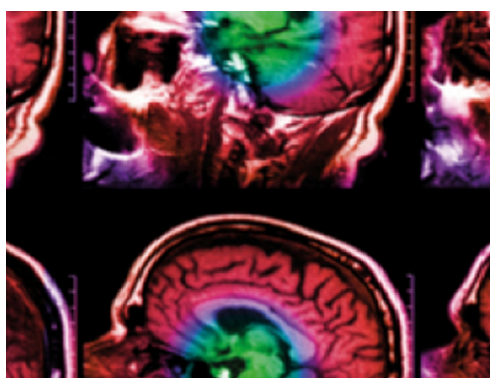


PAPER • OPEN ACCESS

Neural networks-based regularization for large-scale medical image reconstruction

To cite this article: A Kofler *et al* 2020 *Phys. Med. Biol.* **65** 135003

View the [article online](#) for updates and enhancements.



IPEM | IOP

Series in Physics and Engineering in Medicine and Biology

Your publishing choice in medical physics,
biomedical engineering and related subjects.

Start exploring the collection—download the
first chapter of every title for free.



PAPER








OPEN ACCESS

Neural networks-based regularization for large-scale medical image reconstruction

RECEIVED
9 April 2020REVISED
8 May 2020ACCEPTED FOR PUBLICATION
3 June 2020PUBLISHED
1 July 2020

Original Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

A Kofler¹ , M Haltmeier² , T Schaeffter^{3,4,5} , M Kachelrieß⁶ , M Dewey¹ , C Wald¹  and C Kolbitsch^{3,4} ¹ Department of Radiology, Charité - Universitätsmedizin Berlin, Berlin, Germany² Department of Mathematics, University of Innsbruck, Innsbruck, Austria³ Physikalisch-Technische Bundesanstalt, Braunschweig and Berlin, Germany⁴ Division of Imaging Sciences and Biomedical Engineering, King's College London, London United Kingdom⁵ Department of Medical Engineering, Technical University of Berlin, Berlin, Germany⁶ Division of X-ray Imaging and CT, German Cancer Research Center, Heidelberg, GermanyE-mail: andreas.kofler@charite.de

Keywords: deep learning, neural networks, inverse problems, low-dose CT, radial cine MRI

Abstract

In this paper we present a generalized Deep Learning-based approach for solving ill-posed large-scale inverse problems occurring in medical image reconstruction. Recently, Deep Learning methods using iterative neural networks (NNs) and cascaded NNs have been reported to achieve state-of-the-art results with respect to various quantitative quality measures as PSNR, NRMSE and SSIM across different imaging modalities. However, the fact that these approaches employ the application of the forward and adjoint operators repeatedly in the network architecture requires the network to process the whole images or volumes at once, which for some applications is computationally infeasible. In this work, we follow a different reconstruction strategy by strictly separating the application of the NN, the regularization of the solution and the consistency with the measured data. The regularization is given in the form of an image prior obtained by the output of a previously trained NN which is used in a Tikhonov regularization framework. By doing so, more complex and sophisticated network architectures can be used for the removal of the artefacts or noise than it is usually the case in iterative NNs. Due to the large scale of the considered problems and the resulting computational complexity of the employed networks, the priors are obtained by processing the images or volumes as patches or slices. We evaluated the method for the cases of 3D cone-beam low dose CT and undersampled 2D radial cine MRI and compared it to a total variation-minimization-based reconstruction algorithm as well as to a method with regularization based on learned overcomplete dictionaries. The proposed method outperformed all the reported methods with respect to all chosen quantitative measures and further accelerates the regularization step in the reconstruction by several orders of magnitude.

1. Introduction

In inverse problems, the goal is to recover an object of interest from a set of indirect and possibly incomplete observations. In medical imaging, for example, a classical inverse problem is given by the task of reconstructing a diagnostic image from a certain number of measurements, e.g. x-ray projections in computed tomography (CT) or the spatial frequency information (k -space data) in magnetic resonance imaging (MRI). The reconstruction from the measured data can be an ill-posed inverse problem for different reasons. In low-dose CT, for example, the reconstruction from noisy data is ill-posed because of the ill-posedness of the inversion of the Radon transform. In accelerated MRI, on the other hand, the reconstruction from incomplete data is ill-posed since the underlying problem is essentially underdetermined and therefore no unique solution exists without integrating prior information.

In order to constrain the space of possible solutions, a typical approach is to impose specific a-priori chosen properties on the solution by adding a regularization (or penalty) term to the problem. Well known

choices for the regularization are for example given by the popular total variation-minimization and sparse regularization approaches, where the solution is transformed using a sparsifying transform such as the Wavelet-transform or the Fourier-transform (Lustig *et al* 2008) or a finite-differences filter (Block *et al* 2007) and the L_1 -norm of the latter is minimized. While the aforementioned methods use hand-crafted priors, other methods learn the regularization directly within the reconstruction of the images where the regularization is imposed patch-wise by the sparse approximation using a dictionary which is learned in an unsupervised manner during the reconstruction (Wang and Ying 2014, Xu *et al* 2012). However, these learning-based methods are usually time consuming since the regularization is adaptive and learned during an iterative reconstruction scheme. Further, in the specific dictionary learning framework, the regularization requires training of a dictionary and sparse coding of all patches of the current image estimate at each iteration. This is computationally demanding and makes the application in the clinical routine challenging.

Recently, Convolutional Neural Networks (CNNs) have been applied in the field of inverse problems, either as direct full inversion methods (Zhu *et al* 2018), as post processing methods (Jin *et al* 2017, Schwab *et al* 2019, Han and Ye 2018), as learned iterative schemes (Adler and Öktem 2017, 2018, Gupta *et al* 2018, Chun and Fessler 2018, Chun *et al* 2019), or as learned regularizers (Schlemper *et al* 2018, Kofler *et al* 2018, Aggarwal *et al* 2018, Qin *et al* 2018, 2019). When used as post-processing methods, the networks are trained to denoise or remove artefacts from images obtained by the direct reconstruction of the noisy or incomplete data. Although a wide range of different network architecture has been proposed, e.g. Han and Ye (2018), Yang *et al* (2018), a major concern is that the estimated output of the CNN might lack data-consistency. In order to ensure that the obtained image is consistent with the acquired raw data, methods have been proposed where the constructed networks define unrolled iterative schemes which employ the forward and the adjoint operators. These methods can be interpreted as learned iterative schemes and have been successfully applied to different imaging modalities (Adler and Öktem 2017, 2018, Gupta *et al* 2018, Hammernik *et al* 2018, Hauptmann *et al* 2018, Schlemper *et al* 2018, Kofler *et al* 2018, Aggarwal *et al* 2018, Qin *et al* 2018). Thereby, the subnetworks containing trainable parameters can be thought of regularizers which are learned by end-to-end training of the whole network cascade. Due to the integration of the forward and the adjoint operators, iterative or cascaded networks seem to be a choice for various image reconstruction task. However, the main advantage of these methods at the same time represents the computational bottleneck of the approaches. The fact that the forward and the adjoint operators are integrated as layers in the networks requires that the whole object of interest has to be processed at once. Since CNNs typically increase the input size by extracting several feature maps per layer, end-to-end training might be infeasible for some high-dimensional problems, including high-resolution 3D CT volumes or non-Cartesian MR acquisitions. Further, approaches based on iterative reconstruction where different CNNs are applied at different iterates have been considered, see e.g. Hauptmann *et al* (2018), Chun and Fessler (2018), Chun *et al* (2019). They are similar to cascaded/iterative networks with the difference that each network has to be trained separately in a greedy fashion. Since training every network also implies the generation of a new dataset where the forward and the backward operators have to be used, training times increase substantially.

In order to overcome these limitations, we propose to decouple the regularization of the solution from ensuring consistency with the measured data. We present a general framework to use CNNs as learned regularizers and still ensure data-consistency of the obtained solution. In particular, we consider high-dimensional problems where either the object of interest or the measured data are high-dimensional (high-resolution 3D CT) or the evaluation of the forward or the adjoint operators is computationally expensive (dynamic 2D non-Cartesian radial MR acquisition). In Kofler *et al* (2018), the authors studied the performance of cascaded networks of different lengths, i.e. with different number of alternation of CNNs and data-consistency layers, but with an approximately equal overall number of trainable parameters. It was observed that the performance of the different networks was comparable for all studied lengths. Also, in Hyun *et al* (2018), the authors demonstrated that for Cartesian MR image reconstruction, a single k -space correction significantly improved the obtained image quality. There, the k -space correction corresponds to strict data-consistency, i.e. by a replacement of the k -space coefficients estimated by the CNN with the measured ones. A similar approach was proposed in Schwab *et al* (2019) for general inverse problems. Based on these observations, we propose a three-steps approach for large-scale medical image reconstruction, where the construction of cascaded or iterative networks is not possible due to memory constraints. Also, reducing the number of times ones needs to apply the forward and the adjoint operators is desirable in order to shorten the reconstruction times in the clinical routine. The first step of the proposed scheme involves the direct reconstruction from the measurements. The second step requires the processing of the initial reconstruction with a pre-trained CNN. The third and final step consists in minimizing a functional given as a data-fidelity term which is regularized by the previous output of the CNN but is otherwise independent of any network. The fact that instead of imposing strict data-consistency, a regularized functional is minimized,

makes the method general and applicable to problems which are ill-posed due to noisy measurements (e.g. low-dose CT) or ones where incomplete data is available (accelerated MRI) as well.

This paper is organized as follows. In section 2, we formally introduce the inverse problem of image reconstruction and motivate our proposed approach for the solution of large-scale ill-posed inverse problems. We demonstrate the feasibility of our method by applying it to 3D low-dose cone beam CT and 2D radial cine MRI in section 3. We further compare the proposed approach to an iterative reconstruction method given by total variation-minimization (TV) and a learning-based method (DIC) using Dictionary Learning-based priors in section 4. We then conclude the work with a discussion and conclusion in section 5 and section 6.

2. Large-scale image reconstruction with CNN-priors

In this section, we present the proposed deep learning scheme for solving large-scale, possibly non-linear, inverse problems. For the sake of clarity, we do not focus on a functional analytical setting but consider discretized problems of the form

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}, \quad (1)$$

where $\mathbf{A}: X \rightarrow Y$ is a discrete possibly non-linear forward operator between finite dimensional Hilbert spaces, $\mathbf{y} \in Y$ is the measured data, $\mathbf{z} \in Y$ the noise and $\mathbf{x} \in X$ the unknown object to be recovered. The operator \mathbf{A} could for example model the measurement process in various imaging modalities such as the x-ray projection in CT or the Fourier encoding in MRI. Depending on the nature of the underlying imaging modality one is considering, problem (1) can be ill-posed for different reasons. For example, in low-dose CT, the measurement data is inherently contaminated by noise. In cardiac MRI, k -space data is often undersampled in order to speed up the acquisition process. This leads to incomplete data and therefore to an undetermined problem with an infinite number of theoretically possible solutions.

In order to constrain the space of solutions of interest, a typical approach is to impose specific a-priori chosen properties on the solution \mathbf{x} by adding a regularization (or penalty) term $\mathcal{R}(\mathbf{x})$ and using Lagrange multipliers. Then, one solves the relaxed problem

$$D(\mathbf{A}\mathbf{x}, \mathbf{y}) + \lambda \mathcal{R}(\mathbf{x}) \rightarrow \min, \quad (2)$$

where $D(\cdot, \cdot)$ is an appropriately chosen data-discrepancy measure and $\lambda > 0$ controls the strength of the regularization. The choice of $D(\cdot, \cdot)$ depends on the considered problem. For the examples presented in Sections 3 and 4 we choose the discrepancy measure as the squared norm distance in the case of radial cine MRI and the Kullback-Leibler divergence in the case of low dose CT, respectively. A particular feature of proposed the large-scale approach is the one of a regularizer depending on the actual data, see (4) below.

2.1. CNN-based regularization

Clearly, the regularization term $\mathcal{R}(\mathbf{x})$ significantly affects the quality and the characteristics of the solution \mathbf{x} . Here, we propose a generalized approach for solving high-dimensional inverse problems by the following three steps: First, an initial guess of the solution is provided by a direct reconstruction from the measured data, i.e. $\mathbf{x}_{\text{ini}} = \mathbf{A}^\dagger \mathbf{y}$, where $\mathbf{A}^\dagger: Y \rightarrow X$ denotes some reconstruction operator. Then, a CNN is used to remove the noise or the artefacts from the direct reconstruction \mathbf{x}_{ini} in order to obtain another intermediate reconstruction \mathbf{x}_{CNN} which is used as a CNN-prior in a generalized Tikhonov functional

$$F_{\mathbf{y}, \mathbf{x}_{\text{CNN}}, \lambda}(\mathbf{x}) := D(\mathbf{A}\mathbf{x}, \mathbf{y}) + \lambda \|\mathbf{x} - \mathbf{x}_{\text{CNN}}\|_2^2 \rightarrow \min. \quad (3)$$

As a third and final step, the CNN-Tikhonov functional (3) is minimized resulting in the proposed CNN-based reconstruction.

Note that the regularization of the problem, i.e. obtaining the CNN-prior, is decoupled from the step of ensuring data-consistency of the solution via minimization of (3). This means that, in our case, the regularization term $\mathcal{R}(\mathbf{x})$ in (2) depends on the data \mathbf{y} and is of the form

$$\mathcal{R}(\mathbf{x}) = \mathcal{R}(\mathbf{x}, f_\Theta(\mathbf{A}^\dagger \mathbf{y})), \quad (4)$$

where f_Θ denotes a function which decomposes the initial reconstruction $\mathbf{A}^\dagger \mathbf{y}$ into patches, processes them with a pre-trained CNN u_Θ with trainable parameters Θ and recomposes the patches in order to provide the CNN image-prior \mathbf{x}_{CNN} . Therefore, for minimizing (3) we only have to apply f_Θ one single time. Thus, the training of the network u_Θ is facilitated since it only has to learn to map the manifold given by the initial reconstructions to the manifold of the ground truth images. This also allows to use deeper and more

sophisticated CNNs as the ones typically used in iterative networks. Given the high-dimensionality of the considered problems, network training is further carried out on sub-portions of the image samples, i.e. on patches or slices which are previously extracted from the images or volumes. This is motivated by the fact that in most medical imaging applications, one has typically access to datasets with only a relatively small number of subjects. The images or volumes of these subjects, on the other hand, are elements of a high-dimensional space. Therefore, one is concerned with the problem of having topologically sparse training data with only very few data points in the original high-dimensional image space. Working with sub-portions of the image samples increases the number of available data points and at the same time decreases its ambient dimensionality.

2.2. Large-scale CNN-prior

Suppose we have access to a finite set of N ground truth samples $(\mathbf{x}_k)_{k=1}^N$ and corresponding initial estimates $(\mathbf{x}_{\text{ini},k})_{k=1}^N$. We are in particular interested in the case where N is relatively small and the considered samples \mathbf{x}_k have a relatively large size, which is the case for most medical imaging applications. For any sample $\mathbf{x} \in X$ we consider its decomposition in $N_{\mathbf{p},s}$ possibly overlapping patches

$$\mathbf{x} = \mathbf{W}_{\mathbf{p},s} \sum_{j=1}^{N_{\mathbf{p},s}} (\mathbf{R}_j^{\mathbf{p},s})^T \mathbf{R}_j^{\mathbf{p},s} \mathbf{x}, \quad (5)$$

where $\mathbf{R}_j^{\mathbf{p},s}$ and $(\mathbf{R}_j^{\mathbf{p},s})^T$ extract and reposition the patches at the original position, respectively, and the diagonal operator $\mathbf{W}_{\mathbf{p},s}$ accounts for weighting of regions containing overlaps. The entries of the tuples \mathbf{p} and s specify the size of the patches and the strides in each dimension and therefore the number of patches $N_{\mathbf{p},s}$ which are extracted from a single image.

We aim for improved estimates $\mathbf{x}_{\text{CNN},k} = f_\theta(\mathbf{x}_{\text{ini},k}) \approx \mathbf{x}_k$ via a trained network function f_θ to be constructed. Since the operator norm of $\mathbf{W}_{\mathbf{p},s}$ is less or equal to one, by the triangle inequality, we can estimate the average error

$$e_N := \sum_{k=1}^N \|\mathbf{x}_k - \mathbf{x}_{\text{CNN},k}\|_2 \leq \sum_{k=1}^N \sum_{j=1}^{N_{\mathbf{p},s}} \|\mathbf{R}_j^{\mathbf{p},s} \mathbf{x}_k - \mathbf{R}_j^{\mathbf{p},s} \mathbf{x}_{\text{CNN},k}\|_2 =: e_{N,N_{\mathbf{p},s}}. \quad (6)$$

Inequality (6) suggests that it is beneficial estimating each patch of the sample \mathbf{x}_k by a neural network u_θ applied to $\mathbf{R}_j^{\mathbf{p},s} \mathbf{x}_{\text{ini},k}$ rather than estimating the whole sample at once. The neural network u_θ is trained on a subset of pairs

$$\mathcal{D} = \left\{ \left(\mathbf{R}_j^{\mathbf{p},s}(\mathbf{x}_{\text{ini},k}), \mathbf{R}_j^{\mathbf{p},s}(\mathbf{x}_k) \right) : (k,j) \in \mathcal{I}_{N,N_{\mathbf{p},s}} \right\}, \quad (7)$$

of all possible patches extracted from the N samples in the dataset, where $\mathcal{I}_{N,N_{\mathbf{p},s}} := \{1, \dots, N\} \times \{1, \dots, N_{\mathbf{p},s}\}$. During training, we optimize the set of parameters θ to minimize the L_2 -error between the estimated output of the patches and the corresponding ground truth patch by minimizing

$$\mathcal{L}(\theta) = \frac{1}{N_{\text{train}}} \sum_{(\mathbf{z}_{\text{ini}}, \mathbf{z}) \in \mathcal{D}} \|u_\theta(\mathbf{z}_{\text{ini}}) - \mathbf{z}\|_2^2, \quad (8)$$

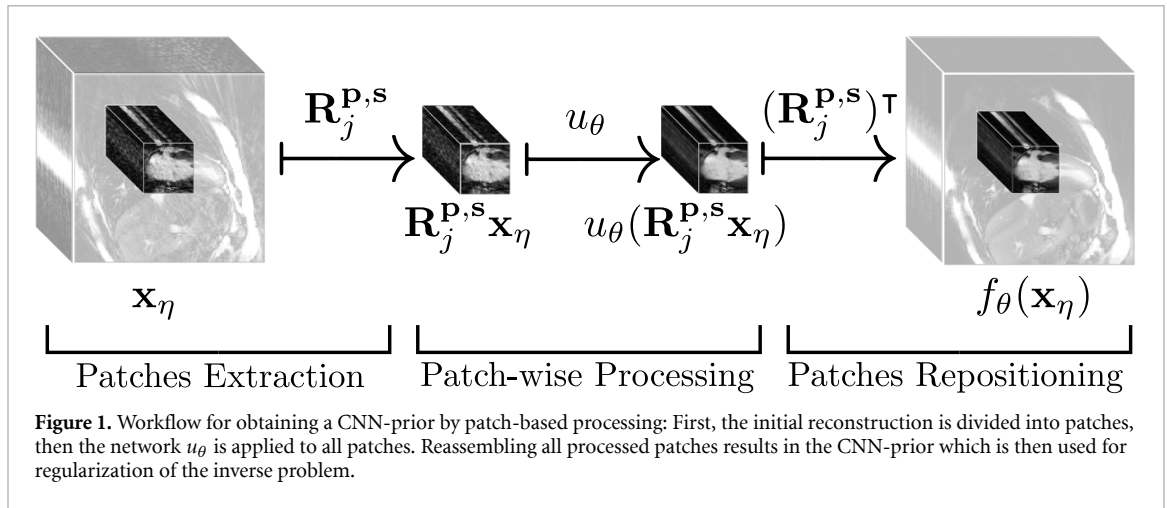
where N_{train} is the number of training patches.

Denote by f_θ the composite function which decomposes a sample image or volume \mathbf{x} into patches, applies a neural network u_θ to each patch, and reassembles the sample from them. This results in the proposed CNN-prior \mathbf{x}_{CNN} given by

$$\mathbf{x}_{\text{CNN}} := f_\theta(\mathbf{x}_{\text{ini}}) = \mathbf{W}_{\mathbf{p},s} \sum_j (\mathbf{R}_j^{\mathbf{p},s})^T (u_\theta(\mathbf{R}_j^{\mathbf{p},s}(\mathbf{x}_{\text{ini}}))), \quad (9)$$

where $\mathbf{x}_{\text{ini}} = \mathbf{A}^\dagger \mathbf{y}$ is the initial reconstruction obtained from the measured data.

Remark 1. Since in (6), the right-hand-side is an upper bound of the left-hand-side, inequality (6) guarantees that the set of parameters found by minimizing (8) is also suitable for obtaining the prior \mathbf{x}_{CNN} . Therefore, u_θ is powerful enough to deliver a CNN-prior to regularize the solution of (3). Figure 1 illustrates the process of extracting patches from a volume using the operator $\mathbf{R}_j^{\mathbf{p},s}$, processing it with a neural network u_θ and repositioning it at the original position using the transposed operator $(\mathbf{R}_j^{\mathbf{p},s})^T$. The example is shown for a 2D cine MR image sequence.



2.3. Reconstruction algorithm

After having found the CNN prior (9), as a final reconstruction step, the optimality condition for minimization problem (3) is solved with an iterative method dependent on the specific application. Minimizing this functional increases data-consistency compared to \mathbf{x}_{CNN} since the discrepancy to the measured data is used again. The solution of (3) is then the final CNN-based reconstruction of the proposed method. Algorithm 1 summarizes the proposed three-step reconstruction scheme.

Algorithm 1: Proposed CNNs-based large-scale image reconstruction algorithm.

Data: trained network u_θ , function f_θ , noisy or incomplete measured data \mathbf{y} , regularization parameter $\lambda > 0$

Output: reconstruction \mathbf{x}_{REC}

- 1) $\mathbf{x}_{\text{ini}} \leftarrow \mathbf{A}^\dagger \mathbf{y}$
 - 2) $\mathbf{x}_{\text{CNN}} \leftarrow f_\theta(\mathbf{x}_{\text{ini}})$
 - 3) $\mathbf{x}_{\text{REC}} \leftarrow \arg \min_{\mathbf{x}} D(\mathbf{A}\mathbf{x}, \mathbf{y}) + \lambda \|\mathbf{x} - \mathbf{x}_{\text{CNN}}\|_2^2$
- Return \mathbf{x}_{REC}
-

Note that the regularizer $\mathcal{R}(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_{\text{CNN}}\|_2^2$ is strongly convex. Therefore, if the discrepancy term $D(\mathbf{A}\mathbf{x}, \mathbf{y})$ is convex, then the Tikhonov functional (3) is strongly convex and can be efficiently minimized by most gradient based iterative schemes including Landweber's iteration and Conjugate Gradient type methods. The specific strategy for minimizing (3) depends on the considered application. In the case of an ill-conditioned inverse problem with noisy measurements, it might be beneficial that (3) is only approximately minimized. For example, for the case of low-dose CT, early stopping of the Landweber iteration is applied as additional regularization method due to the semi-convergence property of the Landweber iteration (Strand 1974). Such a strategy is used for the numerical results presented in section 4.

2.4. Convergence analysis

Another benefit of our approach is that minimization of the Tikhonov functional (3) corresponds to convex variational regularization with a quadratic regularizer. Therefore, one can use well known stability, convergence and convergence rates results (Engl et al 1996, Scherzer et al 2009, Grasmair 2010, Li et al 2020). Consequently, opposed to most existing neural network based reconstruction algorithms, the proposed framework is build on the solid theoretical fundament for regularizing inverse problems. As an example of such results we have the following theorem.

Theorem 1. Let $\mathbf{A}: X \rightarrow Y$ be linear, $\mathbf{x}_{\text{CNN}} \in X$, $\mathbf{y}_0 \in \mathbf{A}(X)$, and $\mathbf{y}_\delta \in Y$ satisfy $\|\mathbf{y}_\delta - \mathbf{y}_0\| \leq \delta$. Then the following hold:

1. For all $\delta, \lambda > 0$, the quadratic Tikhonov functional

$$F_{\mathbf{y}_\delta, \mathbf{x}_{\text{CNN}}, \lambda}(\mathbf{x}) := \|\mathbf{A}\mathbf{x} - \mathbf{y}_\delta\|_2^2 + \lambda \|\mathbf{x} - \mathbf{x}_{\text{CNN}}\|_2^2 \quad (10)$$

has a unique minimizer $\mathbf{x}_{\delta, \lambda}$.

2. The exact data equation $\mathbf{A}\mathbf{x} = \mathbf{y}_0$ has a unique \mathbf{x}_{CNN} -minimizing solution $\mathbf{x}_0 \in \arg \min\{\|\mathbf{x} - \mathbf{x}_{\text{CNN}}\|_2^2 : \mathbf{A}\mathbf{x} = \mathbf{y}_0\}$.
3. If the parameter choice $\lambda = \lambda(\delta)$ satisfies $\lambda, \delta^2/\lambda \rightarrow 0$ as $\delta \rightarrow 0$, then $\lim_{\delta \rightarrow 0} \|\mathbf{x}_0 - \mathbf{x}_{\delta, \lambda}\| = 0$.

Proof. The change of variables

- $\bar{\mathbf{x}} := \mathbf{x} - \mathbf{x}_{\text{CNN}}$
- $\bar{\mathbf{y}}_0 := \mathbf{y}_0 - \mathbf{A}\mathbf{x}_{\text{CNN}}$
- $\bar{\mathbf{y}}_\delta := \mathbf{y}_\delta - \mathbf{A}\mathbf{x}_{\text{CNN}}$

reduces (10) to standard Tikhonov regularization $\|\mathbf{A}\bar{\mathbf{x}} - \bar{\mathbf{y}}_\delta\|^2 + \lambda\|\bar{\mathbf{x}}\|_2^2 \rightarrow \min_{\bar{\mathbf{x}}}$ for the inverse problem $\mathbf{A}\bar{\mathbf{x}} = \bar{\mathbf{y}}_0$. Therefore, Items (1.) – (3.) follow from standard results that can be found for example in (Engl *et al* 1996, section 5). □

Theorem 1 also holds in the infinite-dimensional setting (Engl *et al* 1996, section 5) reflecting the stability of the proposed CNN regularization. Similar results hold for non-linear problems and general discrepancy measures (Grasmair 2010). Moreover, one can derive quantitative error estimates similar to Scherzer *et al* (2009), Grasmair (2010), Li *et al* (2020). Such theoretical investigations, however, are beyond the scope of this paper.

3. Experiments

In the following, we evaluated our proposed method on two different examples of large-scale inverse problems given by 3D low-dose CT and 2D undersampled radial cine MRI. We compared our proposed method to the well-known TV-minimization-based and dictionary learning-based approaches presented in Block *et al* (2007), Wang and Ying (2014) and Tian *et al* (2011), Xu *et al* (2012), which we abbreviate by TV and DIC, respectively. Further details about the comparison methods are discussed later in the paper.

3.1. 2D radial cine MRI

Here we applied our method to image reconstruction in undersampled 2D radial cine MRI. Typically, MRI is performed using multiple receiver coils and therefore, the inverse problem is given by

$$\mathbf{E}_I \mathbf{x} = \mathbf{y}_I, \quad (11)$$

where $\mathbf{x} \in \mathbb{C}^N$ with $N = N_x \cdot N_y \cdot N_t$ is an unknown complex-valued image sequence. The encoding operator \mathbf{E}_I is given by $\mathbf{E}_I = \mathbf{S} \circ \mathbf{E} \circ \mathbf{C}$ where

$$\mathbf{C} = [\mathbf{C}_1, \dots, \mathbf{C}_{n_c}]^T, \quad (12)$$

$$\mathbf{E} = \text{diag}(\mathbf{F}, \dots, \mathbf{F}), \quad (13)$$

$$\mathbf{S} = \text{diag}(\mathbf{S}_I, \dots, \mathbf{S}_I). \quad (14)$$

Here, \mathbf{C}_i denotes the i th coil sensitivity map, n_c is the number of coil-sensitivity maps, \mathbf{F} the 2D frame-wise operator and \mathbf{S}_I with $I \subset J = \{1, \dots, N_{\text{rad}}\}$, $|I| := m \leq N_{\text{rad}}$, a binary mask which models the undersampling process of the N_{rad} Fourier coefficients sampled on a radial grid. The vector $\mathbf{y}_I \in \mathbb{C}^M$ with $M = m \cdot n_c$ corresponds to the measured data. Here, we sampled the k -space data along radial trajectories chosen according to the golden-angle method (Winkelmann *et al* 2006). Note that problem (11) is mainly ill-posed not due to the presence of noise in the acquisition, but because the data acquisition is accelerated and hence only a fraction of the required measurements is acquired.

If we assume a radial data-acquisition grid, problem (11) is a large-scale inverse problem mainly because of two reasons. First, the measurement vector \mathbf{y}_I corresponds to n_c copies of the Fourier encoded image data multiplied by the corresponding coil sensitivity map. Second, the adjoint operator \mathbf{E}_I^H consists of two computationally demanding steps. The radially acquired k -space data is first properly re-weighted and interpolated to a Cartesian grid, for example by using Kaiser-Bessel functions (Rasche *et al* 1999). Then, a 2D inverse Fourier operation is applied to the image of each cardiac phase and the final image sequence is obtained by weighting the images from each estimated coil-sensitivity map and combining them to a single image sequence. We refer to the reconstruction obtained by $\mathbf{x}_I = \mathbf{E}_I^H \mathbf{y}_I$ as the non-uniform fast Fourier-transform (NUFFT) reconstruction. Therefore, in radial multi-coil MRI, the measured k -space data is high-dimensional and the application of the encoding operators \mathbf{E}_I and \mathbf{E}_I^H is further more computationally demanding than sampling on a Cartesian grid, see e.g Smith *et al* (2019). This makes the construction of cascaded networks which also process the k -space data (Han *et al* 2019) or repeatedly employ

the forward and adjoint operators (Schlemper *et al* 2018, Qin *et al* 2018) computationally challenging. Therefore, the separation of the regularization given by the CNNs from the data-consistency step is necessary in this case due to computational reasons.

As proposed in section 2, we solve a regularized version of problem (11) by minimizing

$$F_{\mathbf{y}_I, \mathbf{x}_{\text{CNN}}, \lambda}(\mathbf{x}) = \|\mathbf{E}_I \mathbf{x} - \mathbf{y}_I\|_2^2 + \lambda \|\mathbf{x} - \mathbf{x}_{\text{CNN}}\|_2^2, \quad (15)$$

where \mathbf{x}_{CNN} is obtained a-priori by using an already trained network. For this example, for obtaining the CNN-prior \mathbf{x}_{CNN} , we adopted the XT,YT approach presented in Kofler *et al* (2019) which has been reported to achieve comparable results to the 3D U-net (Hauptmann *et al* 2019). Using the XT,YT approach has the main advantage of only using 2D convolutional layers while still exploiting the spatio-temporal information of the cine MR images. The XT,YT approach works by first extracting all x, t - and y, t -slices from the image, subsequently processing them with a modified version of the 2D U-net and then properly reassembling them to obtain an estimate of the artefact-free image. Since the network only has to learn the manifold of the x, t - and y, t -spatio-temporal slices, the approach was shown to be applicable even when only little training data is available. Since the XT,YT method was previously introduced to only process real-valued data (i.e. the magnitude images), we followed a similar strategy by processing the real and imaginary parts of the image sequences separately but using the same real-valued network u_θ . This further increases the amount of training data by a factor of two. More precisely, let \mathbf{R}_j^{xt} and \mathbf{R}_j^{yt} denote the operators which extract the j th two-dimensional spatio-temporal slices in xt - and yt -direction from a 3D volume $(\mathbf{R}_j^{xt})^\top$ and $(\mathbf{R}_j^{yt})^\top$ their respective transposed operations which reposition the spatio-temporal slices at their original position.

By u_θ we denote a 2D U-net as the one described in Kofler *et al* (2019) which is trained on spatio-temporal slices, i.e. on a dataset of pairs which consist of the spatio-temporal slices in xt - and yt -direction of both the real and imaginary parts of the complex-valued images. The network u_θ was trained to minimize the L_2 -error between the ground truth image and the estimated output of the CNN. Our dataset consists of radially acquired 2D cine MR images from $n = 19$ subjects (15 healthy volunteers and 4 patients with known cardiac dysfunction) with 30 images covering the cardiac cycle. The ground truth images were obtained by kt -SENSE reconstruction using $N_\theta = 3400$ radial lines. We retrospectively generated the radial k -space data \mathbf{y}_I by sampling the k -space data along $N_\theta = 1130$ radial spokes using $n_c = 12$ coils. Note that sampling $N_\theta = 3400$ already corresponds to an acceleration factor of approximately ~ 3 and therefore, $N_\theta = 1130$ corresponds to an accelerated data-acquisition by an approximate factor of ~ 9 . The forward and the adjoint operators \mathbf{E}_I and \mathbf{E}_I^H were implemented using the ODL library (Adler *et al* 2017). The complex-valued CNN-regularized image sequence \mathbf{x}_{CNN} was obtained by

$$\begin{aligned} \mathbf{x}_{\text{CNN}} &= f_\theta(\mathbf{x}_I) \\ &= \frac{1}{2} \left[\sum_j (\mathbf{R}_j^{xt})^\top (u_\theta(\mathbf{R}_j^{xt}(\text{Re } \mathbf{x}_I))) + (\mathbf{R}_j^{yt})^\top (u_\theta(\mathbf{R}_j^{yt}(\text{Re } \mathbf{x}_I))) \right. \\ &\quad \left. + i \left((\mathbf{R}_j^{xt})^\top (u_\theta(\mathbf{R}_j^{xt}(\text{Im } \mathbf{x}_I))) \right) + i \left((\mathbf{R}_j^{yt})^\top (u_\theta(\mathbf{R}_j^{yt}(\text{Im } \mathbf{x}_I))) \right) \right]. \end{aligned}$$

Given \mathbf{x}_{CNN} , functional (15) was minimized by setting its derivative with respect to \mathbf{x} to zero and applying the pre-conditioned conjugate gradient (PCG) method to iteratively solve the resulting system. PCG was used to solve the system $\mathbf{H}\mathbf{x} = \mathbf{b}$ with

$$\begin{aligned} \mathbf{H} &= \mathbf{E}_I^H \mathbf{E}_I + \lambda \mathbf{I}, \\ \mathbf{b} &= \mathbf{x}_I + \lambda \mathbf{x}_{\text{CNN}}. \end{aligned} \quad (17)$$

Since the XT,YT method gives access to a large number of training samples, training the network u_θ for 12 epochs was sufficient. The CNN was trained by minimizing the L_2 -norm of the error between labels and output by using the Adam optimizer (Kingma and Ba 2014). We split our dataset in 12/3/4 subjects for training, validation and testing and performed a 4-fold cross-validation. For the experiment, we performed $n_{\text{iter}} = 16$ subsequent iterations of PCG and empirically set $\lambda = 0.1$. Note that due to strong convexity, (15) has a unique minimizer and solving system (17) yields the desired minimizer. The obtained results can be found in Subsection 4.1.

3.2. 3D low-dose computed tomography

The current generation of CT scanners performs the data-acquisition by emitting x-rays along trajectories in the form of a cone-beam for each angular position of the scanner. Therefore, for each angle ϕ of the rotation, one obtains an x-ray image which is measured by the detector array and thus, the complete sinogram data

can be identified with a 3D array of shape $(N_\phi, N_{r_x}, N_{r_y})$. Thereby, N_ϕ corresponds to the number of angles the rotation of the scanner is discretized by and N_{r_x} and N_{r_y} denote the number of elements of the detector array. The values of these parameters vary from scanner to scanner but are in the order of $N_\phi \approx 1000$ for a full rotation of the scanner and $N_{r_x} \times N_{r_y} \approx 320 \times 800$ for a 320-row detector array, which is for example used for cardiac CT scans (Dewey *et al* 2009). The volumes obtained from the reconstructions are typically given by an in-plane number of pixels of $N_x \times N_y = 512 \times 512$ and varying number of slices N_z , dependent on the specific application. For this example, we consider a similar set-up as in Adler and Öktem (2017). The non-linear problem is given by

$$\mathbf{y}_\eta = \mathbf{T}\mathbf{x} + \boldsymbol{\eta} = p \exp\{-\mu \mathbf{R}\mathbf{x}\} + \boldsymbol{\eta}, \quad (18)$$

where p denotes the average number of photons per pixel, μ is the linear attenuation coefficient of water, \mathbf{R} corresponds to the discretized version of a ray-transform with cone-beam geometry and the vector $\boldsymbol{\eta}$ denotes the Poisson-distributed noise in the measurements. Following our approach, we are interested in solving

$$F_{\mathbf{y}_\eta, \mathbf{x}_{\text{CNN}}, \lambda}(\mathbf{x}) = D_{\text{KL}}(\mathbf{T}\mathbf{x}, \mathbf{y}_\eta) + \lambda \|\mathbf{x} - \mathbf{x}_{\text{CNN}}\|_2^2 \rightarrow \min, \quad (19)$$

where D_{KL} denotes the Kullback-Leibler divergence which corresponds to the log-likelihood function for Poisson-distributed noise. According to the previously introduced notation, the prior \mathbf{x}_{CNN} is given by $\mathbf{x}_{\text{CNN}} = f_\theta(\mathbf{x}_\eta)$, where f_θ denotes a CNN-based processing method with trainable parameters θ and $\mathbf{x}_\eta = \mathbf{R}^\dagger(-\mu^{-1} \ln(p^{-1} \mathbf{y}_\eta))$ with \mathbf{R}^\dagger being the filtered back-projection (FBP) reconstruction.

Since our object of interest \mathbf{x} is a volume, it is intuitive to choose a NN which involves 3D convolutions in order to learn the filters by exploiting the spatial correlation of adjacent voxels in x -, y - and z -direction. In this particular case, u_θ denotes a 3D U-net similar to the one presented in Hauptmann *et al* (2019). Due to the large dimensionality of the volumes \mathbf{x} , the network u_θ cannot be applied to the whole volume. Instead, following our approach, the volume was divided into patches to which the network u_θ is applied. Therefore, the output \mathbf{x}_{CNN} was obtained as described in (9), where u_θ operates on 3D patches given by the vector $\mathbf{p} = (128, 128, 16)$, which denotes the maximal size of 3D patches which we were able to process by a 3D U-net. The strides used for the extraction and the reassembling of the volumes used in (9) is empirically chosen to be $\mathbf{s} = (16, 16, 8)$.

Training of the network u_θ was performed on a dataset of pairs according to (7), where we retrospectively generated the measurements \mathbf{y}_η by simulating a low-dose scan on the ground truth volumes. For the experiment, we used 16 CT volumes from the randomized DISCHARGE trial (Napp *et al* 2017) which we cropped to a fixed size of $512 \times 512 \times 128$. The simulation of the low-dose scan was performed as described in Adler and Öktem (2017) by setting $p = 10\,000$ and $\mu = 0.02$. The operator \mathbf{R} is assumed to perform $N_\phi = 1000$ projections which are measured by a detector array of shape $N_{r_x} \times N_{r_y} = 320 \times 800$. For the implementation of the operators, we used the ODL library (Adler *et al* 2017). The source-to-axis and source-to-detector distances were chosen according to the DICOM files. Since the dataset is relatively small, we performed a 7-fold cross-validation where for each fold we split the dataset in 12 patients for training, 2 for validation and 2 for testing. The number of training samples N_{train} results from the number of patches times the number of volumes contained in the training set. We trained the network u_θ for 115 epochs by minimizing the L_2 -norm of the error between labels and outputs. For training, we used the Adam optimizer (Kingma and Ba 2014). With the described configuration of \mathbf{p} and \mathbf{s} , the resulting number of patches to be processed in order to obtain the prior \mathbf{x}_{CNN} is therefore given by $N_{\mathbf{p}, \mathbf{s}} = 9375$. In this example, the solution \mathbf{x}_{REC} to problem (19) was then obtained by performing $n_{\text{iter}} = 4$ iterations of Landweber's method where we further used the filtered-back projection \mathbf{R}^\dagger as a left-preconditioner to accelerate the convergence of the scheme. For the derivation of the gradient of (19) with respect to \mathbf{x} , we refer to Adler and Öktem (2017). The regularization parameter was empirically set to $\lambda = 1$. The results can be found in subsection 4.2.

3.3. Reference methods

Here we discuss the methods of comparison in more detail and report the times needed to process and reconstruct the images or volumes. The data-discrepancy term $D(\cdot, \cdot)$ was again chosen according to the considered examples as previously discussed. The TV-minimization approach used for comparison is given by solving

$$D(\mathbf{A}\mathbf{x}, \mathbf{y}) + \lambda \|\mathbf{G}\mathbf{x}\|_1 \rightarrow \min_{\mathbf{x}} \quad (20)$$

where \mathbf{G} denotes the discretized version of the isotropic first order finite differences filter in all three dimensions, i.e. in x -, y - and t -direction for the MR example and in x -, y - and z -direction for the CT example. The solution of problem (20) was obtained by introducing an auxiliary variable \mathbf{z} and alternating

between solving for \mathbf{x} and \mathbf{z} . For the solution of one of the sub-problems, an iterative shrinkage method was used, see Chambolle (2005) for more details. The second resulting sub-problem was solved by iteratively solving a system of linear equations, either by Landweber for the CT example or by PCG for the MRI example, as mentioned before.

The dictionary learning-based method used for comparison is given by the solution of the problem

$$D(\mathbf{Ax}, \mathbf{y}) + \lambda \|\mathbf{x} - \mathbf{x}_{\text{DIC}}\|_2^2 \rightarrow \min_{\mathbf{x}} \quad (21)$$

where, in contrast to our proposed method, \mathbf{x}_{DIC} was obtained by the patch-wise sparse approximation of the initial image estimate using an already trained dictionary \mathbf{D} . Therefore, using a similar notation as in (9), the prior \mathbf{x}_{DIC} is given by

$$\mathbf{x}_{\text{DIC}} = \mathbf{W}_{\mathbf{p},\mathbf{s}} \sum_j (\mathbf{R}_j^{\mathbf{p},\mathbf{s}})^\top \mathbf{D} \gamma_j, \quad (22)$$

where the dictionary \mathbf{D} was previously trained by 15 iterations of the iterative thresholding and K residual means algorithm (ITKRM) (Schnass 2018) on a set of ground truth images which were given by the high-dose images for the CT example and the kt -SENSE reconstructions from $N_\theta = 3400$ radial lines for the MRI example. Here again, the dictionary \mathbf{D} operates on three-dimensional patches, for the CT example/ x, y, z -patches and the MR example (x, y, t) -patches. Note that for each fold, for training the dictionary \mathbf{D} , we only used the data which we included in the training set for our method. This means we trained a total of seven dictionaries for the CT example and four dictionaries for the MRI example. For each iteration of ITKRM, we randomly selected a subject to extract 10 000 3D training patches. The corresponding sparse codes γ_j were then obtained by solving

$$\min_{\{\gamma_j\}} \sum_j \|(\mathbf{R}_j^{\mathbf{p},\mathbf{s}}) \mathbf{x}_{\text{ini}} - \mathbf{D} \gamma_j\|_2^2 + \|\gamma_j\|_0, \quad (23)$$

which is a sparse coding problem and was solved using orthogonal matching pursuit (OMP) (Tropp and Gilbert 2007). Thereby, the image \mathbf{x}_{ini} corresponds to either the FBP-reconstruction \mathbf{x}_f for the CT example or to the NUFFT-reconstruction \mathbf{x}_f for the MRI example. In both cases, we used patches of shape given by $\mathbf{p} = (4, 4, 4)$ and strides given by $\mathbf{s} = (2, 2, 2)$. The number of atoms K and the sparsity levels were set to $K = 4 \cdot d$, with $d = 4 \cdot 4 \cdot 4$ and $S = 16$. Note that, in contrast to Xu *et al* (2012) and Wang *et al* (2004), Caballero *et al* (2014), the dictionary and the sparse codes were not learned during the reconstruction, as the sparse coding step of all patches would be too time consuming for very large-scale inverse problems, such as the CT example. Instead, the dictionary and the sparse codes were used to generate the prior \mathbf{x}_{DIC} which makes the method also more similar and comparable to ours. The parameter λ is set as previously stated in the manuscript, depending on the considered example.

3.4. Quantitative measures

For the evaluation of the reconstructions we report the normalized root mean squared error (NRMSE) and the peak signal-to-noise ratio (PSNR) as error-based measures and the structural similarity index measure (SSIM) (Wang *et al* 2004) and the Haar Wavelet-based perceptual similarity index measure (HPSI) (Reisenhofer *et al* 2018) as image-similarity-based measures. The reported statistics were obtained by calculating the measures of the images in the xy -plane and averaging them over the different folds.

4. Results

4.1. Results for 2D radial cine MRI

Figure 2 shows an example of the results obtained with our proposed method. Figure 2(A) shows the initial NUFFT-reconstruction \mathbf{x}_f obtained from the undersampled k -space data \mathbf{y}_f . The CNN-prior \mathbf{x}_{CNN} obtained by the XT,YT network can be seen in figure 2(B) and (shows) a strong reduction of undersampling artefacts but also blurring of small structures as indicated the yellow arrows. The CNN-prior \mathbf{x}_{CNN} is then used as a prior in functional (15) which is subsequently minimized in order to obtain the solution \mathbf{x}_{REC} which can be seen in figure 2(C). Figure 2(D) shows the kt -SENSE reconstruction from the complete sampling pattern using $N_\theta = 3400$ radial spokes for the acquisition. From the point-wise error images, we clearly see that the NRMSE is further reduced after performing the further iterations to minimize the CNN-prior-regularized functional. Further, fine details are recovered as can be seen from the yellow arrows in figure 2(C).

Figure 3 shows a comparison of all different reported methods. As can be seen from the point-wise error in figure 3(B), the TV-minimization (Block *et al* 2007) method was able to eliminate some artefacts but less

Table 1. Quantitative measures for the 2D radial cine MRI example. The measures are obtained as averages over the four different folds.

| | NUFFT | \mathbf{x}_{CNN} | \mathbf{x}_{REC} | TV | DIC |
|-------|----------|---------------------------|---------------------------|----------|----------|
| PSNR | 36.802 3 | 42.564 7 | 48.775 2 | 41.696 8 | 45.474 3 |
| NRMSE | 0.122 8 | 0.061 2 | 0.030 2 | 0.069 3 | 0.044 2 |
| SSIM | 0.664 9 | 0.787 6 | 0.952 | 0.863 5 | 0.917 5 |
| HPSI | 0.967 9 | 0.991 0 | 0.998 5 | 0.987 8 | 0.995 9 |

Table 2. Quantitative measures for the 3D low-dose CT example. The measures are obtained as averages over the seven different folds.

| | FBP | \mathbf{x}_{CNN} | \mathbf{x}_{REC} | TV | DIC |
|-------|----------|---------------------------|---------------------------|---------|----------|
| PSNR | 30.005 2 | 40.354 6 | 39.626 4 | 33.946 | 34.780 7 |
| NRMSE | 0.165 7 | 0.049 8 | 0.053 8 | 0.105 1 | 0.093 8 |
| SSIM | 0.425 | 0.575 5 | 0.581 3 | 0.498 5 | 0.546 5 |
| HPSI | 0.939 4 | 0.982 1 | 0.981 9 | 0.950 3 | 0.958 1 |

accurately compared to both learning-based methods, see figure 3(C) and (figure) 3(D). Table 1 lists the obtained quantitative measures for all methods averaged over the 4 different folds. From table 1, we see that the DIC method yielded better results than TV with respect to all reported measures. Our proposed solution \mathbf{x}_{REC} further surpassed the dictionary learning-based method, by additionally increasing the PSNR and SSIM by approximately 3 dB and 0.04, respectively. The difference with respect to HPSI, on the other hand, is relatively small. Our method also reduced the NRMSE by about 0.014 compared to the DIC method. In addition, from table 1, we see that for this example, even though processing the initial NUFFT-reconstruction with a CNN improved image quality with respect to all reported measures, further iterations to minimize the CNN-prior regularized functional increased data-consistency and additionally improved the PSNR, SSIM, HPSI and NRMSE. In fact, the statistics of the CNN-prior show that only post-processing the initial NUFFT-reconstruction leads to results which are inferior to the DIC method with respect to all reported measures.

4.2. Results for 3D low-dose CT

Figure 4 shows all the intermediate results obtained with the proposed method. Figure 4(A) shows the initial FBP-reconstruction which is contaminated by noise. The FBP-reconstruction was then processed using the function f_{θ} described in (9) to obtain the prior \mathbf{x}_{CNN} which can be seen in figure 4(B). From the point-wise error, we see that patch-wise post-processing with the 3D U-net removed a large portion of the noise resulting from the low-dose acquisition. Solving problem (19) increases data-consistency since we make use of the measured data \mathbf{y}_{η} . Note that in contrast to the previous example of undersampled radial MRI, the minimization of the functional increased data-consistency of the solution but also contaminated the solution with noise, since the measured data is noisy due to the simulated low-dose scan protocol. Table 2 summarizes the obtained quantitative measures for all intermediate reconstructions of our approach as well as for the TV and the DIC method. In the first three columns of table 2 we see the results obtained for all three intermediate reconstructions of our proposed scheme. The reconstruction metrics improved substantially from the FBP-reconstruction to the estimated prior \mathbf{x}_{CNN} . The difference in terms of PSNR was almost 10 dB, while the NRMSE decreased by approximately 0.11. Further, the similarity measures SSIM and HPSI were increased by about 0.14 and 0.04, respectively. Finally, the estimated solution given by \mathbf{x}_{REC} which was obtained by performing $n_{\text{iter}} = 4$ iterations of Landweber to minimize (19) showed a slight decrease in PSNR and NRMSE which is related to the use of the noisy-measured data. However, fine diagnostic details as the coronary arteries are still visible in the prior \mathbf{x}_{CNN} and in the solution \mathbf{x}_{REC} as indicated by the yellow arrows. SSIM slightly increased while HPSI stayed approximately the same.

Figure 5 shows a comparison of images obtained by the different reconstruction methods. In figure 5(A), we see again the FBP-reconstruction obtained from the noisy data. Figure 5(B) shows the result obtained by the TV-minimization method which removed some of the noise as can be taken from the point-wise error image. The result obtained by the DIC method can be seen in figure 5(C) which further reduced image noise compared to the TV method and surpasses TV with respect to the reported statistics, as can be seen in table 2. Finally, figure 5(D) shows the solution \mathbf{x}_{REC} obtained with our proposed scheme and figure 5(E) shows the ground truth image. The reconstruction using the CNN output as a prior further increased the SSIM of the final result and is visually more similar to the ground truth image. The NRMSE and PSNR on the other hand were slightly reduced due to the use of the noisy measured data. HPSI remained approximately the same as can be seen from table 2.

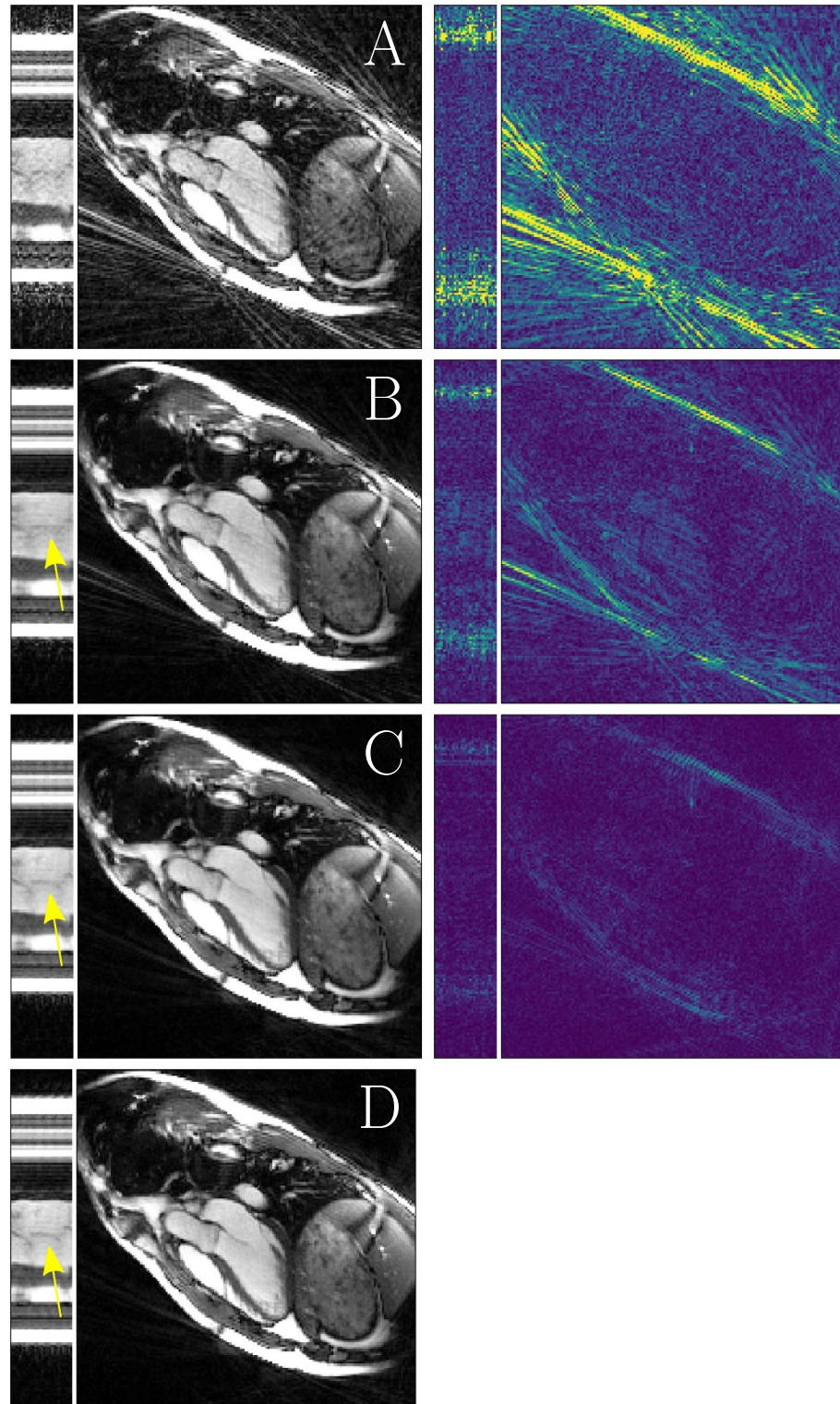


Figure 2. Results for a healthy volunteer showing two slices with different orientations. A: Initial NUFFT-reconstruction \mathbf{x}_I using $N_\theta = 1130$ radial spokes, B: estimated output \mathbf{x}_{CNN} using the spatio-temporal 2D XT,YT U-net, C: solution of the CNNs-based regularized functional \mathbf{x}_{REC} , D: ground truth image reconstruction with *kt*-SENSE and $N_\theta = 3400$ radial spokes. All images are displayed in the same scale. For better visibility, the point-wise error images are magnified by a factor of $\times 3$. The yellow arrows point at details which are smoothed out in the CNN-prior \mathbf{x}_{CNN} but are visible again in the final reconstruction \mathbf{x}_{REC} .

4.3. Reconstruction times

Table 3 summarizes the times for the different components of the reconstructions using all different approaches for both examples. The abbreviations ‘SHRINK’ and ‘LS1’ stand for ‘shrinkage’ and ‘linear system - one iteration’ and denote the times which are needed to apply the iterative shrinkage method for the TV approach and to solve the sub-problems which are solved using iterative schemes, respectively.

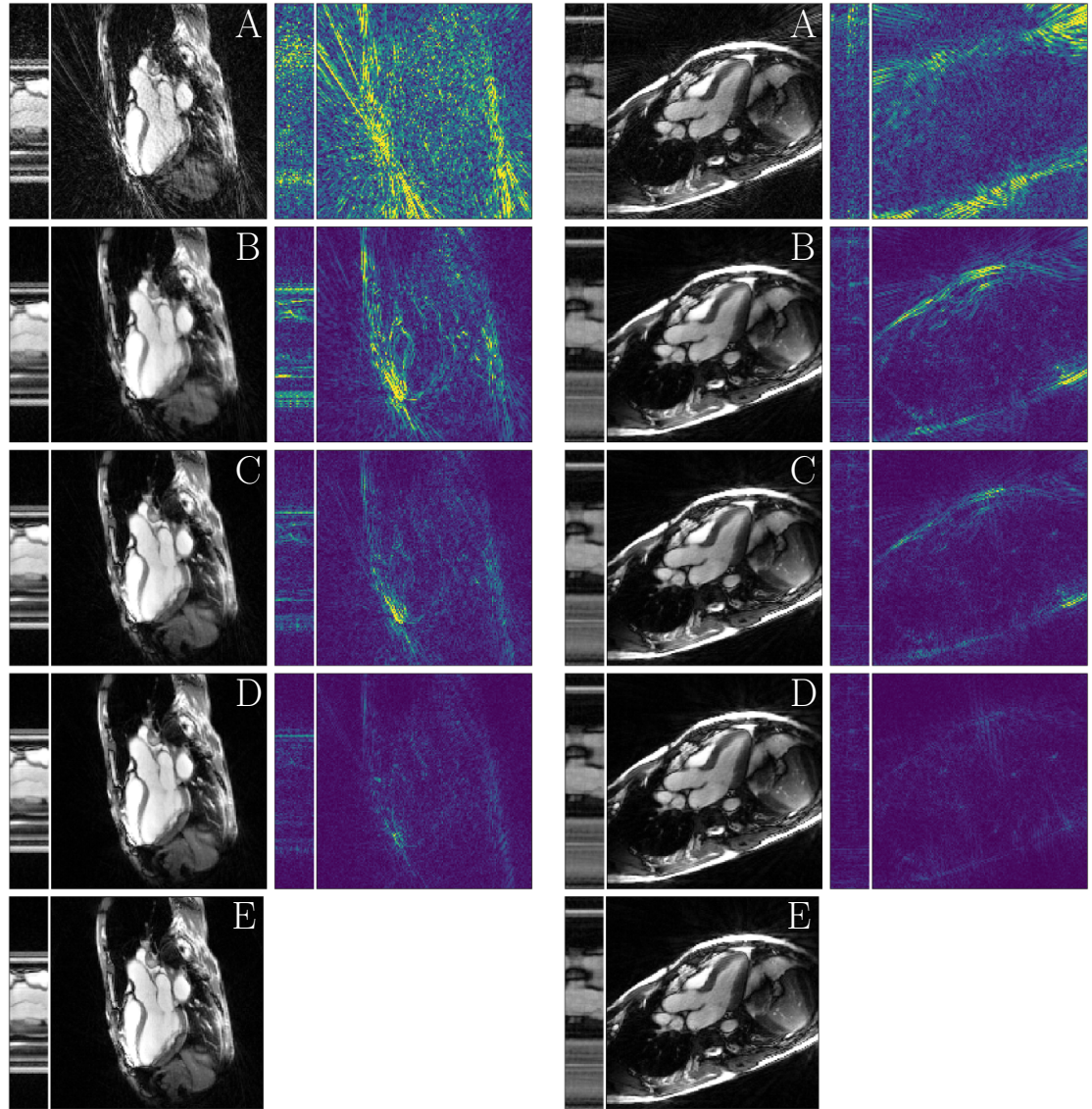
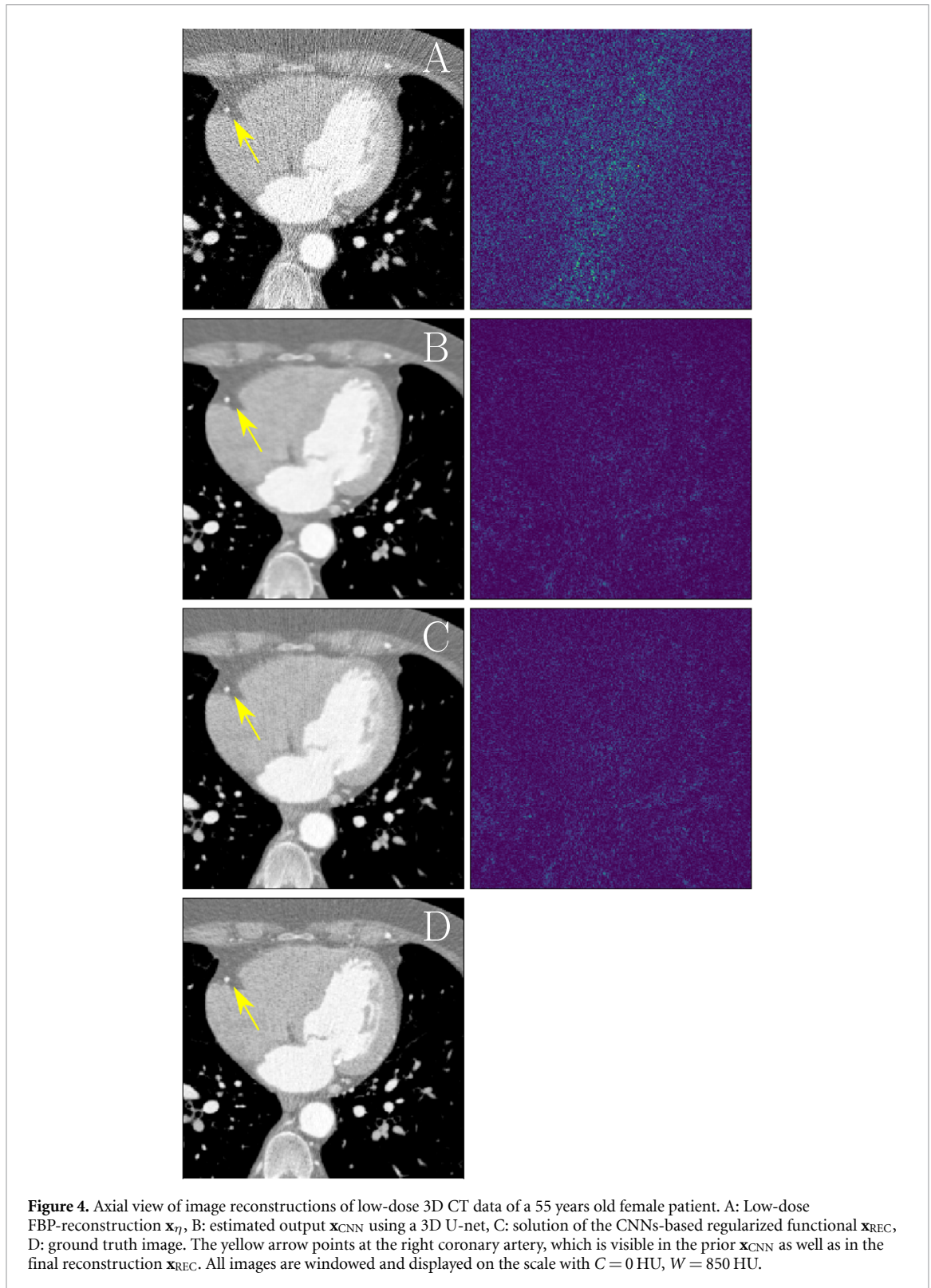


Figure 3. Results for a patient (left panel) and a healthy volunteer (right panel). A: Initial NUFFT-reconstruction \mathbf{x}_I using $N_\theta = 1130$ radial spokes, B: solution of the TV-minimization approach (TV), C: dictionary learning-based regularization solution (DIC), D: CNN-regularized solution \mathbf{x}_{REC} , E: ground truth images obtained by *kt*-SENSE using $N_\theta = 3400$ radial spokes. All images are displayed in the same scale. For better visibility, the point-wise error images are magnified by a factor of $\times 5$. The point-wise error is the lowest for the reconstruction \mathbf{x}_{REC} .

Table 3. Reconstruction and processing times for the different methods for one 3D CT volume and a 2D cine MR image sequence.

| | | 3D Low Dose CT | 2D Radial Cine MRI |
|--|---------------------------|----------------------|------------------------|
| $\mathbf{A}^\dagger \mathbf{y}_\eta$ TV | SHRINK | ≈ 23 s (FBP) | ≈ 11 s (NUFFT) |
| | LS1 | $\ll 1$ s | $\ll 1$ s |
| | Total | ≈ 40 s | $\approx 1:20$ m |
| DIC | \mathbf{x}_{DIC} | ≈ 11 m | ≈ 42 m |
| | LS1 | $\approx 1:24$ h | ≈ 7 m |
| | Total | ≈ 40 s | $\approx 1:20$ m |
| Proposed | \mathbf{x}_{CNN} | $\approx 1:28$ h | ≈ 28 m |
| | LS1 | ≈ 4 m | ≈ 5 s |
| | Total | ≈ 40 s | $\approx 1:20$ m |
| | | ≈ 8 m | ≈ 21 m |

Obviously, in terms of achieved image quality, the advantage of the DIC- and the CNN-based Tikhonov regularization are given by obtaining stronger priors which allow to use a smaller number of iterations to regularize the solution. The advantage of our proposed approach compared to the dictionary learning-based is the highly reduced time to compute the prior which is used for regularization. The reason lies in the fact



that the DIC-based method requires to solve problem (23) to obtain the prior \mathbf{x}_{DIC} , while in our method a CNN is used to obtain the prior \mathbf{x}_{CNN} . Since problem (23) is separable, OMP is applied for each image/volume patch which is prohibitive as the number of overlapping patches in a 3D volume is in the order of $\mathcal{O}(N_x \cdot N_y \cdot N_z)$ or $\mathcal{O}(N_x \cdot N_y \cdot N_t)$, respectively. Obtaining \mathbf{x}_{CNN} , on the other hand, does not involve the solution of any minimization problem but only requires the application of the network u_θ to the different patches. As this corresponds to matrix-vector multiplications with sparse matrices, its computational cost is lower and the calculations are further highly accelerated by performing the computations on a GPU.

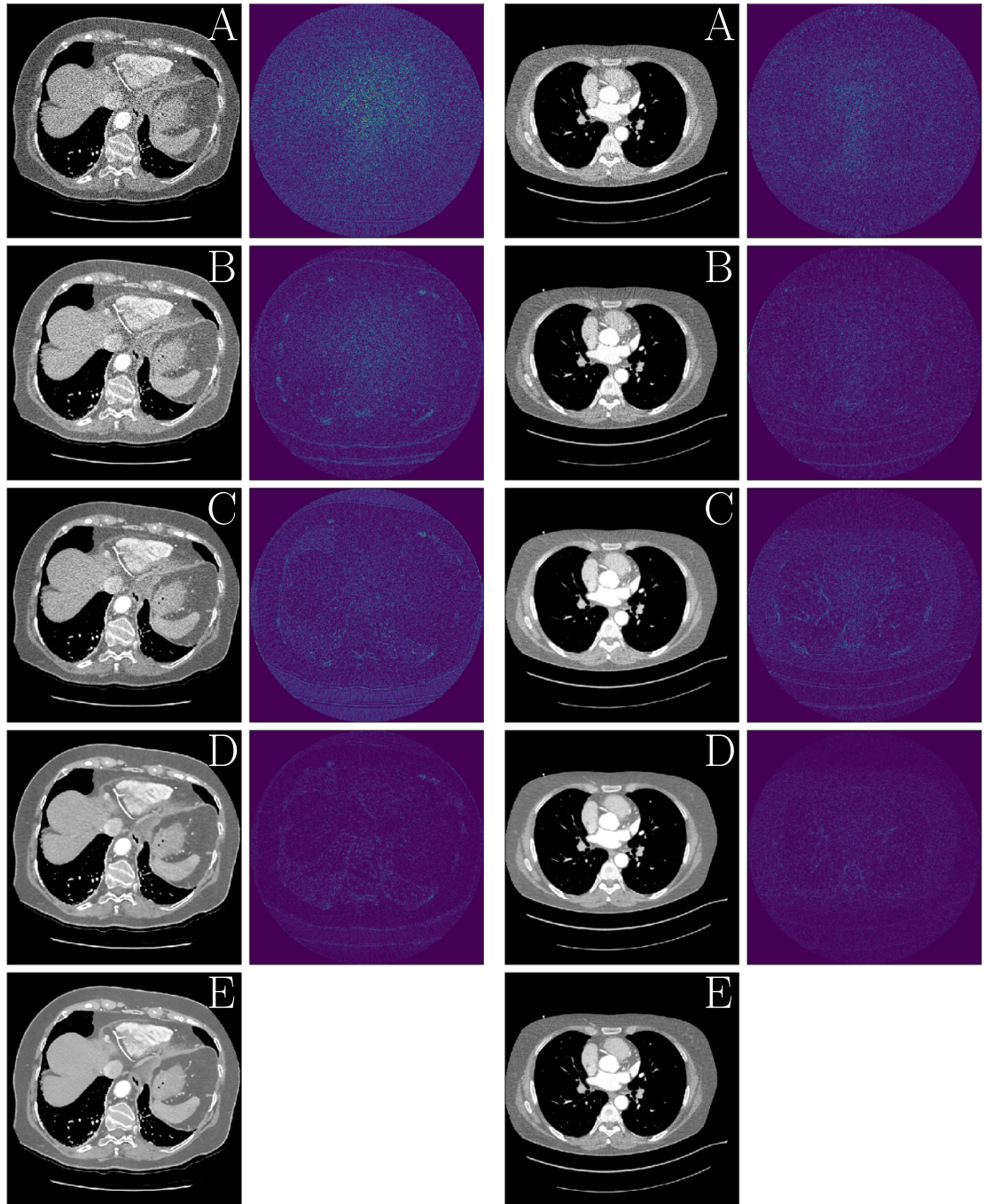


Figure 5. Axial view of image reconstructions of low-dose 3D CT data of two female patients of age 76 (left) and 55 (right). A: Low-dose FBP-reconstruction \mathbf{x}_{76} , B: TV-minimization based reconstruction (TV), C: DIC-regularization based reconstruction (DIC), D: CNN-regularization based reconstruction \mathbf{x}_{REC} , E: ground truth image. All images are windowed and displayed on the same scale with $C = 0$ HU, $W = 800$ HU.

5. Discussion

The proposed three-steps reconstruction scheme provides a general framework for solving large-scale inverse problems. The method is motivated by the observations stated in the ablation study (Kofler *et al* 2018), where the performance of cascades of CNNs with different numbers of intercepting data-consistency layers but approximately fixed number of trainable parameters was studied. First, it was noted that the replacement of simple blocks of convolutional layers by multi-scale CNNs given by U-nets had a visually positive impact on the obtained results. Further, it was empirically shown that the results obtained by cascades of U-nets of different length but with approximately the same number of trainable parameters were all visually and quantitatively comparable in terms of all reported measures. This suggests that, for large-scale problems, where the construction of cascaded networks might be infeasible, investing the same computational effort

and expressive power in terms of number of trainable parameters in one single network might be similarly beneficial to intercepting several smaller sub-networks by data-consistency layers as for example in Schlemper *et al* (2018), Qin *et al* (2018).

Due to the large sizes of the considered objects of interest, the prior \mathbf{x}_{CNN} is obtained by processing patches of the images. Training the network on patches or slices of the images further has the advantage of reducing the computational overhead while naturally enlarging the available training data and therefore being able to successfully train neural networks even with datasets coming from a relatively small number of subjects. Further, as demonstrated in Kofler *et al* (2019), for the case of 2D radial MRI, one can also exploit the low topological complexity of 2D spatio-temporal slices for training the network u_θ . This allows to reduce the network complexity by using 2D- instead of 3D-convolutional layers and still exploiting spatio-temporal correlations and therefore to prevent overfitting. Note that the network architectures we are considering are CNNs and, since they mainly consist of convolutional and max-pooling layers, we can expect the networks to be translation-equivariant and therefore, patch-related artefacts arising from the re-composition of the processed overlapping patches are unlikely to occur in the CNN-prior.

We have tested and evaluated our method on two examples of large-scale inverse problems given by 2D undersampled radial MRI and 3D low-dose CT. For both examples, our method outperformed the TV-minimization method and the dictionary learning-based method with respect to all reported quantitative measures. For the case of 2D undersampled radial cine MRI, using the CNN-prior as a regularizer in the subsequent iterative reconstruction increased the achieved image quality with respect to all reported measures, as can be taken from table 1. For the CT example, due to the inherent presence of noise in the measured data, the quantitative measures of the final reconstruction are only similar to the ones obtained by post-processing the FBP-reconstruction. However, performing a few iterations to minimize functional (19), increased data-consistency of the obtained solution and resulted in a slight re-enhancement of the edges and gave back the CT images their characteristic texture. Future work to qualitatively assess the achieved image quality with respect to clinically relevant features, e.g. the visibility of coronary arteries for the assessment of coronary artery disease in cardiac CT, is already planned.

Using the CNN for obtaining a learning-based prior is faster by several orders of magnitude compared to the dictionary learning-based approach. This is because obtaining the prior with a CNN reduces to a forward pass of all patches, i.e. to multiplications of vectors with sparse matrices, where instead, the sparse coding of all patches involves the solution of an optimization problem for each patch. Further, the time needed for OMP is dependent on the sparsity level and the number of atoms of the dictionary, see Sturm and Christensen (2012). In our comparison, for the 2D radial MRI example, the total reconstruction times of our proposed method and the DIC-based regularization method mainly differ in the step of obtaining the priors \mathbf{x}_{DIC} and \mathbf{x}_{CNN} . Note that, in contrast to Wang and Ying (2014) and Caballero *et al* (2014), in our comparison, the prior \mathbf{x}_{DIC} was only calculated once. In the original works, however, the proposed reconstruction algorithms use an alternating direction method of multipliers (ADMM) which alternates between first training the dictionary \mathbf{D} and sparse coding with OMP and then updating the image estimate. Therefore, the realistic time needed to reconstruct the 2D cine MR images according to Wang *et al* (2004) and Caballero *et al* (2014) is given by the product of the seven minutes needed for one sparse approximation and the number of iterations in the ADMM algorithm and the total time used for PCG for solving the obtained linear systems. Note that for the 3D low-dose CT example, even one patch-wise sparse approximation of the whole volume already takes about one hour and therefore, applying an ADMM type of reconstruction method is computationally prohibitive. Also, note that, even if the size of the image sequences for the MRI example is smaller than the one of the 3D CT volumes, the reconstruction of the 2D cine MR images takes relatively long compared to the CT example due to the fact that we use two different iterative methods (Landweber and PCG) for two different systems with different operators. Further, the number of iterations for the CT example is on purpose smaller than for the MR example, as the measurement data is noisy and early stopping of the iteration can already be thought of as a proper regularization method, see for example Strand (1974). Also, the operators used for the CT examples were implemented by using the operators provided by the ODL library and are therefore optimized for performing calculations on the GPU. On the other hand, for the MRI example, we used our own implementation of a radial encoding operator \mathbf{E} which could be further improved and accelerated.

Clearly, one difficulty of the proposed method is the one shared by all iterative reconstruction schemes with regularization: the need to choose the hyper-parameter λ which balances the contribution of the regularization and the data-fidelity term can highly affect the achieved image quality, especially when the data is contaminated by noise. In cascaded networks, the parameter λ can on the other hand be learned as well during training. Further, some other hyper-parameters as the number of iterations to minimize Tikhonov functional have to be chosen as well. In this work, we empirically chose λ but point out that an exhaustive parameter search might yield superior results.

A limitation of presented work is that the applied CNNs were task-specific and trained for only one undersampling factor in the MR example and for one dose-reduction level in the CT example. However, note that this issue can be easily overcome by for example including a greater variety of samples in the training data. Further, more sophisticated approaches could involve regularization of the CNN based on generative adversarial networks (GANs) (Rick Chang *et al* 2017). In Rick Chang *et al* (2017), it was reported that it is possible to train a single CNN to solve arbitrary inverse problems. Using a similar training strategy for the CNN and combining it with our approach could in particular overcome the problem arising from the use of different acceleration factors in MR or dose-reduction levels in CT.

The proposed method is related to the ones presented in Schlemper *et al* (2018), Qin *et al* (2018), Kofler *et al* (2018) in the sense that steps 2 and 3 in Algorithm 1 are iterated in a cascaded network which represents the different iterations. However, in Schlemper *et al* (2018) and Qin *et al* (2018), the encoding operator is given by a Fourier transform sampled on a Cartesian grid and therefore is an isometry. Thus, assuming a single-coil data-acquisition, given \mathbf{x}_{CNN} , the solution of (3) has a closed-form solution which is also fast and cheap to compute since it corresponds to performing a linear combination of the acquired k -space data and the one estimated from the CNN outputs and subsequently applying the inverse Fourier transform. In the case where the operator \mathbf{A} is not an isometry, one usually needs to either solve a system of linear equations in order to obtain a solution which matches the measured data or, alternatively, rely on another formulation of the functional (3) which is suitable for more general, also non-orthogonal operators (Kofler *et al* 2018). However, if the operator \mathbf{A} and its adjoint \mathbf{A}^H are computationally demanding to apply as in the case of radial multi-coil MRI, or if the objects of interest are high-dimensional, e.g. 3D volumes in low-dose CT, the construction of cascaded or iterative networks is prohibitive with nowadays available hardware. In contrast, in the proposed approach, since the regularization is separated from the data-consistency step, large-scale problems can be tackled as well. Further, in Shan *et al* (2019) and Shan *et al* (2019), efficient solutions for making the CNN applicable to images with different noise levels in low-dose CT were proposed and compared to commercial algorithms for low-dose CT image reconstruction. By employing a CNN which can successfully remove different levels of noise or artefacts from images, the procedure proposed in Algorithm 1 can also be iterated.

By separating the application of the CNN, the regularization and further iterations needed to obtain the reconstruction, one can also choose to employ more complex and sophisticated NNs to obtain the CNN-prior \mathbf{x}_{CNN} as it is typically the case for cascaded or iterative networks. For example, in Schlemper *et al* (2018) or Adler and Öktem (2017), the CNNs were given by simple blocks of fully convolutional neural networks with residual connection. In contrast, in Kofler *et al* (2018), the CNNs were replaced by more sophisticated U-nets Ronneberger *et al* (2015), Jin *et al* (2017). However, the examples in Kofler *et al* (2018), Adler and Öktem (2017, 2018), Gupta *et al* (2018) all use two-dimensional CT geometries, which do not correspond to the ones used in clinical practice. Therefore, particularly for large-scale inverse problems where the construction of iterative networks is infeasible, our method represents a valid alternative to obtain accurate reconstructions.

While in this work we used relatively simple neural network architectures based on a plain U-net as in Jin *et al* (2017), further focus could be put on the choice of the network u_θ , also by using more sophisticated approaches, e.g. improved versions of the U-net Han and Ye (2018) or generative adversarial networks for obtaining a more accurate prior to be further used in the proposed reconstruction scheme.

6. Conclusion

We have presented a general framework for solving large-scale ill-posed inverse problems in medical image reconstruction. The strategy consists in strictly separating the application of the CNN, the regularization of the solution and the step needed to ensure data-consistency by solving the problem in three stages. First, an initial guess of the solution is obtained by the direct reconstruction from the measured data. As a second step, the initial solution is patch-wise processed by a previously trained CNN in order to obtain a CNN image-prior which is then used in a Tikhonov-regularized functional to obtain the final reconstruction in a third step. The strict separation of the steps of obtaining a CNN-prior and then subsequently minimizing a Tikhonov-functional allows to tackle large-scale problems. For both shown examples of 2D undersampled radial MRI and 3D low-dose CT, the proposed method outperformed the total variation-minimization method and the dictionary learning-based approach with respect to all reported quantitative measures. Since the reconstruction scheme is a general one, we expect the proposed method to be successfully applicable to other imaging modalities as well.

Acknowledgment

A Kofler and M Dewey acknowledge the support of the German Research Foundation (DFG), project number GRK2260, BIOQIC. We thank V Wieske for providing clinically relevant cases for the 3D low-dose CT experiments.

ORCID iDs

A Kofler  <https://orcid.org/0000-0001-9169-2572>
M Haltmeier  <https://orcid.org/0000-0001-5715-0331>
T Schaeffter  <https://orcid.org/0000-0003-1310-2631>
M Kachelrieß  <https://orcid.org/0000-0001-9351-4761>
M Dewey  <https://orcid.org/0000-0002-4402-2733>
C Wald  <https://orcid.org/0000-0003-2373-9492>
C Kolbitsch  <https://orcid.org/0000-0002-4355-8368>

References

- Adler J, Kohr H and Oktem O 2017 Operator discretization library (<https://github.com/odgroup/odl>)
- Adler J and Öktem O 2017 Solving ill-posed inverse problems using iterative deep neural networks *Inverse Problems* **33** 124007
- Adler J and Öktem O 2018 Learned primal-dual reconstruction *IEEE Trans. Med. Imaging* **37** 1322–32
- Aggarwal H K, Mani M P and Jacob M 2018 Modl: Model-based deep learning architecture for inverse problems *IEEE Trans. Med. Imaging* **38** 394–405
- Block K T, Uecker M and Frahm J 2007 Undersampled radial mri with multiple coils. iterative image reconstruction using a total variation constraint *Magn. Reson. Med.* **57** 1086–98
- Caballero J, Price A N, Rueckert D and Hajnal J V 2014 Dictionary learning and time sparsity for dynamic MR data reconstruction *IEEE Trans. Med. Imaging* **33** 979–94
- Chambolle A 2005 Total variation minimization and a class of binary MRF models *Int. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* (Berlin: Springer) pp 136–52
- Chun I Y, Zheng X, Long Y and Fessler J A 2019 BCD-net for low-dose CT reconstruction: Acceleration, convergence and generalization *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 31–40
- Chun Y and Fessler J A 2018 Deep BCD-net using identical encoding-decoding CNN structures for iterative image recovery 2018 *IEEE 13th Image, Video and Multidimensional Signal Processing Workshop (IVMSP)* (Piscataway, NJ: IEEE) pp 1–5
- Dewey M, Zimmermann E, Deissenrieder F, Laule M, Dübel H P, Schlattmann P, Knebel F, Rutsch W and Hamm B 2009 Noninvasive coronary angiography by 320-row computed tomography with lower radiation exposure and maintained diagnostic accuracy: comparison of results with cardiac catheterization in a head-to-head pilot investigation *Circulation* **120** 867–75
- Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* vol 375 (Berlin: Springer)
- Grasmair M 2010 Generalized Bregman distances and convergence rates for non-convex regularization methods *Inverse Problems* **26** 115014
- Gupta H, Jin K H, Nguyen H Q, McCann M T and Unser M 2018 CNN-based projected gradient descent for consistent CT image reconstruction *IEEE Trans. Med. Imaging* **37** 1440–53
- Hammernik K, Klatzer T, Kobler E, Recht M P, Sodickson D K, Pock T and Knoll F 2018 Learning a variational network for reconstruction of accelerated MRI data *Magn. Reson. Med.* **79** 3055–71
- Han Y, Sunwoo L and Ye J C 2019 *k*-space deep learning for accelerated MRI *IEEE Trans. Med. Imaging* **39** 377–386
- Han Y and Ye J C 2018 Framing U-net via deep convolutional framelets: Application to sparse-view CT *IEEE Trans. Med. Imaging* **37** 1418–29
- Hauptmann A, Arridge S, Lucka F, Muthurangu V and Steeden J A 2019 Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning—proof of concept in congenital heart disease *Magn. Reson. Med.* **81** 1143–56
- Hauptmann A, Lucka F, Betcke M, Huynh N, Adler J, Cox B, Beard P, Ourselin S and Arridge S 2018 Model-based learning for accelerated, limited-view 3-D photoacoustic tomography *IEEE Trans. Med. Imaging* **37** 1382–93
- Hyun C M, Kim H P, Lee S M, Lee S and Seo J K 2018 Deep learning for undersampled MRI reconstruction *Phys. Med. Biol.* **63** 135007
- Jin K H, McCann M T, Froustey E and Unser M 2017 Deep convolutional neural network for inverse problems in imaging *IEEE Trans. Image Process.* **26** 4509–22
- Kingma D P and Ba J 2014 Adam: A method for stochastic optimization (arXiv:1412.6980)
- Kofler A, Dewey M, Schaeffter T, Wald C and Kolbitsch C 2020 Spatio-temporal deep learning-based undersampling artefact reduction for 2D radial cine MRI with limited training data *IEEE Trans. Med. Imaging* **39** 703–17
- Kofler A, Haltmeier M, Kolbitsch C, Kachelrieß M and Dewey M 2018 A U-nets cascade for sparse view computed tomography *Int. Workshop on Machine Learning for Medical Image Reconstruction* (Berlin: Springer) pp 91–9
- Li H, Schwab J, Antholzer S and Haltmeier M 2020 *Inverse Problems* Nett: Solving inverse problems with deep neural networks
- Lustig M, Donoho D L, Santos J M and Pauly J M 2008 Compressed sensing MRI *IEEE Signal Process. Mag.* **25** 72
- Napp A E et al 2017 Computed tomography versus invasive coronary angiography: design and methods of the pragmatic randomised multicentre discharge trial *Eur. Radiol.* **27** 2957–68
- Qin C, Schlemper J, Caballero J, Price A N, Hajnal J V and Rueckert D 2018 Convolutional recurrent neural networks for dynamic MR image reconstruction *IEEE Trans. Med. Imaging* **38** 280–90
- Qin C, Schlemper J, Duan J, Seegoolam G, Price A, Hajnal J and Rueckert D 2019 *k-t* NEXT: Dynamic MR image reconstruction exploiting spatio-temporal correlations *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (Lecture Notes in Computer Science, vol 11765)* (Berlin: Springer) pp 505–13
- Rasche V, Proksa R, Sinkus R, Bornert P and Eggers H 1999 Resampling of data between arbitrary grids using convolution interpolation *IEEE Trans. Med. Imaging* **18** 385–92

- Reisenhofer R, Bosse S, Kutyniok G and Wiegand T 2018 A Haar wavelet-based perceptual similarity index for image quality assessment *Signal Process. Image Commun.* **61** 33–43
- Rick Chang J, Li C L, Poczos B, Vijaya Kumar B and Sankaranarayanan A C 2017 One network to solve them all—solving linear inverse problems using deep projection models *Proc. of the IEEE Int. Conf. on Computer Vision (Venice, 22–29 October 2017)* (Piscataway, NJ: IEEE) pp 5888–97
- Ronneberger O, Fischer P and Brox T 2015 U-net: Convolutional networks for biomedical image segmentation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (Lecture Notes in Computer Science, vol 9351)* (Berlin: Springer) pp 234–41
- Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F 2009 *Variational Methods in Imaging* (Berlin: Springer)
- Schlemper J, Caballero J, Hajnal J V, Price A N and Rueckert D 2018 A deep cascade of convolutional neural networks for dynamic MR image reconstruction *IEEE Trans. Med. Imaging* **37** 491–503
- Schnass K 2018 Convergence radius and sample complexity of ITKM algorithms for dictionary learning *Appl. Comput. Harmon. Anal.* **45** 22–58
- Schwab J, Antholzer S and Haltmeier M 2019 Deep null space learning for inverse problems: convergence analysis and rates *Inverse Problems* **35** 025008
- Shan H, Kruger U and Wang G 2019 A novel transfer learning framework for low-dose CT *Proc. SPIE* **11072** 110722Y
- Shan H, Padole A, Homayounieh F, Kruger U, Khera R D, Nitiwarangkul C, Kalra M K and Wang G 2019 Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction *Nature Machine Intelligence* **1** 269–76
- Smith D S, Sengupta S, Smith S A and Brian Welch E 2019 Trajectory optimized NUFFT: Faster non-Cartesian MRI reconstruction through prior knowledge and parallel architectures *Magn. Reson. Med.* **81** 2064–71
- Strand O N 1974 Theory and methods related to the singular-function expansion and Landweber's iteration for integral equations of the first kind *SIAM J. Numer. Anal.* **11** 798–825
- Sturm B L and Christensen M G 2012 Comparison of orthogonal matching pursuit implementations 2012 *Proc. of the 20th European Signal Conf. (EUSIPCO)* (Piscataway, NJ: IEEE) pp 220–4
- Tian Z, Jia X, Yuan K, Pan T and Jiang S B 2011 Low-dose CT reconstruction via edge-preserving total variation regularization *Phys. Med. Biol.* **56** 5949
- Tropp J A and Gilbert A C 2007 Signal recovery from random measurements via orthogonal matching pursuit *IEEE Trans. Information Theory* **53** 4655–66
- Wang Y and Ying L 2014 Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary *IEEE Trans. Biomed. Eng.* **61** 1109–20
- Wang Z, Bovik A C, Sheikh H R and Simoncelli E P et al 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Process.* **13** 600–12
- Winkelmann S, Schaeffter T, Koehler T, Eggers H and Doessel O 2006 An optimal radial profile order based on the golden ratio for time-resolved MRI *IEEE Trans. Med. Imaging* **26** 68–76
- Xu Q, Yu H, Mou X, Zhang L, Hsieh J and Wang G 2012 Low-dose x-ray CT reconstruction via dictionary learning *IEEE Trans. Med. Imaging* **31** 1682–97
- Yang Q et al 2018 Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss *IEEE Trans. Med. Imaging* **37** 1348–57
- Zhu B, Liu J Z, Cauley S F, Rosen B R and Rosen M S 2018 Image reconstruction by domain-transform manifold learning *Nature* **555** 487