

Nr. 87  
9. Jun 2022

Preprint-Series: Department of Mathematics - Applied Mathematics

## Convolutional Dictionary Learning by End-To-End Training of Iterative Neural Networks

A. Kofler, C. Wald, T. Schaeffter, M. Haltmeier, C. Kolbitsch



AppliedMathematics

---

Technikerstraße 13 - 6020 Innsbruck - Austria  
Tel.: +43 512 507 53803 Fax: +43 512 507 53898  
<https://applied-math.uibk.ac.at>

# Convolutional Dictionary Learning by End-To-End Training of Iterative Neural Networks

Andreas Kofler

*Physikalisch-Technische Bundesanstalt*  
Braunschweig and Berlin, Germany  
andreas.kofler@ptb.de

Christian Wald

*Department of Radiology*  
*Charité - Universitätsmedizin Berlin*  
Berlin, Germany  
christian.wald@charite.de

Tobias Schaeffter

*Physikalisch-Technische Bundesanstalt*  
Braunschweig and Berlin, Germany  
tobias.schaeffter@ptb.de

Markus Haltmeier

*Department of Mathematics*  
*University of Innsbruck*  
Innsbruck, Austria  
markus.haltmeier@uibk.ac.at

Christoph Kolbitsch

*Physikalisch-Technische Bundesanstalt*  
Braunschweig and Berlin, Germany  
christoph.kolbitsch@ptb.de

**Abstract**—Sparsity-based methods have a long history in the field of signal processing and have been successfully applied to various image reconstruction problems. The involved sparsifying transformations or dictionaries are typically either pre-trained using a model which reflects the assumed properties of the signals or adaptively learned during the reconstruction - yielding so-called blind Compressed Sensing approaches. However, by doing so, the transforms are never explicitly trained in conjunction with the physical model which generates the signals. In addition, properly choosing the involved regularization parameters remains a challenging task. Another recently emerged training-paradigm for regularization methods is to use iterative neural networks (INNs) - also known as unrolled networks - which contain the physical model. In this work, we construct an INN which can be used as a supervised and physics-informed online convolutional dictionary learning algorithm. We evaluated the proposed approach by applying it to a realistic large-scale dynamic MR reconstruction problem and compared it to several other recently published works. We show that the proposed INN improves over two conventional model-agnostic training methods and yields competitive results also compared to a deep INN. Further, it does not require to choose the regularization parameters and - in contrast to deep INNs - each network component is entirely interpretable.

**Index Terms**—Iterative Neural Networks, Sparsity, Convolutional Dictionary Learning, Compressed Sensing, Cardiac Cine MRI

## I. INTRODUCTION

Recently, image reconstruction has become an important application area of machine learning (ML) and has attracted the interest of researchers in different communities. The most prominent type of ML-approaches are so-called iterative neural networks (INNs) or unrolled networks, see e.g. [1], [2], [4]. INNs correspond to iterative reconstruction schemes of finite length and typically consists of regularizing blocks, e.g. convolutional layers, and so-called data-consistency blocks which make use of the forward and adjoint operators. Integrating the physical model into the network reduces the maximum error-bound [5] and improves the network in terms of generalization

properties [6]. In addition, INNs were reported to be the most robust with respect to adversarial attacks compared to model-agnostic networks [7]. Despite their success, *deep* INNs are still black-boxes from a theoretical point of view. There is still no systematic way to design the employed convolutional blocks or to analyze them which raises concerns about their application in a sensitive field such as medical imaging. Nevertheless, the possibility to integrate the physical model into the learning process is a unique feature of INNs which distinguishes them from other learning-based methods which, on the other hand, often come with theoretical advantages such as convergence guarantees [8].

Convolutional dictionary learning (CDL) is a well-established approach which has been extensively applied to different tasks such as image denoising [8], image inpainting [9] and image reconstruction [10]. However, these methods typically learn the dictionary by solving the CDL-problem and are not necessarily tailored to the reconstruction algorithm they are subsequently used with. In addition, properly tuning the regularization parameters is crucial (as we shall see later in the experiments) and is often a tedious and difficult task.

In this work, we exploit the benefits of INNs - i.e. the inclusion of the physical model into the architecture - to construct a physics-informed and entirely interpretable ML-based regularization method using CDL. The proposed method corresponds to an unrolled reconstruction scheme using a learned convolutional dictionary. The filters of the convolutional dictionary as well as the involved regularization parameters - can then be trained in a supervised and physics-aware manner. We apply our method to a realistic large-scale dynamic MR image reconstruction problem and demonstrate that i) the filters obtained by the NN-training yield better reconstructions compared to the ones obtained by decoupled pre-training of the filters as in [8] and [3] and ii) that the proposed method shows competitive results compared to a state-of-the-art method with deep iterative NNs [2].

## II. METHODS

Our data measurement model corresponds to a general type of linear inverse problem

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

where  $\mathbf{A}$  models the measurement process,  $\mathbf{x}$  denotes the image,  $\mathbf{e}$  denotes Gaussian random noise and  $\mathbf{y}$  the measured data. The goal is to recover the unknown image  $\mathbf{x}$  from the observed measured data  $\mathbf{y}$ . Because such problems can be ill-posed for a variety of reasons, e.g. undersampling, poorly-conditioned systems or combinations thereof, the reconstruction requires the use of regularization methods which take additional information into account.

### A. Problem Formulation

Similar to [10], we consider the following regularized reconstruction problem

$$\min_{\mathbf{x}, \{\mathbf{s}_k\}_k} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\lambda}{2} \left\| \mathbf{x} - \sum_{k=1}^K d_k * \mathbf{s}_k \right\|_2^2 + \alpha \sum_{k=1}^K \|\mathbf{s}_k\|_1, \quad (2)$$

where  $\{d_k\}_{k=1}^K$  denotes a set of *fixed* convolutional filters with  $\|d_k\|_2 = 1$  for all  $k$  and  $\lambda, \alpha > 0$ . Because the variables  $\mathbf{s}_k$  are linked with the convolution operator within the  $L_2$ -norm and are also present within the non-smooth  $L_1$ -norm, directly minimizing (2) is difficult. We therefore introduce auxiliary variables  $\{\mathbf{u}_k\}_k$  under the constraint  $\mathbf{u}_k = \mathbf{s}_k$  for all  $k$ , yielding the equivalent problem

$$\min_{\mathbf{x}, \{\mathbf{s}_k\}_k, \{\mathbf{u}_k\}_k} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\lambda}{2} \left\| \mathbf{x} - \sum_{k=1}^K d_k * \mathbf{s}_k \right\|_2^2 + \alpha \sum_{k=1}^K \|\mathbf{u}_k\|_1 \quad \text{s.t.} \quad \forall k : \mathbf{u}_k - \mathbf{s}_k = \mathbf{0}. \quad (3)$$

Our aim is to set up a reconstruction algorithm based on (3) that defines a reconstruction network whose regularization parameters and filters  $\{d_k\}_k$  are trained in an end-to-end manner.

### B. Reconstruction Algorithm

The proposed solution strategy is to define an alternating minimization scheme by minimizing (3) with respect to one of the variables by keeping the others fixed. For fixed  $\mathbf{x}$ , problem (3) takes the form of

$$\min_{\{\mathbf{s}_k\}_k, \{\mathbf{u}_k\}_k} \frac{\lambda}{2} \left\| \mathbf{x} - \sum_{k=1}^K d_k * \mathbf{s}_k \right\|_2^2 + \alpha \sum_{k=1}^K \|\mathbf{u}_k\|_1 \quad \text{s.t.} \quad \forall k : \mathbf{u}_k - \mathbf{s}_k = \mathbf{0}, \quad (4)$$

for which we use the alternating direction method of multipliers (ADMM) [11] to update  $\{\mathbf{s}_k\}_k$  and  $\{\mathbf{u}_k\}_k$ . For  $0 \leq j \leq T$ , the update-rules are given by

$$\begin{aligned} \{\mathbf{s}_k\}_k^{(j+1)} &= \arg \min_{\{\mathbf{s}_k\}_k} \frac{\lambda}{2} \left\| \mathbf{x} - \sum_{k=1}^K d_k * \mathbf{s}_k \right\|_2^2 \\ &+ \frac{\beta}{2} \sum_{k=1}^K \|\mathbf{u}_k^{(j)} - \mathbf{s}_k + \mathbf{z}_k^{(j)}\|_2^2, \end{aligned} \quad (5)$$

$$\begin{aligned} \{\mathbf{u}_k\}_k^{(j+1)} &= \arg \min_{\{\mathbf{u}_k\}_k} \alpha \sum_{k=1}^K \|\mathbf{u}_k\|_1 \\ &+ \frac{\beta}{2} \sum_{k=1}^K \|\mathbf{u}_k - \mathbf{s}_k^{(j+1)} + \mathbf{z}_k^{(j)}\|_2^2, \end{aligned} \quad (6)$$

$$\{\mathbf{z}_k\}_k^{(j+1)} = \{\mathbf{z}_k\}_k^{(j)} + \{\mathbf{u}_k\}_k^{(j+1)} - \{\mathbf{s}_k\}_k^{(j+1)}, \quad (7)$$

where  $\{\mathbf{z}_k\}_k$  are the dual-variables of the ADMM algorithm.

1) *Update of  $\{\mathbf{s}_k\}_k$* : Updating  $\{\mathbf{s}_k\}_k$  corresponds to the most challenging problem and is achieved by solving the sub-problem in the Fourier-domain. Define

$$\mathbf{x}^f := \mathbf{F}\mathbf{x}, \mathbf{d}_k^f := \mathbf{F}d_k, \mathbf{u}_k^f := \mathbf{F}\mathbf{u}_k, \mathbf{s}_k^f := \mathbf{F}\mathbf{s}_k, \mathbf{z}_k^f := \mathbf{F}\mathbf{z}_k. \quad (8)$$

Then, by making use of the Fourier-convolution theorem, updating  $\{\mathbf{s}_k\}_k$  according to (5) is equivalent to solving

$$\min_{\{\mathbf{s}_k^f\}_k} \frac{1}{2} \left\| \mathbf{x}^f - \sum_{k=1}^K d_k^f \cdot \mathbf{s}_k^f \right\|_2^2 + \frac{\gamma}{2} \sum_{k=1}^K \|\mathbf{u}_k^f - \mathbf{s}_k^f + \mathbf{z}_k^f\|_2^2, \quad (9)$$

where the multiplication  $d_k^f \cdot \mathbf{s}_k^f$  is meant component-wise and  $\gamma =: \beta/\lambda$ . Problem (9) can be re-written in a more compact form as

$$\min_{\mathbf{s}^f} \frac{1}{2} \left\| \mathbf{x}^f - \mathbf{D}^f \mathbf{s}^f \right\|_2^2 + \frac{\gamma}{2} \|\mathbf{u}^f + \mathbf{z}^f - \mathbf{s}^f\|_2^2, \quad (10)$$

where  $\mathbf{D}^f := [\mathbf{D}_1^f, \dots, \mathbf{D}_K^f] = [\text{diag}(d_1^f), \dots, \text{diag}(d_K^f)]$ ,  $\mathbf{s}^f := [\mathbf{s}_1^f, \dots, \mathbf{s}_K^f]^\top$ ,  $\mathbf{u}^f := [\mathbf{u}_1^f, \dots, \mathbf{u}_K^f]^\top$  and  $\mathbf{z}^f := [\mathbf{z}_1^f, \dots, \mathbf{z}_K^f]^\top$ . Solving problem (10) corresponds to solving a linear system

$$((\mathbf{D}^f)^\mathbf{H} \mathbf{D}^f + \gamma \mathbf{I}) \mathbf{s}^f = (\mathbf{D}^f)^\mathbf{H} \mathbf{x}^f + \gamma (\mathbf{u}^f + \mathbf{z}^f) \quad (11)$$

which, because of its special structure, has a computationally inexpensive closed-form solution which can be obtained using the Sherman-Morrison formula, see [12] for more details. Let  $(\mathbf{s}^f)^*$  denote the solution of (10). Then we finally obtain  $\mathbf{s}_k^{(j+1)} := \mathbf{F}^\mathbf{H}(\mathbf{s}_k^f)^*$  for all  $k$  as the solution of (5).

2) *Update of  $\{\mathbf{u}_k\}_k$  and  $\{\mathbf{z}_k\}_k$* : Since the sub-problem in (6) is separable w.r.t.  $k$ , updating  $\{\mathbf{u}_k\}_k$  according to (6) is efficiently implemented by applying the soft-thresholding operator for all  $k$ , i.e.

$$\mathbf{u}_k^{(j+1)} = \mathcal{S}_{\alpha/\beta}(\mathbf{z}_k^{(j)} - \mathbf{s}_k^{(j+1)}), \quad (12)$$

where  $\mathcal{S}_{\alpha/\beta}$  denotes the soft-thresholding operator with threshold  $\alpha/\beta$ . Updating the dual variables according to (7) is trivial.

3) *Update of  $\mathbf{x}$* : Last, for fixed  $\{\mathbf{s}_k\}_k$  and  $\{\mathbf{u}_k\}_k$ , updating  $\mathbf{x}$  corresponds to solving the sub-problem

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\lambda}{2} \left\| \mathbf{x} - \sum_{k=1}^K d_k * \mathbf{s}_k \right\|_2^2 \quad (13)$$

which can be achieved by solving the system  $\mathbf{H}\mathbf{x} = \mathbf{b}$  with

$$\mathbf{H} = \mathbf{A}^\mathbf{H} \mathbf{A} + \lambda \mathbf{I} \quad (14)$$

$$\mathbf{b} = \mathbf{A}^\mathbf{H} \mathbf{y} + \lambda \sum_{k=1}^K d_k * \mathbf{s}_k \quad (15)$$

by any iterative solver, e.g. a conjugate gradient (CG)-method.

### C. Proposed Reconstruction Network

Note that the updates of the respective variables describe a solution strategy for solving problem (3) where the set of convolutional filters  $\{d_k\}_k$  as well as the parameters  $\lambda$ ,  $\alpha$  and  $\beta$  are *fixed*. We construct a neural network whose blocks correspond to the operations required to solve the previously described sub-problems. Then, the filters  $\{d_k\}_k$  as well as  $\alpha$ ,  $\beta$  and  $\lambda$  are free parameters in the reconstruction network that can be learned by back-propagation on a set of input-target image pairs  $\mathcal{D} = \{(\mathbf{x}_0^i, \mathbf{x}_f^i)_{i=1}^M\}$ , where  $\mathbf{x}_f$  denotes a ground-truth image which is used as label during training and  $\mathbf{x}_0 := \mathbf{A}^\dagger \mathbf{y}$  denotes an initial guess of the image which is typically directly obtained from the measured data by some reconstruction operator  $\mathbf{A}^\dagger$ .

Typically, CDL algorithms impose a norm constraint on the filters, i.e.  $\|d_k\|_2 = 1$ , to avoid the scaling ambiguity between the filters and the sparse coefficient maps [12], [8]. Since we train the filters by back-propagation, we project the filters onto the unit-sphere after each filter-update by rescaling them by the corresponding  $L_2$ -norm.

### D. Implementation Details

Although the implementation of the single components of the proposed network might seem straight-forward at a first glance, a closer look at the sub-problems reveals that some of the operations require careful treatment. First, the Fourier-convolution theorem used to obtain (9) only holds when circular padding conditions are assumed. Second, from (9), one can see that the dimensionality of the filters in the Fourier-domain has to match the one of the image, which means that the filters have to be properly zero-padded. Third, to be able to use the Sherman-Morrison formula for solving (10), one needs to construct the right hand side of the system (11), where again, by proper circular and zero-padding, one needs to ensure that  $(\mathbf{D}^f)^\mathbf{H}$  and therefore implicitly  $\mathbf{D}^\mathbf{H}$  are indeed the adjoint operators of  $\mathbf{D}^f$  and  $\mathbf{D}$ , respectively. This means that they need to fulfill  $\langle \mathbf{D}\mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{a}, \mathbf{D}^\mathbf{H}\mathbf{b} \rangle$  and  $\langle \mathbf{D}^f \mathbf{a}^f, \mathbf{b}^f \rangle = \langle \mathbf{a}^f, (\mathbf{D}^f)^\mathbf{H} \mathbf{b}^f \rangle$  for all  $\mathbf{a}$  and  $\mathbf{b}$ . Last, because we treat complex-valued images as two-channelled images and the filters are shared across the real and the imaginary part, these aspects also have to be taken into account when constructing all relevant operators. An implementation of our network is available at [www.github.com/koflera/ConvSparsityNNs](https://www.github.com/koflera/ConvSparsityNNs).

## III. EXPERIMENTS

We applied our proposed method to an accelerated cardiac cine MR image reconstruction problem using radial  $k$ -space trajectories. The operator  $\mathbf{A}$  in (1) takes the form

$$\mathbf{A} := (\mathbf{I}_{N_c} \otimes \mathbf{E})\mathbf{C}, \quad (16)$$

for a complex valued image  $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_{N_t}]^\mathbf{T} \in \mathbb{C}^N$  with  $N = N_x \times N_y \times N_t$ , where  $\mathbf{I}_{N_c}$  denotes an identity operator and  $\mathbf{C}$  contains the  $N_c$  coil-sensitivity maps, i.e.  $\mathbf{C} = [\mathbf{C}_1, \dots, \mathbf{C}_{N_c}]^\mathbf{T}$ , with  $\mathbf{C}_j = \text{diag}(\mathbf{c}_j, \mathbf{c}_j, \dots, \mathbf{c}_j) \in \mathbb{C}^{N \times N}$  and  $\mathbf{c}_j \in \mathbb{C}^{N_x \times N_y}$ . The operator  $\mathbf{E} = \text{diag}(\mathbf{E}_1, \dots, \mathbf{E}_{N_t})$

consists of different 2D radial Fourier-encoding operators  $\mathbf{E}_t$  which for each time point  $t \in \{1, \dots, N_t\}$  sample a 2D image  $\mathbf{x}_t \in \mathbb{C}^{N_x \times N_y}$  along radial lines in Fourier-space. To accelerate the acquisition process, we only acquire a sub-portion of the  $k$ -space coefficients indexed by the set  $I$  and denote the resulting operator by  $\mathbf{A}_I$ . The  $k$ -space trajectories are chosen according to the golden-angle method and  $\mathbf{A}_I$  was implemented using the library `TorchKBNufft` [14]. Further, we used  $k$ -space pre-conditioning, see [13] for more details.

### A. Dataset

We used a dataset of 15 healthy volunteers and four patients with a total of 216 cine MR images of shape  $N_x \times N_y \times N_t = 320 \times 320 \times 30$ . The data was split into 12/3/4 subjects (144/36/36 dynamic images) for training, validation and testing. The test set consisted of the four patients. As in [13], the initial  $k$ -space data with  $N_c = 12$  coils was retrospectively simulated using  $N_\theta = 36$  spokes per cardiac phase and was corrupted by Gaussian noise with a standard deviation of  $\sigma = 0.02$  resulting in an approximate SNR 8.5.

### B. 3D vs 2D

Because training INNs can be computationally demanding, we also consider a variant of problem (2), where instead of using 3D filters, we use 2D filters and impose the regularization for each time-frame, i.e. the regularization term in (2) changes to

$$\min_{\{\mathbf{s}_{k,t}\}_k} \frac{\lambda}{2} \sum_{t=1}^{N_t} \left( \left\| \mathbf{x}_t - \sum_{k,t}^K d_k * \mathbf{s}_{k,t} \right\|_2^2 + \alpha \sum_{k=1}^K \|\mathbf{s}_{k,t}\|_1 \right). \quad (17)$$

This allows to use a larger number of filters during training at the price of not exploiting the temporal dimension for the regularization which is the one with the highest correlation among neighbouring pixels. Note that this formulation only changes the regularization term but not the reconstruction algorithm. Within the network, the temporal dimension of the image  $\mathbf{x}$  can be dynamically merged back and forth to play the role of the batch-dimension depending on the stage of the reconstruction.

### C. Methods of Comparison and Evaluation

Since our proposed approach is a method for training a convolutional dictionary, we compare it to two classical model-agnostic CDL-methods, which we refer to as CDL 2D and CDL 3D. For this, we pre-trained the 2D and 3D convolutional filters using [8] and [3] with  $\alpha = 1.0$  and  $\alpha = 0.2$ , respectively. After having obtained the filters, we fixed them in our network and only trained  $\lambda$ ,  $\alpha$  and  $\beta$  to exclude a possible change of performance which could be attributed to a sub-optimal choice of the regularization parameters. Further, we also compared with the method in [13], which corresponds to the *analysis operator* counterpart of this work and which we abbreviate by NN-CAOL 3D. We abbreviate our proposed methods by NN-CDL 3D and NN-CDL 2D, respectively. Last, we also compared our method to an adaptation of the deep cascade of CNNs presented in [2], which we denote by DnCn3D, see [13] for details. We trained all CDL-methods

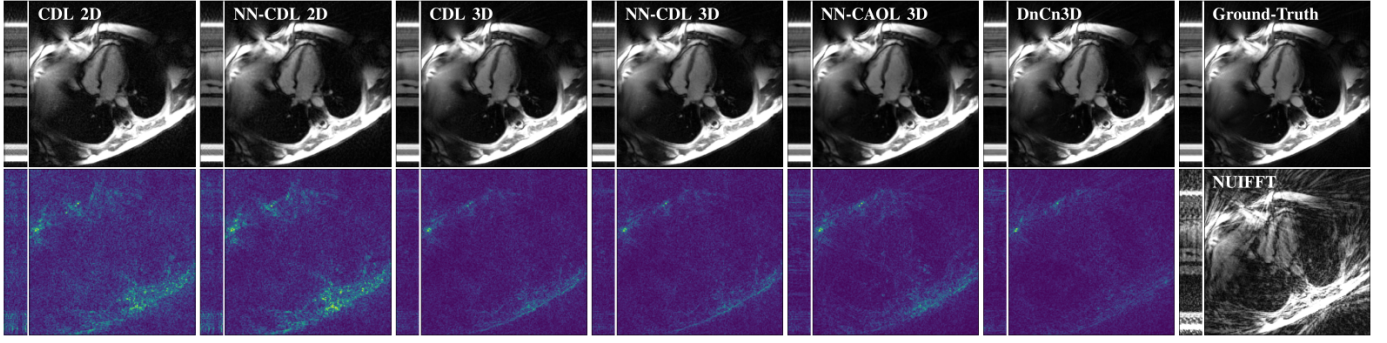


Figure 1. An example of images reconstructed with the different reported methods: CDL 2D [3], NN-CDL 2D, CDL 3D [8], NN-CDL 3D, NN-CAOL 3D [13] and DnCn3D [2]. For CDL 3D, NN-CDL 3D and NN-CAOL 3D, we used  $K = 16$  filters with  $k_f = 7$ , and for NN-CDL 2D we used  $K = 96$  and  $k_f = 9$ . Further, the target image and the initial reconstruction given by a non-uniform inverse FFT (NUFFT) are shown.

for different filter configurations  $K = 16, 24$  with filters of shapes  $k_f \times k_f \times k_f = 7 \times 7 \times 7$  for 3D and  $K = 64, 96$  with  $k_f \times k_f = 9 \times 9$  for 2D. Based on a hyper-parameter selection on the validation set, we finally used  $K = 16$  and  $k_f = 7$  for CDL 3D, NN-CDL 3D and CAOL 3D (see [13]), while for CDL 2D and NN-CDL 2D, we used  $K = 96$  and  $k_f = 9$ . The length of the networks and the number of CG-iterations for solving (14) were set to  $T = 4$  and  $n_{CG} = 12$  for all CDL-methods. Because problem (2) is separable w.r.t. the temporal dimension, we trained all INNs on images only containing  $N_t = 4$  and  $N_t = 12$  cardiac phases for NN-CDL 2D and NN-CDL 3D, respectively. The INNs were trained for 16 epochs by minimizing the squared  $L_2$ -error between the target image and the estimated one using ADAM with an initial learning rate of  $5 \cdot 10^{-4}$ . All reconstructions were evaluated in terms of PSNR, NRMSE and SSIM which were calculated over a central squared ROI of  $160 \times 160$  pixels.

#### IV. RESULTS AND DISCUSSION

Figure 1 shows an example of images reconstructed by all used methods. All methods yielded accurate reconstructions by successfully removing undersampling artefacts and noise. By comparing CDL 2D to NN-CDL 2D and CDL 3D to NN-CDL 3D, we see that using the proposed INN resulted in a more accurate reconstruction, by slightly better removing residual noise. Further, by comparing NN-CAOL 3D with CDL 3D and NN-CDL 3D, we see that the CDL approach seems to give results which are marginally better than for the analysis operator approach. At no surprise, all sparsity-based methods are surpassed by DnCn3D which, however - being an INN with *deep* convolutional blocks which also differ from iteration to iteration - contains eight to eleven times more trainable parameters than NN-CDL 2D and NN-CDL 3D, respectively, and is hardly interpretable. Table I lists the average of the statistics obtained over all 2D images of each cardiac cycle of the test set. The table well-reflects the visually observed results.

Figure 2 shows the training- and the validation-error during the optimization of the proposed network. As we can see, training the filters by back-propagation further reduced the

error compared to using the pre-trained filters obtained by [3] and [8]. The effect is clearly more pronounced for the 3D case. Further, we see how drastic the impact of not carefully chosen  $\beta$  and  $\lambda$  can be on CDL 2D and CDL 3D. Note that, although CDL-2D and CDL-3D in the end achieve results which are comparable with the proposed method, they only do so after a subsequent optimization of  $\lambda$ ,  $\beta$  and  $\alpha$  using the proposed INN. This demonstrates that the proposed INN can also be used as an algorithm for the selection of the regularization parameters.

Table I  
QUANTITATIVE RESULTS FOR ALL DIFFERENT METHODS AVERAGED OVER THE TEST SET.

	NRMSE	PSNR	SSIM	# Params
$\mathbf{A}_I^\#$	0.360	36.449	0.634	-
CDL 2D [3]	0.121	45.557	0.942	3 891
NN-CDL 2D	0.119	45.683	0.946	3 891
CDL 3D [8]	0.081	49.227	0.972	2 747
NN-CDL 3D	0.079	49.503	0.974	2 747
NN-CAOL 3D [13]	0.090	47.934	0.965	2 746
DnCn3D [2]	0.076	50.191	0.973	31 233

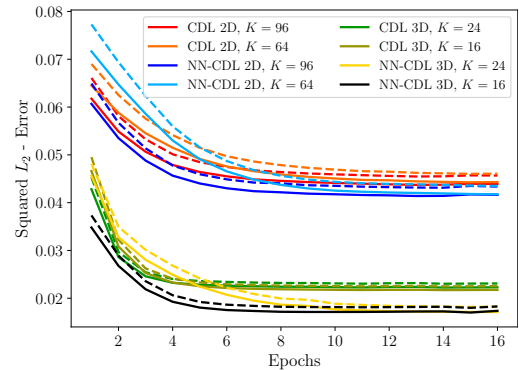


Figure 2. Training- and validation-error (solid and dashed lines) for all CDL-methods. Training the convolutional filters with the proposed INN further reduces the validation-error compared to the de-coupled methods in [3] and [8]. In addition,  $\lambda$ ,  $\alpha$  and  $\beta$  can be trained as well. Note that for CDL 2D and CDL 3D, the filters were kept fixed and only  $\lambda$ ,  $\alpha$  and  $\beta$  were trained.

Because the proposed method is interpretable, we can verify that the learned filters indeed serve the purpose they were

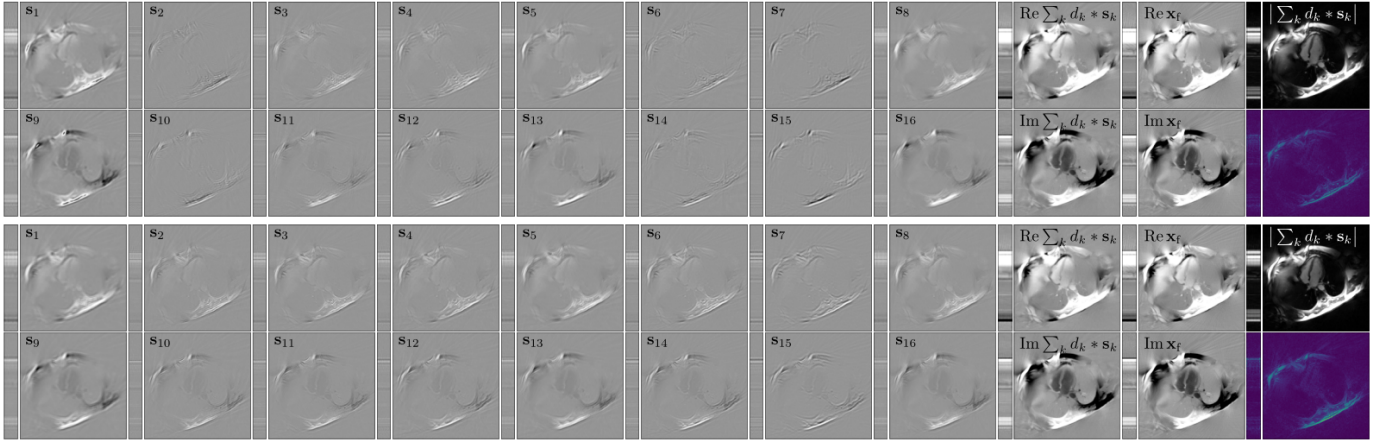


Figure 3. Estimates of the sparse feature maps  $\{s_k\}_k$  for  $K = 16$  and  $k_f = 7$  after having reconstructed the image solving problem (2) using the 3D filters of NN-CDL 3D (top) and the ones of CDL 3D (bottom). Further, the real and the imaginary parts of the approximation  $\sum_k d_k * s_k$  as well as the magnitude image and the corresponding error-images are shown. The sparse feature maps are displayed in a window  $[-15\sigma, 15\sigma]$ , where  $\sigma$  denotes the standard deviation calculated over all 16 feature maps. As can be seen, the obtained sparse feature maps are relatively similar for both approaches. The filters obtained by the proposed NN-CDL slightly better approximate the ground-truth image and removed a larger portion of the undersampling artefacts.

designed for. Figure 3 shows the estimates of  $\{s_k\}_k$  at the last iteration of CDL 3D and NN-CDL 3D. We can see that  $\sum_k d_k * s_k$  indeed gives an approximation of the image and the approximation using the filters of NN-CDL 3D exhibits a lower point-wise error compared to CDL 3D.

## V. CONCLUSION

In this work, we have presented a physics-informed method for supervised training of a convolutional dictionary for image reconstruction applied to accelerated cardiac MRI. The proposed method consists of an iterative neural network (INN) which is rigorously derived from a properly defined sparse reconstruction problem. The filters of the dictionary are trained in a supervised manner. We have compared our method to two unsupervised model-agnostic methods and showed competitive results where, most importantly, our method does not require the choice of the regularization parameters as they are jointly trained with the filters. The proposed INN also achieved results comparable to its analysis operator counterpart and to a method based on deep INNs. On top, each learned component of the proposed reconstruction method is entirely interpretable in contrast to INNs with deep convolutional blocks.

## REFERENCES

- [1] Jonas Adler and Ozan Öktem, “Solving ill-posed inverse problems using iterative deep neural networks,” *Inverse Problems*, vol. 33, no. 12, pp. 124007, 2017.
- [2] Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert, “A deep cascade of convolutional neural networks for dynamic MR image reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, 2018.
- [3] Jialin Liu, Cristina Garcia-Cardona, Brendt Wohlberg, Wotao Yin, “First- and second-order methods for online convolutional dictionary learning,” *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 1589–1628, 2018.
- [4] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll, “Learning a variational network for reconstruction of accelerated MRI data,” *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.
- [5] Andreas K Maier, Christopher Syben, Bernhard Stimpel, Tobias Würfl, Mathis Hoffmann, Frank Schebesch, Weilin Fu, Leonid Mill, Lasse Kling, and Silke Christiansen, “Learning with known operators reduces maximum error bounds,” *Nature machine intelligence*, vol. 1, no. 8, pp. 373–380, 2019.
- [6] Andreas Kofler, Markus Haltmeier, Tobias Schaeffter, and Christoph Kolbitsch, “An end-to-end-trainable iterative network architecture for accelerated radial multi-coil 2d cine MR image reconstruction,” *Medical Physics*, vol. 48, no. 5, pp. 2412–2425, 2021.
- [7] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen, “On instabilities of deep learning in image reconstruction and the potential costs of AI,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30088–30095, 2020.
- [8] Il Yong Chun and Jeffrey A Fessler, “Convolutional dictionary learning: Acceleration and convergence,” *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1697–1712, 2017.
- [9] Vardan Pappyan, Yaniv Romano, Jeremias Sulam, and Michael Elad, “Convolutional dictionary learning via local processing,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5296–5304, 2017.
- [10] Tran Minh Quan and Won-Ki Jeong, “Compressed sensing dynamic mri reconstruction using gpu-accelerated 3d convolutional sparse coding,” in *International conference on medical image computing and computer-assisted intervention*. Springer, pp. 484–492, 2016.
- [11] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato and Jonathan Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [12] Brendt Wohlberg, “Efficient algorithms for convolutional sparse representations,” *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 301–315, 2015.
- [13] Andreas Kofler, Christian Wald, Tobias Schaeffter, Markus Haltmeier, and Christoph Kolbitsch, “Convolutional analysis operator learning by end-to-end training of iterative neural networks,” in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, 2022. doi=10.1109/ISBI52829.2022.9761621.
- [14] Matthew J Muckley, Ruben Stern, Tullie Murrell, and Florian Knoll, “Torchkbuff: A high-level, hardware-agnostic non-uniform fast Fourier transform,” in *ISMRM Workshop on Data Sampling & Image Reconstruction*, 2020.