NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction

Zaccharie Ramzi, Chaithya G.R., Jean-Luc Starck, Philippe Ciuciu

NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction

Zaccharie Ramzi, Chaithya G R, Jean-Luc Starck and Philippe Ciuciu *Senior Member, IEEE*

***Abstract*— Deep Learning has become a very promising avenue for magnetic resonance image (MRI) reconstruction. In this work, we explore the potential of unrolled networks for the non-Cartesian acquisition setting. We design the *NC-PDNet*, the first density-compensated unrolled network and validate the need for its key components via an ablation study. Moreover, we conduct some generalizability experiments to test our network in out-of-distribution settings, for example training on knee data and validating on brain data. The results show that the *NC-PDNet* outperforms the baseline models visually and quantitatively in the 2D settings. Additionally, in the 3D settings, it outperforms them visually. In particular, in the 2D multi-coil acquisition scenario, the *NC-PDNet* provides up to a 1.2 dB improvement in peak signal-to-noise ratio (PSNR) over baseline networks, while also allowing a gain of at least 1 dB in PSNR in generalization settings. We provide the open-source implementation of our network, and in particular the Non-uniform Fourier Transform in TensorFlow, tested on 2D multi-coil and 3D data.**[1]

I. Introduction

Magnetic Resonance Imaging (MRI) is the reference imaging technique used to probe soft tissues in the human body non-invasively. The data acquisition process in MRI is inherently slow due to the sequential measurements collection in the Fourier domain, also called k-space. This slow acquisition causes many issues such as motion-corrupted scans, limited image resolution in a scan time compatible with clinical routine, non-applicability to certain patients and limited patient throughput. To reduce the scan time as a means to offset the limiting factors, the main solution proposed was to diminish the number of measurements in the k-space. It was exploited for parallel imaging (PI) [2], [3] as well as for the use of Compressed Sensing (CS) in MRI [4]. In this last framework, a manually crafted decomposition basis is used to express the sparsity of the image to be reconstructed in a transform domain. From this sparsity prior, an optimization problem, whose solution should be close to the image of interest, can be constructed.

This optimization problem stems from casting the reconstruction as an idealized inverse problem (forgetting magnetic gradient inaccuracies and B_0 field inhomogeneities for example). It can be written in its most general form as:

$$\boldsymbol{y}_\ell = \mathcal{F}_\Omega \boldsymbol{S}_\ell \boldsymbol{x} + \boldsymbol{n}_\ell \quad \forall \ell = 1, \dots, L \tag{1}$$

with L the number of coils used for the acquisition, $\mathcal{F}_\Omega$ the measurements operator, a 2D or 3D Non-uniform Discrete Fourier Transform (NDFT) [5] with locations Ω (if Ω is on a grid, $\mathcal{F}_\Omega = \Omega \boldsymbol{F}$ can be simplified to a regular Fourier Transform $\boldsymbol{F}$ followed by a masking operator parameterized by Ω). Diagonal matrices $(\boldsymbol{S}_\ell)_{\ell=1}^L$ define the sensitivity maps (in the case of a single coil acquisition, $L = 1$ and $\boldsymbol{S}_1$ is just the identity matrix) and $\boldsymbol{y}_\ell \in \mathbb{C}^m$ are the k-space measurements associated with each coil, $\boldsymbol{n}_\ell$ is an additive white Gaussian noise (AWGN), and $\boldsymbol{x} \in \mathbb{C}^n$ the target image to reconstruct.

The reconstructed image $\widehat{\boldsymbol{x}}$ is then recovered by solving the following optimization problem:

$$\arg\min_{\boldsymbol{x} \in \mathbb{C}^n} \sum_{\ell=1}^{L} \frac{1}{2} \|\boldsymbol{y}_\ell - \mathcal{F}_\Omega \boldsymbol{S}_\ell \boldsymbol{x}\|_2^2 + \mathcal{R}(\boldsymbol{x}) \tag{2}$$

where $\mathcal{R}$ is a regularisation term that serves as a prior of the solution. Usually, $\mathcal{R}$ is hand-crafted and can be the ℓ_1-norm of the decomposition of $\boldsymbol{x}$ in some adequate basis like the wavelets, or the gradient of the image (TV regularization).

A prototypical version of an iterative algorithm that can solve the optimization problem (2) is the following proximal gradient descent:

$$\begin{aligned} \boldsymbol{x}_{n+1} &= \boldsymbol{x}_n - \epsilon_n \mathcal{A}^H \left(\mathcal{A}\boldsymbol{x}_n - \mathbb{y}\right) \\ \boldsymbol{x}_{n+1} &= \text{prox}_{\epsilon_n \mathcal{R}}\left(\boldsymbol{x}_{n+1}\right) \end{aligned} \tag{3}$$

where $\mathcal{A} = \left(\boldsymbol{I}_L \otimes \mathcal{F}_\Omega\right)\mathbb{S}$, $\mathbb{S} = [\boldsymbol{S}_1^H, \dots, \boldsymbol{S}_L^H]^T$ and $\mathbb{y} = [\boldsymbol{y}_1^H, \dots, \boldsymbol{y}_L^H]^T$ and ϵ_n is the gradient step size at the n-th iteration and ^{H} is the Hermitian (i.e. transpose conjugate) operator. The following issues can be spotted using this method: (i) iterations have computationally-intensive operators, (ii) the prior $\mathcal{R}$ needs to be manually crafted and ideally its proximal operator has to be closed form. While dictionary learning methods [6] can alleviate the second issue, they do so at an increased computation cost but still with relatively simple (i.e. linear) learned priors.

To circumvent these problems, Deep Learning (DL) methods have become of great interest for MRI reconstruction. Indeed, the unrolled framework [7] allows for the use of less iterations for image reconstruction while enabling the neural network to learn more complex priors on MR images from the data itself. Hence, it tackles both problems (i)-(ii) observed in CS methods at the same time. Moreover, the current deep learning frameworks offer an off-the-shelf efficient implementation on GPU and TPU which also facilitates gains

Zaccharie Ramzi, Chaithya G R and P. Ciuciu are with CEA, NeuroSpin, F-91191 Gif-sur-Yvette, cedex, France and Inria, Parietal, Université Paris-Saclay, F-91120 Palaiseau, France, whereas J.-L. Starck is affiliated with AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité F-91191 Gif-sur-Yvette, France.

[1]This work is an extended version of the work presented at the 2021 ISBI conference [1].

in terms of reconstruction speed. While the first deep learning methods applied to MRI reconstruction were not necessarily using the unrolled framework [8], the recent state-of-the-art methods all rely on this framework [9]–[12]. Most of these approaches are employed to reconstruct images from data collected on a Cartesian grid, and therefore cannot be tested in the more general non-Cartesian acquisition setting, when the measurements are collected off the grid.

Non-Cartesian MRI acquisitions are of interest for multiple purposes. For example, the radial acquisition that happens with projection acquisition can be used for shorter echo times, reduced sensitivity to motion and improved temporal resolution in some applications [13]. The same statement holds for spiral acquisition [13]. As these acquisition setups are used in clinical routines, it is crucial to push the development of DL reconstruction methods for their application to these types of data.

Moreover, the interest in going from 2D imaging to 3D imaging can be justified by a higher signal-to-noise ratio (SNR) requested for targeting higher resolution in clinical use [14]. While multiple approaches exist to efficiently sample in 3D [15], [16], here we focus on the simple stacking strategy (e.g. stack of stars or stack of spirals) [17], [18]. Moreover, due to hardware constraints, we focus on single-coil 3D imaging rather than multi-coil 3D imaging, and leave this research direction open for future work.

In this paper, we contribute to the domain of non-Cartesian MRI reconstruction by:

- introducing the *NC-PDNet*, the first density-compensated unrolled network for non-Cartesian data;
- implementing the first unrolled networks for non-Cartesian multi-coil 2D and single-coil 3D data;
- open sourcing our implementation for all the networks, data processing, training and evaluation[2] in TensorFlow [19], in particular a version of the Non-Uniform Fast Fourier Transform (NUFFT) in the TensorFlow framework, usable for multi-coil and 3D data (with the corresponding density compensation code)[3];
- performing some out-of-distribution performance tests for the trained networks.

II. Related Works

a) Correcting k-space for non-Cartesian reconstruction: This work strongly builds on the seminal paper by *Pipe and Menon* [20] introducing density compensation (DC) for non-Cartesian acquired data. In this work, the authors provide a solution for computing the DC which avoids some requirements that may be difficult to meet in practice, such as the knowledge of the trajectories or the constraint associated with the Nyquist criterion. However, this work does not use a learning reconstruction strategy to complement the application of the DC.

Some other works have explored other possibilities to correct the k-space for non-Cartesian reconstruction, using a pre-conditioning of the k-space in iterative CS procedures in order to accelerate the convergence (not necessarily to obtain better image quality for an unlimited computational budget). The authors in [21] introduced a diagonal pre-conditionner optimized for the Mean Squared Error (MSE) that did not need an inner loop. This was an improvement over the work of [22]. However, these works use classical iterative algorithms which have been shown to be outperformed by end-to-end learned unrolled approaches in various settings [9], [23]–[25].

b) Unrolled neural networks: The idea of unrolling an optimization algorithm was originally introduced for the problem of sparse coding [7]. It has then been applied to many different inverse problems and in particular in MRI [9], [10], [25], [26]. While an unrolled neural network architecture can be created for all the different iterative algorithms solving (2) (e.g. ADMM [27], Neumann Series [28], Dictionary Learning [25] or Alternating Minimization [24]), in this work we focus only on one of them, the XPDNet framework [29], as it is the highest ranked framework with an open-source implementation in the fastMRI 2020 reconstruction challenge [12], giving the name to our network: the *NC-PDNet*. The aim of this paper is not actually to compare different strategies for the unrolled architecture but instead to demonstrate the improvement inherent to the density compensation mechanism for a given unrolled architecture.

c) Neural networks for non-Cartesian reconstruction: The Nonuniform Variational Network [25] was the first unrolled network to be designed for non-Cartesian MRI reconstruction. This network unrolls a proximal gradient descent algorithm, and operates on 192×192 single-coil images with a variable density acquisition scheme. This network, however, does not include a DC scheme, which is most likely due to the fact that the trajectories studied in this work do not present an off-balanced density. Additionally, this network was only applied to single-coil 2D data whose phase was simulated. Moreover, there is no open-source implementation for this work to date. In contrast, our network is applied to both single and multi-coil 2D raw data as well as to 3D data without phase information. A perceptual complex neural network was designed in [30] to tackle 2D + time real-time cine MRI in cardiac imaging. However, this architecture is not an unrolled network and was applied to the gridding-reconstructed MR image. The authors of [30] specify that they did not consider using data consistency-based methods because of the increased computation cost of the NUFFT, specifying that efficient tools like `torchkbnufft` [31] would help in this regard. Here, we show that using a data consistency step, a modeling choice allowed by the efficient implementation of the NUFFT, is actually a critical step to obtain improved results compared to baseline neural networks.

d) Neural networks for 3D reconstruction: In [32], the authors introduced a complex unrolled network for multi-coil 3D + time reconstruction with two unrolled iterations for Cartesian acquisitions. However, due to the memory requirements of the NUFFT, it is not clear how to adapt this solution to the non-Cartesian case. Multi-coil 3D MRI reconstruction is also tackled in [33], where the authors design a memory-efficient algorithm to train unrolled networks, however only for Cartesian acquisitions.

[2] `https://github.com/zaccharieramzi/fastmri-reproducible-benchmark`

[3] `https://github.com/zaccharieramzi/tfkbnufft`

III. Model

A. Unrolled network

In this work, we unroll the proximal gradient descent (3) using the alternate scheme proposed in [23] (that is the primal only version of the work).

Algorithm 1: *NC-PDNet*: Density compensated unrolled neural network

Data: $\mathbb{y}$ the k-space data, Ω the non-Cartesian trajectories, $\boldsymbol{d}$ the pre-computed DC weights, $\mathbb{S}$ the coarse estimates of the sensitivity maps

Result: $\boldsymbol{x}$, the reconstructed magnitude image

$\mathbb{S} = g_{\theta_r}(\mathbb{S})$; // Sensitivity maps refinement

Update $\mathcal{A}$ and $\mathcal{A}^H$;

$\boldsymbol{x} = \mathcal{A}^H \mathbb{y}$;

$\boldsymbol{x}_b = [\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{x}, \boldsymbol{x}, \boldsymbol{x}]$; // Buffer creation, in practice concatenation along the channel dimension

for $i \leftarrow 1$ **to** N_C **do**

  $\mathbb{y}_{res} = \mathcal{A}\boldsymbol{x}_b[0] - \mathbb{y}$; // Data consistency

  $\boldsymbol{x}_{dc} = \mathcal{A}^H \boldsymbol{d} \mathbb{y}_{res}$; // Density compensation

  $\boldsymbol{x}_b = \boldsymbol{x}_b + f_{\theta_i}([\boldsymbol{x}_b, \boldsymbol{x}_{dc}]))$; // Proximity operator learning and non-linear acceleration scheme

end

$\boldsymbol{x} = |\boldsymbol{x}_b[0]|$ // Magnitude computation

We describe the algorithm for the forward pass of the *NC-PDNet* in Alg. 1, in the general multi-coil setting, and an illustration is given in Fig.1. The only adjustment to be made for the single-coil case, is that sensitivity maps $\mathbb{S}$ are not involved.

In Alg. 1, each network f_{θ_i} is a convolutional neural network consisting of 3 convolutions applied successively to the input, with a ReLU activation function applied after each of the first two convolutions (as was done in [23]). We specify here that the number of unrolled iterations N_C is fixed, that each network f_{θ_i} has its own set of weights independent of the other f_{θ_j}, and that the whole network is trained end-to-end. This is different from the plug-and-play framework [34] where the number of iterations is not fixed, each sub-network has the same weights and the sub-network is learned offline.

a) Iterations Buffer: Following the idea of [23] (called "memory" in the original paper [35]), we carry a buffer $\boldsymbol{x}_b$ between the different unrolled iterations rather than a single estimate. This buffer of size N_P allows to learn a nonlinear acceleration scheme as explained in [35].

b) Sensitivity maps refinement: g_{θ_r} is a U-net [36] applied uniformly over all the sensitivity maps for sensitivity maps refinement [37]. The coarse estimates of the sensitivity maps $\mathbb{S}$ are computed using the density compensated NUFFT adjoint of the low frequencies of $\mathbb{y}$, rather than using the ESPIRiT method [38]. Indeed ESPIRiT is computationally intensive, and any approximation could be corrected by the network.

c) Complex values: We use real-valued convolutions in the network rather than custom-made complex convolutions [39]. Indeed, real-valued convolutions turn out to have more capacity than complex convolutions when applied to complex tensors handled by concatenating the real and imaginary parts. The concatenation is done on the channel dimension.

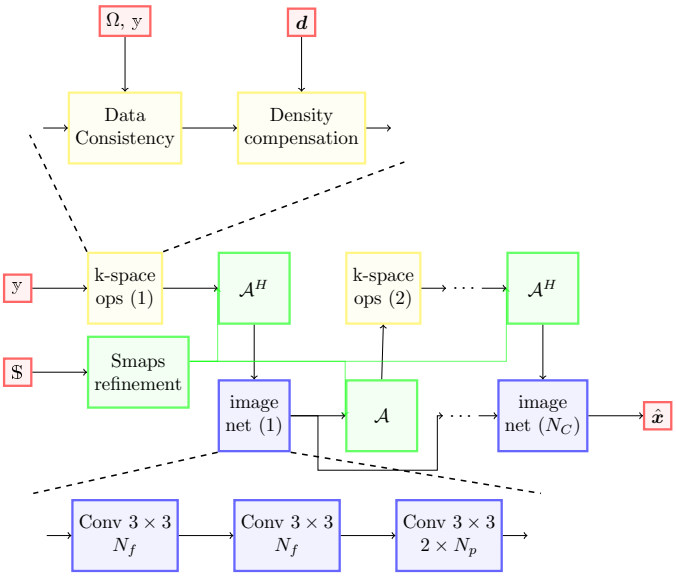

Fig. 1. Illustration of the *NC-PDNet* for multi-coil MRI reconstruction. $\boldsymbol{N_f}$ is the number of filters used for the convolutions, $\boldsymbol{N_P}$ is the buffer size and $\boldsymbol{N_C}$ the number of unrolled iterations.

d) 3D: For the 3D case we replace the 2D convolutions by 3D convolutions with a kernel size of $3 \times 3 \times 3$. Because the activations of the convolutions are much larger in this case, a smaller network is designed to fit into memory when backpropagating. Therefore, we reduce the number of filters for each convolution to $N_f = 16$, and we only consider $N_C = 6$ unrolled iterations.

B. Non-Uniform Fourier Transform

The NDFT is the generalization of the discrete Fourier Transform to positions in the Fourier space that are not necessarily equispaced and on a Cartesian grid. An approximate algorithm to have an efficient computation of the NDFT was introduced in [5], [40]. This algorithm uses an oversampled grid and an optimal interpolation to perform the NDFT efficiently at the cost of an approximation. We refer to this algorithm as the Non-Uniform Fast Fourier Transform (NUFFT), and highlight that unlike the Fast Fourier Transform, it is not an exact algorithm.

While the NUFFT is a more efficient algorithm than the direct application of the definition of the NDFT, it is still a very computationally heavy algorithm compared to the Fourier transform for the same image dimensions. An alternative to the NDFT could be to grid the data in the k-space and simply use the Fast Fourier Transform (FFT) to compute the data consistency. In practice, this would mean using $\mathbb{y}_{grid} = grid(\mathbb{y}, \Omega)$ (where $grid$ is a linear gridding operation) instead of $\mathbb{y}$ and $\mathcal{F}_\Omega$ would simply be the FFT.

a) NDFT vs. NUFFT: We draw the reader attention to the potential confusion between the 2 acronyms. The NDFT is the original transform, while the NUFFT is the approximate fast algorithm to compute the transform.

C. Data Consistency

The data consistency is the step in the unrolled network allowing us to inject the initial measurements $\mathbb{y}$ by comparing them to the current estimate's $\boldsymbol{x}_b[0]$ measurements. The formula we have chosen for data consistency stems from the AWGN model, by taking the gradient of the ℓ_2-norm in (2):

$$\boldsymbol{x}_{dc} = \mathcal{A}^H \boldsymbol{d}(\mathcal{A}\boldsymbol{x}_b[0] - \mathbb{y}) \,. \tag{4}$$

D. Density Compensation

Unlike the Cartesian case, the adjoint operator of the NDFT is not always its inverse operator. Worse, in most cases, the NDFT does not admit an inverse operator. The application of the adjoint operator to the k-space can therefore be very far from the solution to the inverse problem (1).

To circumvent this, DC has been introduced [20]. Indeed, the main problem with the classical non-Cartesian trajectories like radial or spiral, is that they densely sample the center of the k-space. Therefore, when computing the adjoint, a lot of weight is assigned to the densely sampled region at the center of the k-space, resulting in an image with abnormally large values. DC is just the action of using factors that weigh the different sample locations so that they all play an even role during the application of the adjoint.

It is particularly needed for deep learning approaches because the values entering convolutional layers need to have normalized (or close to normalized) values in order to avoid numerical issues.

In practice, for both the radial and the spiral trajectories, we obtain the DC factors $\boldsymbol{d}$ by applying the adjoint and forward operators[4] iteratively for N iterations, starting from ones:

$$\begin{aligned} \boldsymbol{d}_0 &= \mathbf{1} \\ \boldsymbol{d}_{n+1} &= \frac{\boldsymbol{d}_n}{\mathcal{F}_\Omega \mathcal{F}_\Omega^H \boldsymbol{d}_n} \end{aligned} \tag{5}$$

where the division is here pointwise. The final weights are $\boldsymbol{d}_N$. In practice, we took $N = 10$.

IV. Data

We used the *NC-PDNet* on different data settings. In all these settings, the data was retrospectively under-sampled using fixed trajectories. The 2 trajectories we studied in this work were spiral and radial trajectories, with an acceleration factor of 4 compared to the full Cartesian acquisition.

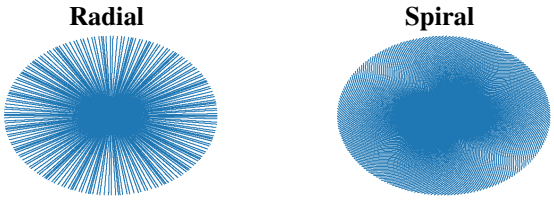

Fig. 2. The 2 k-space trajectories considered in this work. Each of them uses 100 spokes and has a total of 64k measurements.

For the 3D case, we simply stacked these trajectories along the additional dimension. An illustration of the stacking mechanism is shown in Supplementary Material in Fig. S1.

[4]In practice we use the interpolation operator of the NUFFT rather than the whole operator.

A. fastMRI

For the 2D single-coil and multi-coil cases, we used the raw knee data from the fastMRI dataset [41]. The acquisition was performed using a Cartesian 2D Turbo Spin Echo (TSE) sequence and the data was collected over a 15-channel phased array coil. The imaging contrasts were proton-density weighting, half with fat suppression, half without. The sequence parameters were defined as follows: Echo train length 4, matrix size 320×320, in-plane resolution 0.5mm$\times$0.5mm, slice thickness 3mm, no gap between slices. In total, there were 973 volumes (34,742 slices) for the training subset and 199 volumes (7135 slices) for the validation subset. In this study, we chose to jointly consider the two weighting contrasts (Proton Density and Proton Density with Fat Suppression) available present in the fastMRI dataset. The single-coil data was emulated from the multi-coil data, meaning that it was not actually acquired with a single receiver coil. Instead, the single-coil data is the result of a linear combination of multi-coil data in such a way that the reconstructions using root sum-of-squares match [41]. To obtain the non-Cartesian measurements $\mathbb{y}$, we simply used the inverse Fourier transform of the full Cartesian k-space data, denoted $\boldsymbol{y}_{or,\ell}$ for each coil ℓ separately. Formally, for a given sampling pattern Ω:

$$\boldsymbol{y}_\ell = \mathcal{F}_\Omega \boldsymbol{F}^T \boldsymbol{y}_{or,\ell} \quad \forall \ell = 1, \dots, L \tag{6}$$

We did not use the fastMRI data to perform the 3D experiments for two reasons:

- The data in the fastMRI dataset has a limited number of slices (roughly 30 to 40 for each acquisition), and the 3D networks will therefore not be able to fully take advantage of the third dimension.
- The data in the fastMRI dataset was collected in 2D using an anisotropic resolution to maintain a good SNR (in-plane resolution 0.5mm$\times$0.5mm, slice thickness 3mm).

B. OASIS

The Open Access Series of Imaging Studies (OASIS) brain database [42] is a database including MRI scans of 1068 participants, yielding 2168 MR sessions (3400 volumes). Of these 2168, we selected 2164 sessions which feature T1-weighted sequences. 1878 of these were acquired on a 3T scanner, 236 at 1.5T and the remaining are undisclosed (50). We decided to use the OASIS dataset with data recast to matrix size of $176 \times 256 \times 256$, using zero-padding. While the data from the fastMRI dataset is the raw k-space measurements, the data available in the OASIS dataset is the DICOM magnitude image. Therefore the NDFT was applied to the magnitude volume and we did not simulate the phase. The data was separated in a training and a validation set. The split was participant-based, that is a participant cannot have a scan in both sets. The split was of 90% for the training set and 10% for the validation set.

V. Results

A. Experimental setup

a) Metrics: The metrics used for evaluating the networks are peak signal-to-noise ratio (PSNR) and structural similarity

index (SSIM) [43]. They are computed over whole volumes, that is the dynamic range (used for the computation of PSNR and SSIM) is computed over the whole volume. The parameters and definitions used are the same as in the fastMRI paper [41].

b) Comparison: To illustrate the need for all the key components of the *NC-PDNet*, we carried out an ablation study against other neural networks where we removed some aspects of the *NC-PDNet*.

- An unrolled network without DC but only a normalization mechanism (without any normalization the network does not train due to very high values), to show the need for the DC.
- An unrolled network without NUFFT and instead a gridded version of the k-space (the operator is only the much faster FFT), to show the need for a better approximation of the NDFT.
- A U-net [36] applied directly to the density-compensated adjoint of the k-space measurements, to show the need for the unrolling framework (3D convolutions and pooling operations are used in the 3D setting).
- The density-compensated adjoint of the k-space measurements as very naive baseline.

For the 2D settings, the unrolled networks have $N_C = 10$ unrolled iterations, use a buffer size $N_P = 5$ and a number of convolution filters $N_f = 32$. For the 3D setting, we needed to reduce those numbers for the network to fit on a single GPU, and used $N_C = 6$ unrolled iterations, a buffer size $N_P = 2$, and $N_f = 16$. The U-net was trained residually and had a base number of filters of 16.

We carried out the training and evaluation of the variants of the unrolled networks only in the 2D single-coil setting.

c) Generalizability: We also studied the generalizability of the trained networks to other settings. First, we studied the generalizability of the networks to different sampling trajectories. In practice, we evaluated the networks trained on the spiral trajectory in the 2D multi-coil setting on the radial trajectory, and vice-versa. We term this the *reverse setting*.

Second, we studied the generalizability of the networks to different organs and contrasts simultaneously. We took advantage of the 2D multi-coil brain dataset available in the fastMRI dataset to carry out this analysis. This dataset features brain data acquired with four different weighting contrasts (T1, T2, FLAIR and T1 after Gadolinium injection), collected on the same scanners, which allows to test for generalizability solely on organ/contrast.

d) Computational efficiency: The time spent for image reconstruction by a given method is also an important factor to take into account when choosing which algorithm to be used in the clinical realm. Indeed, in clinical applications we need to visualize the reconstructed MR images prior to the end of the exam to permit any potential re-start of the pulse MR sequence in case of the presence of significant artifacts impeding accurate image-based diagnosis. Additionally, this allows the physician to report on the scans directly to the patient once the exam is completed. For that reason, we decided to report the reconstruction times of the different methods and networks.

TABLE I

MEAN PSNR / SSIM ON THE VALIDATION VOLUMES OF THE DIFFERENT APPROACHES FOR BOTH FASTMRI CONTRASTS (KNEE DATASET) IN THE SINGLE-COIL SETTING. THE BEST RESULTS ARE IN BOLD FONT.

Model	Radial	Spiral	# Parameters
PDNet no DC	27.02 / 0.6747	28.02 / 0.6946	156k
Adjoint + DC	27.11 / 0.6471	31.70 / 0.7213	0
PDNet w gridding	31.12 / 0.6887	31.45 / 0.7126	156k
U-net on Adjoint + DC	32.26 / 0.7224	32.82 / 0.7460	481k
NC-PDNet	**32.66 / 0.7327**	**33.08 / 0.7534**	156k

TABLE II

MEAN PSNR / SSIM ON THE VALIDATION VOLUMES OF THE DIFFERENT APPROACHES FOR BOTH FASTMRI CONTRASTS (KNEE DATASET) IN THE MULTI-COIL SETTING. THE BEST RESULTS ARE IN BOLD FONT.

Model	Radial	Spiral	# Parameters
Adjoint + DC	25.91 / 0.6486	31.36 / 0.7197	0
U-net on Adjoint + DC	38.78 / 0.9106	40.02 / 0.9215	481k
NC-PDNet	**40.00 / 0.9191**	**40.68 / 0.9255**	163k

B. Quantitative results

a) 2D single-coil: The quantitative results in Tab. I show that our approach (*NC-PDNet*) outperforms the others. Moreover, these results show that to obtain accurate reconstruction results for data collected along radial and spiral trajectories, combining unrolling and DC, with an accurate NDFT approximation, is instrumental, as using only one of these three ingredients will lead to degraded performance. Moreover, it is worth noting that the quantitative results are slightly improved for spiral trajectory likely because it has a better coverage of the high frequencies in the k-space.

b) 2D multi-coil: For the experiments in the 2D multi-coil setting, we selected only the two best performers of the single-coil experiments for comparison purposes, along with the adjoint baseline for reference. We see in Tab. II that because the adjoint with DC has a very low performance for the radial trajectory, the U-net is not able to correct sufficiently well for aliasing artifacts in the image domain to obtain competitive results with the *NC-PDNet*. This shows that the *NC-PDNet* has the ability to overcome situations where the trajectory might be causing issues to naive methods. We also observed that the gap in spiral imaging is larger in the multi-coil setting than in the the single-coil setting for the *NC-PDNet*. Finally, it can be noted that the sensitivity maps refinement has a relatively low impact on the number of parameters in the *NC-PDNet*, with only 7k additional parameters in θ_r (163k vs 156k parameters in bottom rows of Tab. II vs Tab. I).

c) 3D single-coil: For the 3D setting, we used the same U-net architecture, replacing the 2D convolutions by 3D

TABLE III

MEAN PSNR ON THE VALIDATION VOLUMES OF THE DIFFERENT APPROACHES FOR THE OASIS DATASET (3D SETTING). THE BEST RESULTS ARE IN BOLD FONT.

Model	Radial stacks	Spiral stacks	# Parameters
Adjoint + DC	25.43	30.35	0
U-net on Adjoint + DC	**31.14**	**31.20**	1.6M
NC-PDNet	30.14	30.55	67k

TABLE IV

MEAN PSNR/SSIM ON THE VALIDATION VOLUMES OF THE DIFFERENT APPROACHES FOR BOTH FASTMRI CONTRASTS (KNEE DATASET) IN THE MULTI-COIL REVERSE SETTING. THE BEST RESULTS ARE IN BOLD FONT.

Model	Radial	Spiral	# Parameters
Adjoint + DC	25.91 / 0.6486	31.36 / 0.7197	0
U-net on Adjoint + DC (trained on different trajectory)	27.94 / 0.87	26.35 / 0.8850	481k
***NC-PDNet* (trained on different trajectory)**	**37.86 / 0.9079**	**36.28 / 0.9052**	163k

TABLE V

MEAN PSNR/SSIM ON THE VALIDATION VOLUMES OF THE DIFFERENT APPROACHES FOR ALL BRAIN FASTMRI IMAGING CONTRASTS IN THE MULTICOIL SETTING. THE BEST RESULTS ARE IN BOLD FONT.

Model	Radial	Spiral	# Parameters
Adjoint + DC	27.31 / 0.6028	32.23 / 0.6603	0
U-net on Adjoint + DC (trained on knee data)	37.88 / 0.9234	38.76 / 0.9302	481k
***NC-PDNet* (trained on knee data)**	**39.48 / 0.9368**	**39.81 / 0.9390**	163k

convolutions. Due to this, and the fact that we reduced the size of the *NC-PDNet*, the size ratio between the the U-net and the *NC-PDNet* is significantly increased. As a result, the U-net clearly outperforms the *NC-PDNet* in the 3D setting. More importantly, we see that the *NC-PDNet* is only slightly better than the application of the adjoint with DC in the spiral case, meaning that the capacity of the network is critical for this application.

d) Reverse Trajectories: The quantitative results for the reverse setting shown in Tab. IV reveal contrasting results. The drop in PSNR is significant for both networks (below the adjoint level for the U-net trained on radial and evaluated on spiral), while the drop in SSIM is very much controlled in the *NC-PDNet*'s case (a bit less so for the U-net). It is worth mentioning that while the drop in PSNR is significant for the *NC-PDNet*, the network is still competitive with the regular setting's networks in terms of PSNR values (37.86dB vs 40dB in the radial validation case) while the U-net is not (27.94dB vs 38.78dB in the radial validation case).

e) 2D Brain Data: The quantitative results for the brain data are not to be compared head-to-head to the results obtained for the knee data. Indeed there is a significant mismatch in reconstruction difficulty, the brain data being easier to reconstruct than the knee data metrics-wise. However, we can see that the networks still largely outperform the baseline for both trajectories. We also see that the *NC-PDNet* with DC outperforms the U-net on both trajectories, this time with a bigger gap, suggesting that it generalizes in a better way.

f) Reconstruction times: The reconstruction times given in Tab. VI show that the use of the NUFFT in the unrolled setting induces a 4.5-fold increase in reconstruction time for the 2D setting and an 8-fold increase for the 3D setting[5]. However, the use of the DC merely increases the reconstruction time by only 33ms.

[5]The reconstruction times observed for 3D are much less than those in 2D because of the different image sizes as well as sizes of the networks.

TABLE VI

RECONSTRUCTION TIMES IN MILLISECONDS FOR THE DIFFERENT NETWORKS IN THE DIFFERENT SETTINGS FOR A SINGLE SLICE FOR THE RADIAL (OR RADIAL STACKS) TRAJECTORY.

Model	Single-coil 2D	Multi-coil 2D	3D
PDNet no DC	446	NA	NA
Adjoint + DC	101	135	7
PDNet w gridding	112	NA	NA
U-net on Adjoint + DC	110	145	9
NC-PDNet	479	661	80

C. Qualitative results

For all qualitative results, the images were selected at random, and do not necessarily have the same contrast. For each setting, the missing acquisition trajectory's reconstruction figures (if any) can be found in the Supplementary Material in Figs. S2-S5. The residual images are represented in absolute with a scale relative to each residual to highlight better the errors specific to one reconstruction. A shared scale would result in very uninformative residuals as the adjoint sometimes has a very high error.

a) 2D single-coil imaging: The visual inspection of the reconstructed MR images confirm the quantitative measurements. In particular, one can visualize in Fig. 3 that the image's inner contrast is better recovered by the *NC-PDNet* and U-net, and the structures are sharper for both acquisitions. Finally, the residual map is slightly flatter for the *NC-PDNet* compared to the U-net, especially around structures for the radial acquisition.

b) 2D multi-coil imaging: The qualitative results for the multi-coil setting, available in Fig. 4 confirm the quantitative results. We see on the right part of the knee that a clear artifact is left unaltered by the U-net after the application of the adjoint with DC. This artifact is not present anymore in the image reconstructed by *NC-PDNet*. We notice however that the application of U-net has allowed us to flatten the error map and recover the inner contrast of the image.

c) 3D imaging: The qualitative results for the 3D reconstruction strikingly contrast with the quantitative results. For the spiral acquisition in particular, we see in Fig. 5 that the *NC-PDNet* performs much better compared to the adjoint with DC than what was expected from the quantitative results. Most of the blur observed in the adjoint reconstruction is absent in the unrolled reconstruction, although it is still blurry. This blurring effect might not be well caught by the PSNR metric which is known to be oblivious to some deformations [44]. We also see in Fig. 5 that the *NC-PDNet* erases less details in the reconstruction than the U-net. Much of the quantitative error in this case is due to the loss in the image inner-contrast.

d) Reverse Trajectories: The qualitative results for the reverse setting, shown in Figs. 6-7, allow us to look at the quantitative results from a different perspective. We actually observed that the reconstructions are excellent in terms of image quality and that the degradation compared to the regular setting is minimally visible (except the radial artifacts for the *NC-PDNet* trained on spiral acquisition and evaluated on radial spokes in the right part of the knee). This confirms that the SSIM is the best metric to monitor in order for measuring

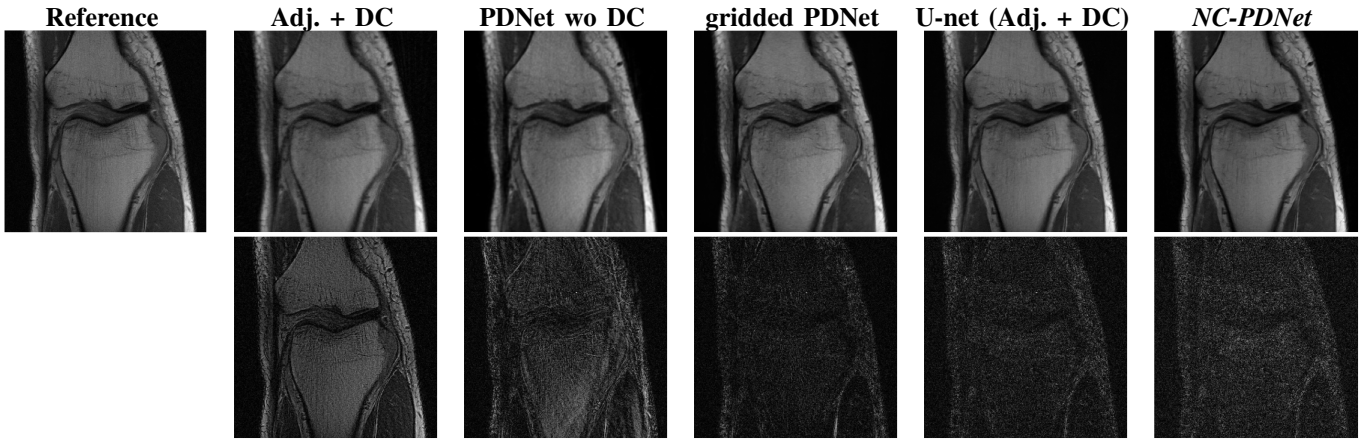

Fig. 3. **2D single-coil radial acquisition (knee fast MRI dataset)**: Reconstruction results for a specific slice (16th slice of `file1001184`, part of the validation set). The top row represents the reconstruction using the different methods, while the bottom one represents the absolute error when compared with the reference.

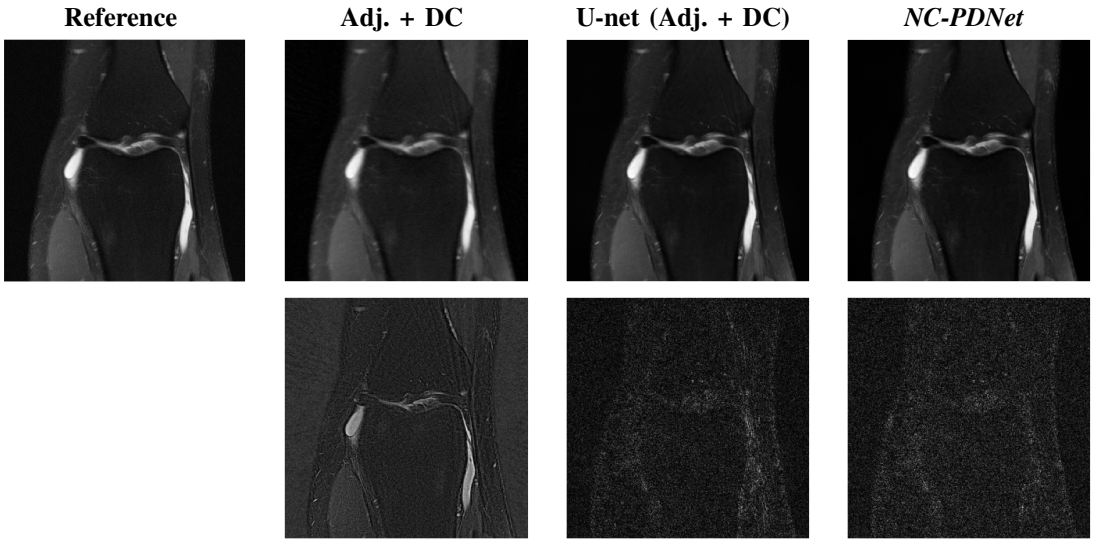

Fig. 4. **2D multi-coil radial acquisition (knee fast MRI dataset)**: Reconstruction results for a specific slice (16th slice of `file1000000`, part of the validation set). The top row represents the reconstruction using the different methods, while the bottom one represents the absolute error when compared with the reference.

success in generalization to other trajectories.

e) Brain Data: The qualitative results for the brain data confirm that the *NC-PDNet* was able to better generalize than the U-net. In particular, for the spiral trajectory in Fig. 8, we can see in the error map that the inter-hemispheric brain region was not reconstructed well by the U-net while the *NC-PDNet* managed to.

VI. Discussion and conclusion

In this work, we demonstrated how to apply the framework of unrolled networks to the problem of non-Cartesian MRI reconstruction. In particular, we showed via an ablation study, the importance of using the mechanism of DC [20] for this setting and how it is instrumental in obtaining the best possible results in a Deep Learning framework. We also managed to show how this mechanism can be extended to challenging acquisition scenarios like multi-coil and 3D reconstruction with very minor adjustments. We conducted additional generalizability experiments to measure the robustness of the trained networks to out-of-distribution settings, and concluded that the visual degradation is minimal for all networks, and that additionally the *NC-PDNet* is quantitatively consistent. In particular, we found out that the generalization to other organs is possible even when using a single organ in training. Finally the modular design of the code allows implementation and improvement of the results.

Some issues still remain open regarding this work. The first is, of course, that of transferability/generalizability. Indeed, these networks have been trained on a specific dataset with given experimental conditions, a given MR systems vendor (Siemens-Healthineers, Erlangen, Germany) and a specific acquisition setup and organ. The question of generalizability from one vendor to another has been partially answered in [12], where it has been observed that the results degrade when the networks trained with one vendor are applied to another. Although we partially addressed the concerns regarding the acquisition setup and the organs with our out-of-distribution

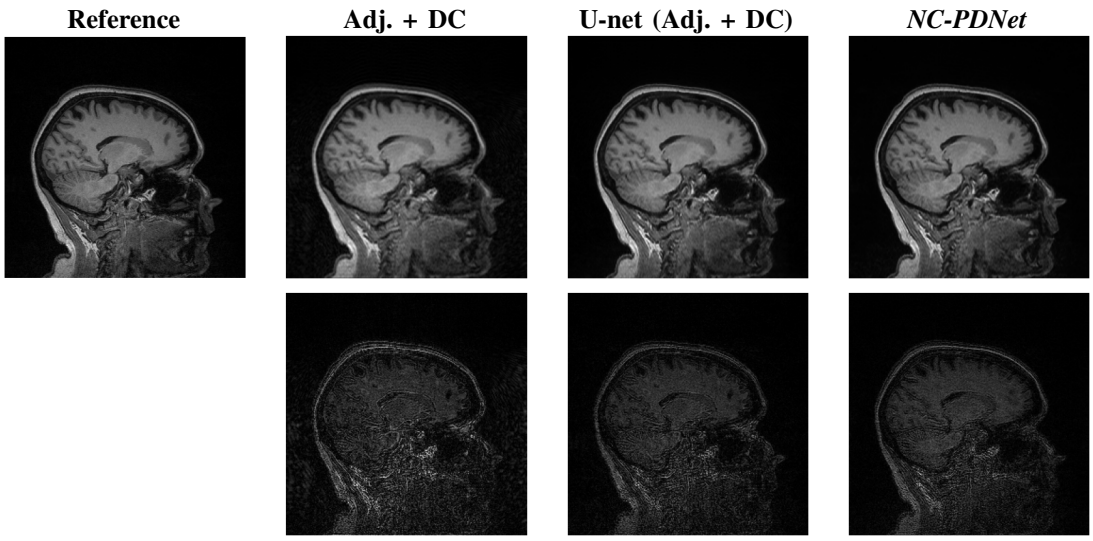

Fig. 5. **3D stack-of-spirals acquisition (OASIS dataset, T1 contrast)**: Reconstruction results for a specific slice (101st slice of `sub-OAS30001_ses-d0129_run-01_T1w`, part of the validation set). The top row represents the reconstruction using the different methods, while the bottom one represents the absolute error when compared with the reference.

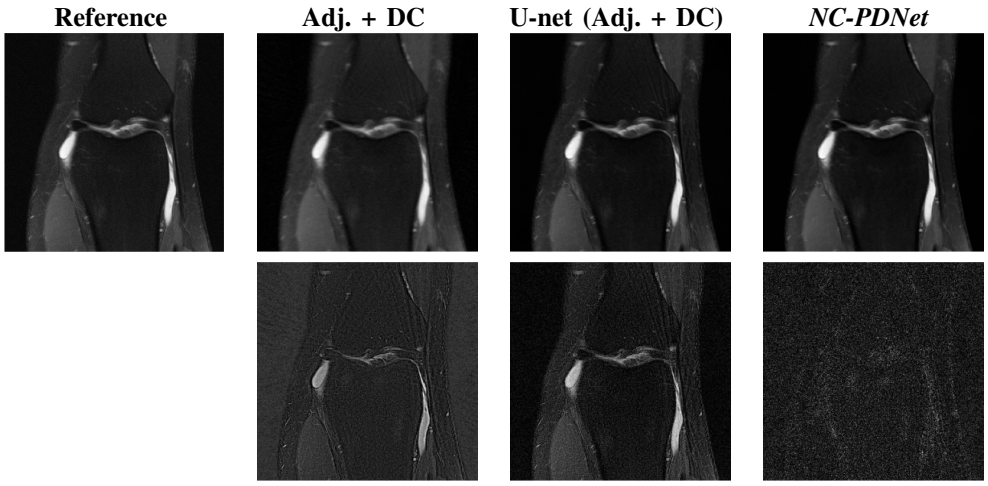

Fig. 6. **2D multi-coil radial acquisition (reverse setting, knee dataset)**: Reconstruction results for a specific slice (16th slice of `file1000000`, part of the validation set) with networks trained with spiral trajectories. The top row represents the reconstruction using the different methods, while the bottom one represents the absolute error when compared with the reference.

experiments, a more systematic study is still required. Another point that should be further investigated is the generalizability to different acceleration factors. Indeed this work focused on a fixed 4-fold acceleration factor, but in clinical use, acceleration factors may vary widely from one exam to the other depending on time constraints.

A more challenging question that can be raised is whether a lack of generalizability is intrinsic to the training process or not: The need for training data generally translates to using retrospective under-sampling to be able to learn how to correct for aliasing artifacts from a "correct" ground truth. How do the networks generalize to prospective under-sampling? This question is very difficult to answer systematically. In the case of prospective acquisition, as no ground truth is available, the methods cannot be compared one another. The only viable option is to have qualitative results, such as those available in [45]. A final issue on transferability arises in this work for the 3D data, which consisted of magnitude volumes (and not of raw k-space data as in the 2D case). How can the networks generalize to complex-valued data? There exists ways to generate an artificial phase information, and they could be of help in order to be closer to the real use-case [46].

One disadvantage of the current implementation of unrolled networks in the non-Cartesian setting is their slowness compared to other baseline networks. This has been diagnosed in this work to be due to the NUFFT. Recent work has been carried out to implement an efficient NUFFT on the GPU, cufiNUFFT [47], although not directly in a framework allowing auto-differentiation, making its use at training time difficult. However, fast implementations could very well be embedded in the network at reconstruction allowing to alleviate the slowness of NUFFT in auto-differentiation frameworks.

Finally, the 3D setting illustrated a major blocker on the road ahead for unrolled networks: Memory requirements for the

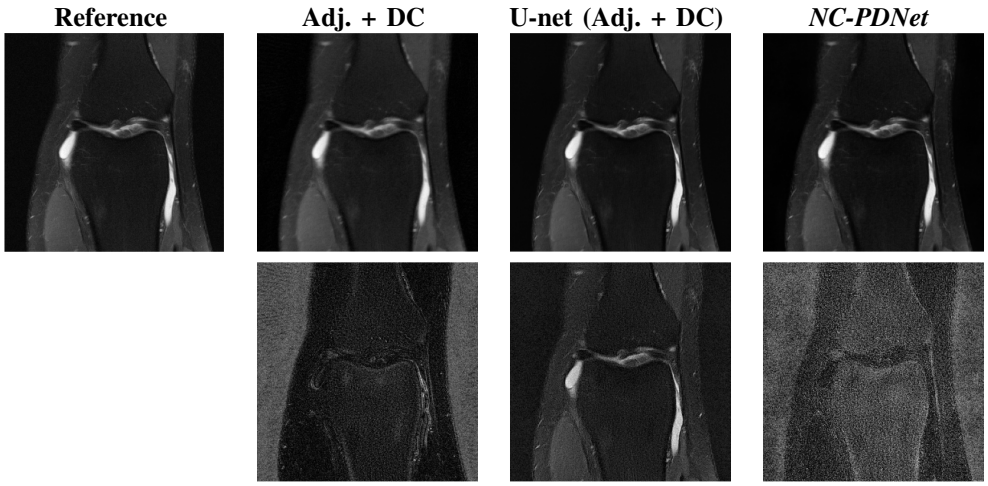

Fig. 7. **2D multi-coil spiral acquisition (reverse setting, knee dataset)**: Reconstruction results for a specific slice (16th slice of `file1000000`, part of the validation set) with networks trained on radial trajectories. The top row represents the reconstruction using the different methods, while the bottom one represents the absolute error when compared with the reference.

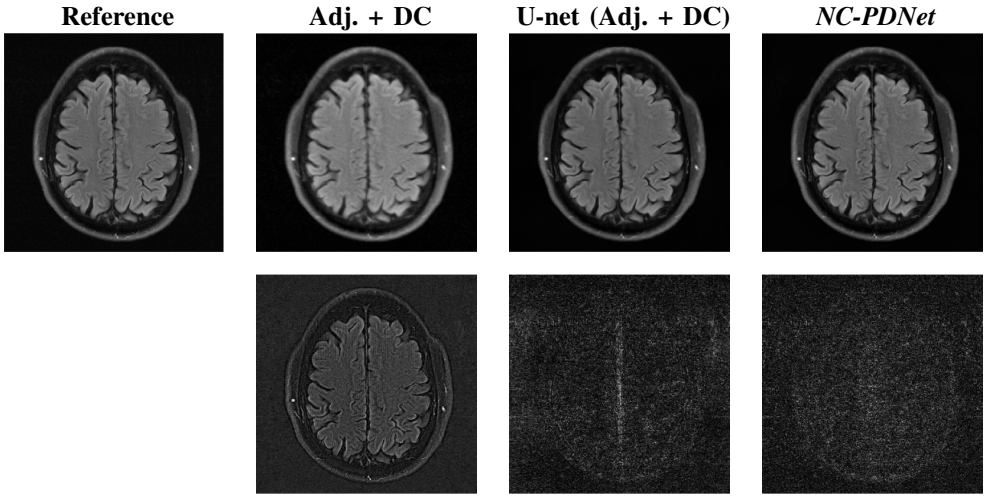

Fig. 8. **2D multi-coil radial acquisition**: Reconstruction results for a specific slice (6th slice of `file_brain_AXFLAIR_200_6002447`) from the brain fastMRI dataset with networks trained on knee data. The top row represents the reconstruction using the different methods, while the bottom one represents the absolute error when compared with the reference.

backpropagation during training. Indeed, none of the networks seemed to overfit, implying that there is room for networks with larger capacity. However, training such networks in the current state with a basic implementation on regular 32GB GPUs is out-of-reach. Some possible workarounds of this problem exist on the implementation side, using multiple GPUs, like model parallelism [48] or pipelining [49]. Some other solutions exist on the modeling side, by using networks that either reduce the memory requirements of the backpropagation like gradient checkpointing [50], invertible networks [33] or by getting rid of the backpropagation altogether like implicit deep learning [51], [52].

Acknowledgements

We acknowledge the financial support of the Cross-Disciplinary Program on Numerical Simulation of CEA for the project entitled SILICOSMIC. We also acknowledge the French Institute of development and resources in scientific computing (IDRIS) for their AI program allowing us to use the Jean Zay supercomputer's GPU partitions.

References

[1] Z. Ramzi, P. Ciuciu, and J.-L. Starck, "Density Compensated Unrolled Networks for Non-Cartesian MRI Reconstruction," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021.

[2] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 11 1999.

[3] M. A. Griswold, P. M. Jakob *et al.*, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 6 2002.

[4] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging Michael," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[5] J. A. Fessler and B. P. Sutton, "Nonuniform Fast Fourier Transforms Using Min-Max Interpolation," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 560–74, 2003.

[6] J. Caballero, A. N. Price, D. Rueckert, and J. V. Hajnal, "Dictionary learning and time sparsity for dynamic MR data reconstruction," *IEEE Transactions on Medical Imaging*, vol. 33, no. 4, pp. 979–994, 2014.

[7] K. Gregor and Y. Lecun, "Learning Fast Approximations of Sparse Coding," in *Proceedings of the 27th International Conference on Machine Learning*, 2010.

[8] B. Zhu, J. Z. Liu *et al.*, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 3 2018.

[9] T. Eo, Y. Jun *et al.*, "KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images," *Magnetic Resonance in Medicine*, vol. 80, no. 5, pp. 2188–2201, 2018.

[10] K. Hammernik, T. Klatzer *et al.*, "Learning a Variational Network for Reconstruction of Accelerated MRI Data," *Magnetic Resonance in Medicine*, vol. 3071, no. 79, pp. 3055–3071, 2017.

[11] J. Schlemper, J. Caballero *et al.*, "A Deep Cascade of Convolutional Neural Networks for MR Image Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, 2018.

[12] M. J. Muckley, B. Riemenschneider *et al.*, "State-of-the-art Machine Learning MRI Reconstruction in 2020: Results of the Second fastMRI Challenge," Tech. Rep., 2020. [Online]. Available: http://arxiv.org/abs/2012.06318

[13] M. A. Bernstein, K. F. King, and X. J. Zhou, "Advanced pulse sequence techniques," in *Handbook of MRI Pulse Sequences*. Elsevier, 2004, pp. 802–954.

[14] K. T. Block, H. Chandarana *et al.*, "Towards Routine Clinical Use of Radial Stack-of-Stars 3D Gradient-Echo Sequences for Reducing Motion Sensitivity," *Journal of the Korean Society of Magnetic Resonance in Medicine*, vol. 18, no. 2, p. 87, 2014.

[15] C. Lazarus, P. Weiss *et al.*, "3D variable-density SPARKLING trajectories for high-resolution T2*-weighted magnetic resonance imaging," *NMR in Biomedicine*, vol. 33, no. 9, p. e4349, 2020.

[16] C. G R, P. Weiss *et al.*, "Globally optimized 3D SPARKLING trajectories for high-resolution T2*-weighted Magnetic Resonance Imaging," Dec. 2020, working paper or preprint. [Online]. Available: https://hal.inria.fr/hal-03090471

[17] H. K. Song and L. Dougherty, "Dynamic MRI with projection reconstruction and KWIC processing for simultaneous high spatial and temporal resolution," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 52, no. 4, pp. 815–824, 2004.

[18] D. R. Thedens, P. Irarrazaval *et al.*, "Fast magnetic resonance coronary angiography with a three-dimensional stack of spirals trajectory," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 41, no. 6, pp. 1170–1179, 1999.

[19] M. Abadi, A. Agarwal *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," Tech. Rep., 2016. [Online]. Available: http://arxiv.org/abs/1603.04467

[20] J. G. Pipe and P. Menon, "Sampling density compensation in MRI: Rationale and an iterative numerical solution," *Magnetic Resonance in Medicine*, vol. 41, no. 1, pp. 179–186, 1999.

[21] F. Ong, M. Uecker, and M. Lustig, "Accelerating Non-Cartesian MRI Reconstruction Convergence Using k-Space Preconditioning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1646–1654, 2020.

[22] J. D. Trzasko, A. Manduca *et al.*, "A Preconditioned ADMM Strategy for Field-Corrected Non-Cartesian MRI Reconstruction," in *ISMRM*, vol. 22, 2014, p. 1535.

[23] J. Adler and O. Öktem, "Learned Primal-Dual Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.

[24] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model based deep learning architecture for inverse problems," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 394–405, 2019.

[25] J. Schlemper, S. Sadegh *et al.*, "Nonuniform Variational Network: Deep Learning for Accelerated Nonuniform MR Image Reconstruction," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 57–64.

[26] F. Knoll, T. Murrell *et al.*, "Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge," *Magnetic Resonance in Medicine*, vol. 84, no. 6, pp. 3054–3070, 2020.

[27] Y. Yang, J. Sun *et al.*, "Deep ADMM-net for compressive sensing MRI," in *Neural Information Processing Systems*, 2016, pp. 10–18.

[28] D. Gilton, G. Ongie, and R. Willett, "Neumann Networks for Linear Inverse Problems in Imaging," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 328–343, 2019.

[29] Z. Ramzi, J.-L. Starck, and P. Ciuciu, "XPDNet for MRI Reconstruction: an application to the 2020 fastMRI challenge," in *ISMRM*, 2021.

[30] D. Shen, S. Ghosh *et al.*, "Rapid reconstruction of highly undersampled, non-Cartesian real-time cine k-space data using a perceptual complex neural network (PCNN)," *NMR in Biomedicine*, vol. 34, no. 1, pp. 1–12, 2021.

[31] M. J. Muckley, R. Stern, T. Murrell, and F. Knoll, "TorchKbNufft: A high-level, hardware-agnostic non-uniform fast fourier transform," in *ISMRM Workshop on Data Sampling & Image Reconstruction*, 2020.

[32] T. Küstner, N. Fuin *et al.*, "CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions," *Scientific Reports*, vol. 10, no. 1, pp. 1–13, 2020.

[33] M. Kellman, K. Zhang *et al.*, "Memory-efficient learning for large-scale computational imaging," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1403–1414, 2020.

[34] T. Meinhardt, M. Moeller, and D. Cremers, "Learning Proximal Operators : Using Denoising Networks for Regularizing Inverse Imaging Problems," in *ICCV*, 2017, pp. 1781–1790.

[35] J. Adler and O. Oktem, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Problems*, vol. 33, no. 12, pp. 1–24, 2017.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.

[37] A. Sriram, J. Zbontar *et al.*, "End-to-End Variational Networks for Accelerated MRI Reconstruction," in *MICCAI*, 2020.

[38] M. Uecker, P. Lai *et al.*, "ESPIRiT-An Eigenvalue Approach to Autocalibrating Parallel MRI: Where SENSE meets GRAPPA HHS Public Access," *Magn Reson Med*, vol. 71, no. 3, pp. 990–1001, 2014.

[39] E. K. Cole, J. Y. Cheng, J. M. Pauly, and S. S. Vasanawala, "Complex-Valued Convolutional Neural Networks for MRI Reconstruction," Tech. Rep. [Online]. Available: https://arxiv.org/abs/2004.01738

[40] P. J. Beatty, D. G. Nishimura, and J. M. Pauly, "Rapid gridding reconstruction with a minimal oversampling ratio," *IEEE Transactions on Medical Imaging*, vol. 24, no. 6, pp. 799–808, 2005.

[41] J. Zbontar, F. Knoll *et al.*, "fastMRI: An Open Dataset and Benchmarks for Accelerated MRI," Tech. Rep., 2018. [Online]. Available: https://arxiv.org/pdf/1811.08839.pdf

[42] P. J. LaMontagne, T. L. S. Benzinger *et al.*, "OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease," Tech. Rep., 2019. [Online]. Available: https://www.medrxiv.org/content/10.1101/2019.12.13.19014902v1

[43] Z. Wang, A. C. Bovik, H. Rahim Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, 2004.

[44] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proceedings - International Conference on Pattern Recognition*, 2010, pp. 2366–2369.

[45] Z. Ramzi, A. Vignaud, J.-L. Starck, and P. Ciuciu, "Is good old GRAPPA dead?" in *ISMRM*, 2021.

[46] J. Schlemper, M. Salehi, and M. Sofka, "Mutli-coil Magnetic Resonance Imaging Using Deep Learning," 2020. [Online]. Available: https://patents.google.com/patent/US20200294287A1/en

[47] Y.-h. Shih, J. Blaschke, and A. H. Barnett, "cuFINUFFT : a load-balanced GPU library for general-purpose nonuniform FFTs," Tech. Rep., 2021. [Online]. Available: https://arxiv.org/abs/2102.08463

[48] N. Shazeer, Y. Cheng *et al.*, "Mesh-tensorflow: Deep learning for supercomputers," in *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[49] Y. Huang, Y. Cheng *et al.*, "GPipe: Efficient training of giant neural networks using pipeline parallelism," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[50] T. Chen, B. Xu, C. Zhang, and C. Guestrin, "Training Deep Nets with Sublinear Memory Cost," Tech. Rep., 2016. [Online]. Available: http://arxiv.org/abs/1604.06174

[51] S. Bai, J. Zico Kolter, and V. Koltun, "Deep equilibrium models," in *NeurIPS*, 2019.

[52] D. Gilton, G. Ongie, and R. Willett, "Deep Equilibrium Architectures for Inverse Problems in Imaging," Tech. Rep., 2021. [Online]. Available: http://arxiv.org/abs/2102.07944