Computationally Efficient IMplicit Training Strategy for UNrolled NEtworks (IMUNNE): A preliminary analysis using accelerated real-time cardiac cine MRI

Nikolay Iakovlev, Florian A. Schiffers, Santiago L. Tapia, Daming Shen, KyungPyo Hong, Michael Markl, Daniel C. Lee, Aggelos K. Katsaggelos, *Life Fellow, IEEE,* and Daniel Kim

Abstract*—*Objective: **Highly-undersampled, dynamic MRI reconstruction, particularly in multi-coil scenarios, is a challenging inverse problem. Unrolled networks achieve state-of-the-art performance in MRI reconstruction but suffer from long training times and extensive GPU memory cost.** ***Methods:*** **In this work, we propose a novel training strategy for IMplicit UNrolled NEtworks (IMUNNE) for highly-undersampled, multi-coil dynamic MRI reconstruction. It formulates the MRI reconstruction problem as an implicit fixed-point equation and leverages gradient approximation for backpropagation, enabling training of deep architectures with fixed memory cost. This study represents the first application of implicit network theory in the context of real-time cine MRI. The proposed method is evaluated using a prospectively undersampled, real-time cine dataset using radial k-space sampling, comprising balanced steady-state free precession (b-SSFP) readouts. Experiments include a hyperparameter search, head-to-head comparisons with a complex U-Net (CU-Net) and an alternating unrolled network (Alt-UN), and an analysis of robustness under noise perturbations; peak signal-to-noise ratio, structural similarity index, normalized root mean-square error, spatio-temporal entropic difference, and a blur metric were used.** ***Results:*** **IMUNNE produced significantly and slightly better image quality compared to CU-Net and Alt-UN, respectively. Compared with Alt-UN, IMUNNE significantly reduced training and inference times, making it a promising approach for highly-accelerated, multi-coil real-time cine MRI reconstruction.** ***Conclusion:*** **IMUNNE strategy successfully applies unrolled networks to image reconstruction of highly-accelerated, real-time radial cine MRI.** ***Significance:*** **Implicit training enables rapid, high-quality, and cost-effective CMR exams by reducing training and inference times and lowering memory cost associated with advanced reconstruction methods.**

***Index Terms*—Compressed sensing, deep learning, image reconstruction, implicit network, unrolled network**

The authors gratefully acknowledge funding support from the National Institutes of Health (R01HL116895, R01HL151079, R01HL168859, 1R01HL167148-01A1), the Radiological Society of North America (EILTC2302), and the American Heart Association (19IPLOI34760317, 949899).

Nikolay Iakovlev is with the Department of Radiology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611 USA (e-mail: nikolay.iakovlev@northwestern.edu).

Daniel Kim, KyungPyo Hong, and Michael Markl are with the Department of Radiology, Northwestern University Feinberg School of Medicine, Chicago, USA.

Florian A. Schiffers is with the Department of Computer Science, Northwestern University, Evanston, USA.

Aggelos K. Katsaggelos and Santiago L. Tapia are with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, USA.

Daming Shen was with the Department of Radiology, Northwestern University Feinberg School of Medicine, Chicago, USA. He is now with GE Healthcare Inc., Chicago, USA.

Daniel C. Lee is with the Department of Radiology and also with the Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, USA.

I. Introduction

CINE cardiovascular magnetic resonance (CMR) is the gold standard imaging test for evaluation of cardiac function and volumes [1], [2]. The standard pulse sequence is the retrospective, electrocardiogram gated, breath-held cine MRI with balanced steady state free precession (b-SSFP) readout [3]. In patients with arrhythmia and/or dyspnea, breath-hold, b-SSFP cine MRI may yield nondiagnostic image quality due to ghosting image artifacts [4]. In such scenarios, a real-time cine MRI may mitigate these artifacts. Standard real-time cine MRI using parallel imaging [5], [6], however, produces poor spatio-temporal resolution. One approach to achieve relatively high spatio-temporal resolution in real-time cine MRI is highly accelerating the scan using compressed sensing (CS) [7]. For data acceleration with CS, the incoherent k-space sampling pattern could be achieved with either Cartesian [8], [9] or non-Cartesian [10]–[12] approaches, with each having advantages and disadvantages. In this study, we focus on radial k-space sampling with tiny golden angles, which has the following advantages [13]: (a) incoherent k-space sampling; (b) retrospectively rebin to arbitrary acceleration rate to achieve patient-specific temporal resolution based on heart rate. A major hindrance to clinical translation with CS accelerated real-time cine with radial k-space sampling is the lengthy reconstruction time, owing to the iterative nature and also to the need to conduct nonuniform fast Fourier transform (NUFFT) [10], [14].

In recent years, deep learning (DL) reconstruction techniques have witnessed a remarkable surge of interest due to their promising ability to overcome the inherent challenges of iterative reconstruction methods. The most straightforward

DL-based reconstruction techniques are the so-called post-processing or denoising networks [14]–[16]. These algorithms "learn" a function that maps a zero-filled MR image that is contaminated with aliasing artifacts to its de-aliased counterpart. The a-priori learning phase requires a large dataset consisting of pairs of aliased images and their corresponding "ground truth." This approach, however, allows much faster inference than what is achievable with an iterative technique. Nevertheless, performance of purely data-driven denoisers depends highly on the size and the quality of the training dataset at hand. Furthermore, in general, there are no theoretical guarantees for performance or generalizability of the trained network. This a common problem in DL [17]. Closely related is the issue of adversarial attacks or instability with respect to small perturbations in the input [18], [19].

State-of-the-art reconstruction methods aim to combine the advantages of physics-based iterative and data-driven techniques [20], [21]. It has been shown that incorporating forward and adjoint operators into convolutional neural network (CNN)-based reconstruction increases stability and robustness against adversarial attacks [22]. Another study showed that including the forward model lowers the maximum error bound of the network [23]. An entire class of hybrid architectures has been derived from iterative algorithms by enhancing them with learned components. These networks are called variational, cascaded or unrolled networks as they often resemble an "unrolled" cascade of blocks, each consisting of a physics prior and a CNN-based regularizer. Therefore, each block successively refines the initial solution by simultaneously enforcing consistency with the measured k-space and regularizing the solution according to a data-driven term. Integrating prior knowledge into the architecture as a differentiable form of the imaging operator, enables significant reduction in the necessary training data [23], [24]. For several important iterative algorithms, such as ADMM [25] or ISTA [26] corresponding unrolled architectures have been successfully developed [27]–[29].

Although unrolled networks are effective with Cartesian acquisition schemes [30], [31], they require extensive computational capabilities in the case of non-trivial forward and adjoint operators (e.g., NUFFT [32]) to compute and propagate the error gradients "through" the model operators. Additionally, end-to-end training requires storing intermediate gradients of each block during backpropagation within the GPU VRAM. This computational bottleneck is a major hindrance to apply unrolled networks to multi-dimensional MRI with computationally demanding imaging operators [30], [32].

In this paper, we propose a novel strategy called IMplicit UNrolled NEtwork (IMUNNE) that leverages implicit network theory [33]–[35] to overcome the memory and training time limitations associated with unrolled networks, while retaining their advantages. Implicit networks have shown promising results in various machine learning tasks [36]–[38], and by incorporating implicit network theory into the framework of unrolled networks, we aim to reduce the computational burden and enhance the applicability of unrolled networks to multi-dimensional MRI reconstruction with radial k-space sampling.

Our research focuses on multi-coil 2D real-time cardiac cine MRI obtained using radial k-space sampling, where we evaluate the IMUNNE approach extensively. We compare the performance of our proposed method with two state-of-the-art methods: (a) a complex U-Net (CU-Net) [14], [30]; (b) end-to-end alternating unrolled network for non-Cartesian dynamic MRI reconstruction (Alt-UN) [32]. The main contributions of our work are as follows:

- We propose a novel strategy to train unrolled networks in the context of highly-undersampled, multi-coil dynamic MRI reconstruction. In particular, we reinterpret the end-to-end cascaded structure as an implicit network and apply Jacobian-free backpropagation [33] strategy to approximate the error gradient for weight updates. This strategy drastically reduces the computational demand and memory requirements during training, thereby enabling more complex operators and larger input sizes.
- We demonstrate this by applying IMUNNE as a reconstruction algorithm on a multi-coil real-time cine MRI dataset, using b-SSFP readouts. The method shows superior performance compared to state-of-the-art reconstruction techniques on the given dataset, which suggests versatility and adaptability of the method for real-time cine MRI using radial k-space sampling.
- We demonstrate that training IMUNNE does not require a large dataset. Moreover, IMUNNE consistently performs well when evaluated on noise-corrupted data. These benefits are typical of end-to-end unrolled networks and remain true for IMUNNE, even though the method avoids full end-to-end training and only approximates the correct error gradient.

The structure of the paper is organized as follows: In Section II, we provide a formal introduction to the reconstruction problem and present the details of the CNN-block used in our proposed method, along with the training strategy employed. Section III presents a comprehensive set of experiments conducted to validate the computational efficacy, accuracy and robustness of our approach. We compare our results with CU-Net – a purely data-driven approach – and Alt-UN – an end-to-end trainable unrolled network for real-time cine MRI using radial k-space sampling. Additionally, in Section IV, we discuss the main advantages and limitations of our work. There, we also provide a theoretical comparison of our method with related works. Finally, in Section V, we conclude our findings and discuss future research directions.

II. Methodology

A. Theoretical Background

The inverse problem of undersampled MRI reconstruction can be described as the following minimization problem:

$$\min_{\boldsymbol{x}} \left|\left|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\right|\right|_2^2 + \mathcal{R}(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{A}$ is the NUFFT operator, $\boldsymbol{x}$ the unknown image, $\boldsymbol{y}$ the undersampled k-space data impacted by noise, and $\mathcal{R}$ a regularizer function restricting the solution space. Traditionally, this minimization problem is solved by iterative approximation algorithms, such as the conjugate gradient (CG) method. In our work, we are primarily concerned with multi-coil, real-time cine MRI using radial k-space sampling. In particular, $\boldsymbol{x} \in \mathbb{C}^d$, with $d = d_x \times d_y \times d_t$ being the two spatial dimensions and the temporal dimension, respectively; $\boldsymbol{y} \in \mathbb{C}^{d_x \times d_{sp} \times d_t \times n_c}$, n_c denoting the number of coils, d_{sp} being the number of radial spokes per time frame; and $\boldsymbol{A}$ given by the concatenation of the operator $\boldsymbol{C}: \mathbb{C}^d \to \mathbb{C}^{d \times n_c}$, which multiplies the cine image with each of the coil sensitivity maps, and the NUFFT applied to each of the coil images. As a formula:

$$\boldsymbol{A} \coloneqq NUFFT \circ \boldsymbol{C}. \tag{2}$$

For highly accelerated sampling, i.e., $d_{sp} \ll d_y$, the Nyquist sampling criterion is not satisfied, and thus the problem is ill-conditioned. The higher the undersampling rate, the more the solution depends on the correct choice of regularization [39]. Traditionally, iterative minimization algorithms have employed different regularization strategies such as wavelet transform [40] or discrete cosine transform [41], [42]. In the context of real-time cine MRI, temporal total variation with L1 norm is widely used to enforce gradient sparsity in the time dimension [43]. However, for higher undersampling rates, the manual choice of the regularizer might prove insufficient to eliminate aliasing artifacts [44].

Unrolled networks [30], [32], [44], [45] can be seen as an iterative method which "learns" the optimal regularizer from a representative dataset. Assuming a differentiable regularizer $\mathcal{R}$, the gradient descent iterates for an optimal point of (1) are as follows:

$$\boldsymbol{x}_{k+1} \coloneqq \boldsymbol{x}_k - s \cdot [\boldsymbol{A}^{\perp}(\boldsymbol{A}(\boldsymbol{x}_k) - \boldsymbol{y}) + \nabla\mathcal{R}(\boldsymbol{x}_k)] \tag{3}$$

with a scalar step size $s > 0$.

Next, we express the gradient of the regularizer $\nabla\mathcal{R}$ as a neural network U_θ with a trainable parameter set $\Theta \coloneqq (\theta_1, \dots, \theta_p)$. The update operator $R_{\theta,s}$ to produce the next iterate $\boldsymbol{x}_{i+1} \coloneqq R_{\theta,s}(\boldsymbol{x}_i)$ is then given by:

$$R_{\theta,s} \coloneqq \boldsymbol{I}_d(\cdot) - s \cdot [\boldsymbol{A}^{\perp}(\boldsymbol{A}(\cdot) - \boldsymbol{y}) + U_\theta(\cdot)] \tag{4}$$

where $\boldsymbol{I}_d$ denotes the identity map.

Instead of iterating until convergence, "unrolled" networks apply the update operator $R_{\theta,s}$ for a fixed number of times n, whereby an initial input is successively refined by each of the unrolled blocks (Fig. 1). This setup defines a differentiable algorithm that is solely dependent on the parameter set Θ, which can therefore be trained end-to-end using conventional error backpropagation [24], [46] to find the optimal parameters Θ. However, conventional error backpropagation to train unrolled networks limits the unrolling to only very small numbers of blocks since the memory requirements grow linearly with the network depth [30], [32], [33]. This problem is particularly exacerbated by the presence of computationally demanding operators, such as NUFFT, and/or multi-dimensionality of real-time cine MRI data (i.e., multi-coil, 2D, time) [32].

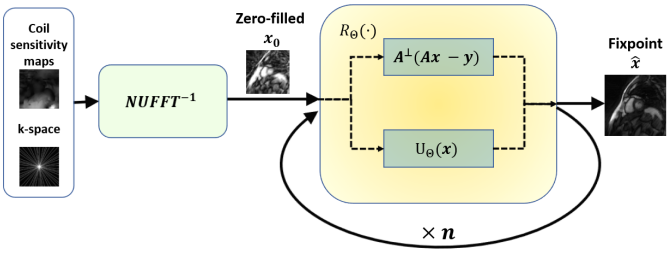

Fig. 1. Schematic illustration of the unrolled structure interpreted as an implicit network: Zero-filled input $\boldsymbol{x}_0$, which was obtained by the inverse NUFFT operator, is refined by the update R_θ until convergence. R_θ consists of a data-consistency term and learned regularizer.

B. Implicit Networks

Another way to approach the minimization problem (1) is to consider the fact that the gradient of the (continuously differentiable) objective function attains zero at an optimal point $\hat{\boldsymbol{x}}$. This has the consequence that $\hat{\boldsymbol{x}}$ is a fixed-point of the update operator $R_{\theta,s}$:

$$R_{\theta,s}(\hat{\boldsymbol{x}}) = \hat{\boldsymbol{x}} - s \cdot \underbrace{[\boldsymbol{A}^{\perp}(\boldsymbol{A}(\hat{\boldsymbol{x}}) - \boldsymbol{y}) + U_\theta(\hat{\boldsymbol{x}})]}_{=\nabla\left[||\boldsymbol{A}\hat{\boldsymbol{x}}-\boldsymbol{y}||_2^2+\mathcal{R}(\hat{\boldsymbol{x}})\right]=\boldsymbol{0}} = \hat{\boldsymbol{x}} \tag{5}$$

Under certain assumptions for $R_{\theta,s}$, we can formulate the MRI reconstruction problem in the following form of an implicit fixed-point equation:

$$Find\ \hat{\boldsymbol{x}}\ with\ \hat{\boldsymbol{x}} = R_{\theta,s}(\hat{\boldsymbol{x}}). \tag{6}$$

The type of neural networks that find a solution of an implicit equation in the form of (6) are called implicit networks [33]–[35]. In contrast to classical feed-forward networks that produce an explicit deterministic output, implicit networks approximate the solution of an implicit equation via a fixed-point iteration similar to (3).

Given the inconsistent use of the term *implicit* in deep-learning literature, it is important to clarify that the implicit networks mentioned earlier are categorically different from *implicit neural representation networks* [47], [48]. This distinction is crucial, particularly because neural representation networks have already been extensively applied to dynamic MRI reconstruction [49]–[51]. Our research is entirely unrelated to implicit neural representation networks. Therefore, whenever we refer to an *implicit network*, we are specifically referencing the type of networks that solve a fixed-point equation [33].

The unrolled network structure defined by (4) can thus be viewed as an implicit network and the unrolled blocks correspond to the fixed-point iteration (Fig. 2), which solves

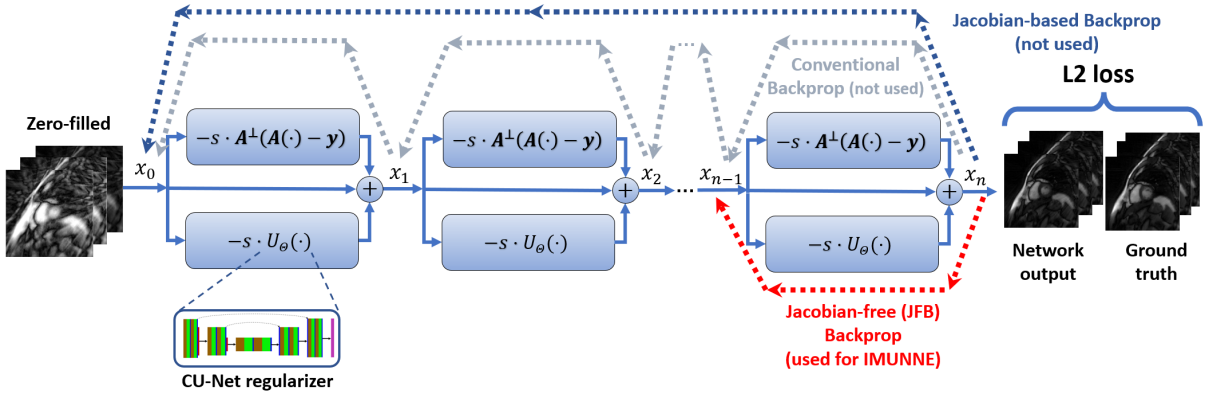

Fig. 2. Unrolled architecture defined by the update operator $R_{\theta,s}$: n identical unrolled blocks consisting of data consistency (top), skip connection (center), and a learned regularizer term (bottom). Zero-filled initial estimate x_0 is refined consecutively by each block. For the sake of brevity, the figure does not show the measured k-space and the coil-sensitivity maps which are part of the input for the data consistency blocks. The last block produces the final image x_n. End-to-end training can be achieved by either conventional or Jacobian-based, i.e. analytical, backpropagation. In contrast to end-to-end training, IMUNNE interprets the unrolled scheme as an implicit network and applies Jacobian-free backpropagation [33], i.e., error is propagated only through the last block, thereby approximating the true gradient. Due to weight sharing, this is sufficient to update all regularizers. In our experiments, we use a complex U-Net to implement the learned regularizer U_{Θ}.

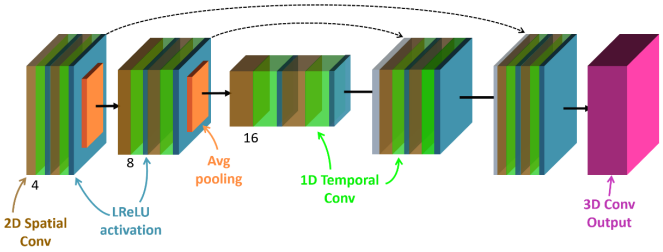

Fig. 3. Complex U-Net (CU-Net) architecture: encoding/decoding stages enhanced with skip connections (dashed line). Each stage consists of 2D spatial followed by 1D temporal convolutional layers (cf. [30]). We use leaky ReLU as activation function and average pooling. The final 3D convolution layer produces the final output. We employed CU-Net both as IMUNNE regularizer (U_{θ}) and as a standalone denoiser network with the same number of parameters.

the implicit problem stated in (6). Furthermore, if $R_{\theta,s}$ is a *contraction*, i.e., it is γ-Lipschitz with $\gamma \in [0,1)$, a classical result in functional analysis guarantees linear convergence of the sequence defined in (3) [52].

One substantial advantage of implicit networks is that they use analytical gradients for backpropagation, and thus enable training of very deep architectures with fixed memory cost. Furthermore, analytical gradients allow network training when the number of unrolls is a priori unknown. In contrast, conventional backpropagation computes gradients for each block sequentially in the reverse order and, therefore, requires a fixed number of unrolls. Previous works used the costly Jacobian operator for the analytical weight update [34], [35]. However, a recent study showed that the true gradient can be approximated by backpropagating through last iteration only using Jacobian-free backprop (Fig. 2) [33]. The authors of [33] provide a theoretical error bound of the approximation and experimentally show that this approach even alleviates some of the constraints on the operator $R_{\theta,s}$, e.g., the requirement for it to be a contraction.

In our work, we apply this theoretical result heuristically to train an unrolled network "implicitly", i.e., performing backpropagation through the last unrolled iteration only using multi-coil, real-time cine MRI datasets. Given the fixed-point iteration, weight sharing is used between the n regularizers and, therefore, Jacobian-free backpropagation (i.e. only through the last block) is sufficient to calculate the updated weights for the entire architecture. This setup would not be possible with conventional backpropagation due to memory constraints of currently available GPU cards. To the best of our knowledge, this study represents the first application of implicit network theory to enable unrolled training in the context of multi-dimensional MRI.

C. *Proposed Training Method*

In this work, we introduce IMUNNE, a simple scheme of gradient backpropagation through last iteration only. For architectures with identical unrolled blocks, we have explained in the last subsection how they can be reformulated as an implicit network and therefore benefit from Jacobian-free backpropagation. Fig. 2 depicts the architecture and the IMUNNE training scheme. In the following, we will use the abbreviation IMUNNE for both the architecture and the training strategy.

The structure is comprised of an unrolled CS scheme. The network receives zero-filled cine image $\boldsymbol{x}_0$ as input and passes it through a computational block n times. In accordance with (4), the block computes the sum of the following three components: Model operator term involving NUFFT; a skip connection; and a data-driven regularizer term incorporating a CU-Net [14], [30], [53] which acts as the regularizer U_{Θ}. Each time the cine image set is processed by the block, it is being refined, until after n steps it yields the network output $\boldsymbol{x}_n$. In contrast to CS, where the gradient step size s is determined by the line search method for each iteration, we treat s as a hyper-parameter and keep it fixed for all unrolled blocks. More elaborate schemes to initialize (or even train) s in combination with IMUNNE might be an interesting direction for future investigation, which is beyond the scope of this initial work.

The regularizer block was realized by a CU-Net from the CINENet architecture [30] which we adapted to process 2D+time input (Fig. 3). The authors showed the advantage of

using a cascade of spatial convolutions followed by 1D temporal convolutions to better capture the spatial and dynamic features and to reduce the number of trainable parameters. In contrast to the original work, our version of the CU-Net consists of 3 encoding/decoding stages enhanced with skip connections. Each stage consists of 2D spatial followed by 1D temporal convolutional layers [30]. We used leaky ReLU as activation function and average pooling for dimensionality reduction. The final 3D convolution layer produces the network output.

IMUNNE was trained in a supervised manner: the L_2 difference between $\boldsymbol{x}_n$ and the ground truth image $\boldsymbol{x}$ was computed as the final loss function. We used the ADAM [54] optimizer with default parameters. Due to memory limitations, we set the batch size to be 1.

D. Experiments with Patient Data

To assess the computational efficacy and accuracy of IMUNNE, we conducted several preliminary experiments with in-vivo patient data described in the next section. First, we performed an extensive hyper-parameter grid search to find optimal values for the step size s and the number of unrolled blocks n. Finally, we compared IMUNNE to CU-Net and Alt-UN. In the subsequent sections, we present comprehensive details pertaining to the utilized datasets, the evaluation methodologies employed, the implementation specifics of the proposed method, and the computational environment in which the experiments were conducted.

E. Datasets

We used a retrospective real-time cine dataset obtained using radial k-space sampling [10]. The dataset contained 49 real-time cine scans using b-SSFP readout acquired from a cohort of 40 patients (9 patients had follow-up exams on different days; 30 men; mean age = 68 ± 9.7 years) with atrial fibrillation, as previously described [14]. All patients provided informed consent in writing and agreed to future analysis of data. This study was performed in accordance with protocols approved by our institutional review board (Northwestern University; protocol number: STU00205545 [approval date: 10/08/2019]) and was Health Insurance Portability and Accountability Act (HIPAA) compliant.

Each scan consisted of 11-18 slices per patient and 40 cardiac frames per slice. We randomly split the dataset into 10 cases for training, 10 cases for validation and 29 cases for testing. This split was motivated by several reasons: First, allocating approximately 20% of the data for training enabled us to manage the long training times. This is reinforced by the fact that unrolled networks require less training data in general [23], [24]. Second, assigning 60% of the dataset to testing achieves a higher level of statistical fidelity during evaluation. Finally, this setup accurately simulates the practical constraints of working in a clinical context with limited data availability. The sample unit for training was a single cine series resulting in 140 training, 137 validation and 410 testing cine sets. Due to reasons listed in [14], we obtained the ground truth using Golden-angle radial sparse parallel MRI (GRASP) [55]. The coil sensitivity maps were self-calibrated from time-average data using the method described by Walsh et al. [56]. We applied coil compression to 8 virtual coils using PCA [57]. The reconstructed coil sensitivity maps were subsequently used within training and evaluation of the unrolled networks (IMUNNE and Alt-UN). For the initial reconstruction and throughout the experiments, we applied a simple ramp filter as density compensation for the radial trajectory.

Pertinent imaging parameters included: field of view (FOV) = 288 × 288 mm, matrix size = 160 × 160, spatial resolution = 1.8 mm × 1.8 mm, slice thickness = 8 mm, TE = 1.4 ms, TR = 2.7 ms, receiver bandwidth = 975 Hz/pixel, tiny golden angle sequence = 23.62814° [58], temporal resolution = 29.7 ms, and flip angle = 52°. We used 11 radial spokes per cardiac frame, corresponding to an effective acceleration factor of 15 (compared to fully sampled Cartesian counterpart).

F. Hyperparameter Search

To find an optimal network configuration, we first performed a hyper-parameter 2D grid search for different numbers of unrolled blocks ($n = 2,3,4,5,6$) and fixed step sizes ($s = 0.5, 0.6, \dots, 1.3, 1.4$). To faithfully compare the different configurations, we trained each of them exactly with 500 epochs.

G. Comparison with CU-Net and Alt-UN

We compared our IMUNNE strategy to a standalone CU-Net acting as a denoiser reconstruction network. For a fair comparison, we used the same configuration, particularly the same number of trained parameters, as in the CU-Net within the unrolled architecture. In addition to the dataset described in the previous section, we also performed cross-validation of CU-Net trained on the extended training set comprising 80% of the data and evaluated on the remaining 20%. Moreover, we compared IMUNNE to Alt-UN, an end-to-end UN specifically targeting multidimensional non-Cartesian MRI reconstruction [32]. In contrast to our architecture, this approach alternates data-consistency and CNN-based regularization blocks within the end-to-end reconstruction pipeline to reduce the GPU memory requirements.

Before training, the network weights were initialized using the default PyTorch initialization routine (e.g. uniform zero-mean initialization for convolutional layers; negative slope = 0.01 for leaky ReLU). For the sake of simplicity, we used the same learning rate of 5e-04 with a learning rate decay factor of 0.95 every 30 epochs for all training instances. Both CU-Net and IMUNNE were trained for 500 epochs, respectively. Alt-UN was trained according to the training scheme prescribed by the authors of the original publication [32]: the regularization block was pretrained followed by 150 epochs of fine-tuning of the end-to-end architecture.

Finally, we performed the same comparison experiments using noise-corrupted k-space measurements as input. Specifically, a given k-space measurement $\mathrm{ksp_m}$ was corrupted by adding a scaled noise vector sampled from a complex normal distribution [59]. The scaling parameter for the noise was $0.1 \times \max\left(\mathrm{abs}(\mathrm{ksp_m})\right)$, where the right factor

stands for the maximum absolute value of $\mathrm{ksp_m}$. Subsequently, we used the distorted data to evaluate the stability of the networks which had been trained on the original uncorrupted data, as described above.

H. Evaluation Metrics

To evaluate and compare the performance of IMUNNE between different hyper-parameter settings and with other methods, we calculated the peak signal-to-noise ratio (PSNR), the structural similarity index (SSIM) [60] and the normalized root mean-square error (NRMSE) of network outputs with respect to the corresponding ground truth. A reference-free blur metric [61] (0: sharpest; 1: blurriest) was calculated to assess the preservation of fine details. To assess sharpness in spatial domain, we computed the metric for each frame of the (x, y, t) cine series and report the mean value as "2D Blur". To quantify temporal blurring, we follow the procedure in [62] and calculate the metric for the 1D-FFT-transformed series (x, y, f) which we refer to as "FFT Blur". Finally, we report the combined spatio-temporal reduced reference entropic difference (ST-RRED) [63], [64] to evaluate to overall video quality. All measures were calculated over a central region of 100×100 pixels, in order to focus on the heart region. Other relevant performance factors that we report are the training time and the evaluation time per instance.

I. Implementation and Computational Environment

To train the different network architectures, we used a Linux GPU workstation (NVIDIA A100-PCIE, 40GB VRAM GPU; AMD EPYC 7702P 64-Core CPU, 512GB RAM). All architectures were implemented using PyTorch with GPU acceleration. To incorporate backpropagatable NUFFT operators, we used the TorchKbNufft [65] library. The code implementation of the proposed method was made publicly available on GitHub[1].

For Alt-UN [32], we used the implementation which was provided by the authors as part of the original publication and only adapted it slightly to fit our datasets. However, we opted to eliminate the residual connection within the image domain since our datasets were reconstructed using an adaptable regularization weight, leading to different scaling magnitude between the zero-filled inputs and the reconstructed images. A visual comparison of the evaluation results, both with and without the skip connection, is shown in Fig. S3. In particular, the network version with the residual connection produced poor image quality. Otherwise, training was performed using the same hyperparameters as reported in the original publication with a trainable weight λ.

For both IMUNNE and Alt-UN, the network input consisted of four components listed in the following with the corresponding dimensions and their respective specific values (x, y: 2D spatial coordinates; f number of time frames; c: number of coils; nsp: number of radial k-space spokes per time frame):

- Zero-filled input / Ground truth data: 320 x 320 x 40; (x, y, t)
- Measured k-space data: 320 x 8 x 11 x 40: (x, c, nsp, t)
- Coil sensitivity maps: 8 x 320 x 320: (c, x, y).

Since CU-Net does not require k-space, the network input consisted only of the zero-filled image, spatially cropped to the central region of 160 x 160 x 40.

All subsequent evaluation experiments were conducted using the centred ROI of 100 x 100 x 40 pixels.

III. Results

In this section, we present the results obtained from our experiments with patient data, focusing on the training behavior and reconstruction performance of the proposed method.

A. Hyperparameter Choice

The first experiment involved finding the optimal values for the step size s and the number of unrolled blocks n for the IMUNNE architecture. Except for the blur metric, all other evaluation metrics indicate $n = 5$ and $s = 0.8$ to be an optimal pair (see also supplementary Fig. S4). In the case of 4 unrolls the blur metric shows improvement for some large s values. However, since artificial sharpness can be introduced by an increased presense of noise [14], we continued to use the parameters $(n, s) = (5, 0.8)$ for subsequent experiments.

A more extensive grid search would be required to thoroughly understand asymptotic behavior of the metrics for $n \to \infty$ and $s \to 0$. Heuristically, one might expect better results to the degree $R_{\Theta,s}$ is a contraction while simultaneously increasing the number of iterations. It is straightforward to see from (4) that this corresponds to the case of very small values of s and large values of n. While this is an iteresting theoretical consideration, it is practically infeasible to test this scenario due to the following reasons:

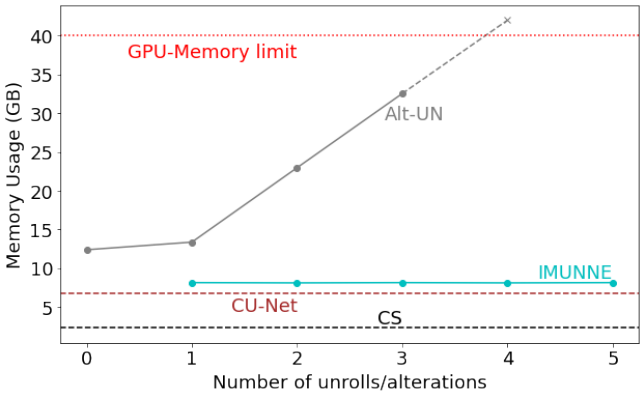

Fig. 4. Comparison of memory footprints of Alt-UN, IMUNNE and CU-Net during training, and GPU-accelerated CS reconstruction: The memory requirement for IMUNNE is constant regardless of the number of iterations, whereas Alt-UN shows a linear increase. Having the same trainable base, CU-Net and IMUNNE only differ due to additional storage of k-space and sensitivity maps within IMUNNE. The initial gap between Alt-UN and IMUNNE is caused by the difference in the number of trainable parameters (Alt-UN: 129K; IMUNNE 21.7K). Note that the number of unrolls/alterations is not applicable to CU-Net and CS which are, therefore, indicated as horizontal lines in the plot.

[1] https://github.com/niakovlev/IMUNNE

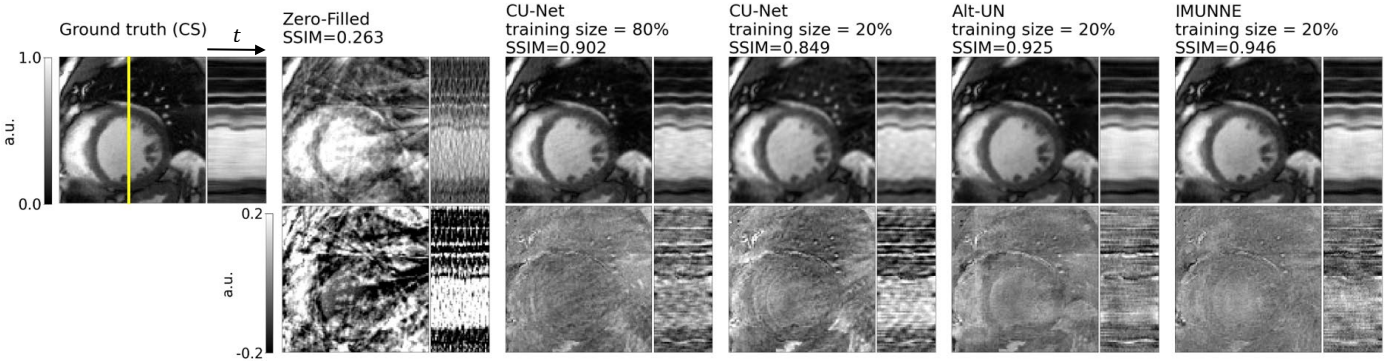

Fig. 5. Representative example from b-SSFP testing dataset. The upper row depicts the 15th frame along with the temporal profile along the vertical middle line (e.g., like M-mode in echocardiography; yellow line in the left image) of the cine series reconstructed with (from left to right): Compressed Sensing (ground truth image); inverse NUFFT of the undersampled k-space; CU-Net trained on 80% of the data; CU-Net; Alt-UN; IMUNNE. The bottom row shows the corresponding scaled difference images for each method. For dynamic display, see Video S1 in Supplemental Materials.

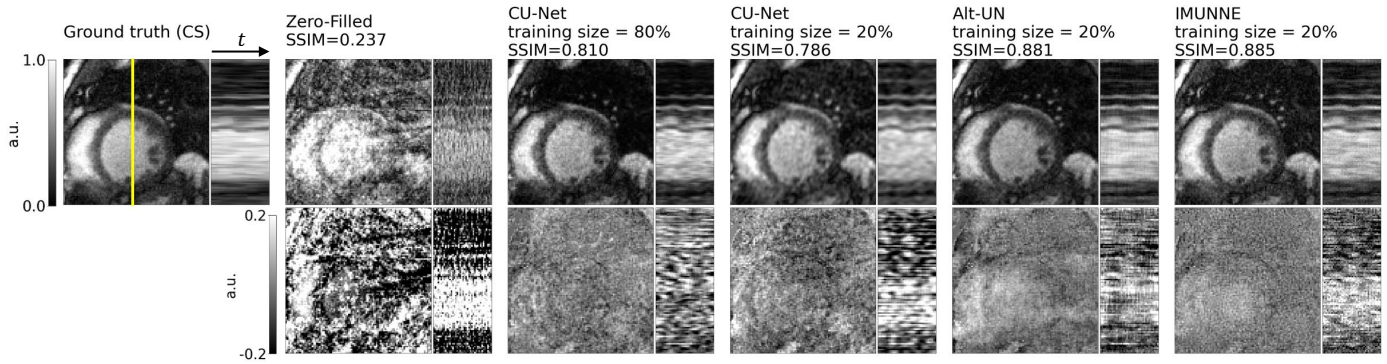

Fig. 6. Same representative example from b-SSFP testing dataset with noise-corrupted k-space as input. The upper row depicts the 15th frame along with the temporal profile along the vertical middle line (e.g., like M-mode in echocardiography; yellow line in the left image) of the cine series reconstructed with (from left to right): Compressed Sensing (ground truth image); inverse NUFFT of the undersampled k-space; CU-Net trained on 80% of the data; CU-Net; Alt-UN; IMUNNE. The bottom row shows the corresponding scaled difference images for each method. For dynamic display, see Video S2 in Supplemental Materials.

1. In order to have a practically relevant advantage to CS in terms of computational speed, n must be significantly lower than the number of iterations in CS (30 in our case).
2. In order to compensate for a low n, we need to increase s to allow the network to make "larger leaps" towards the optimum. This brings the operator $R_{\Theta,s}$ further away from being a contraction.
3. Due to lengthy training times (~3-5 days) for just one hyperparameter pair (n, s), it is infeasible to perform a extensive grid search for larger number of unrolls, since both training and the inference time increase proportionally with n.

Although these practical constraints (small n; large s) lead to a deviation from the theoretically optimal scenario (large n; small s), the optimal values found through the limited grid search, nevertheless, produced promising results, as we illustrate in the following sections.

B. Memory benchmarks

To demonstrate the memory-efficiency of IMUNNE, we performed an ablation study with different numbers of unrolled blocks (IMUNNE) and different numbers of alterations (Alt-UN; parameter M in the original publication [32]), respectively. We measured the peak allocated GPU memory (in megabytes) for each of the networks during training. Our results clearly indicate the advantage of IMUNNE (Fig. 4): constant memory requirement despite growing number of iterations since training is performed only using the last unrolled block. On the other hand, Alt-UN, being an end-to-end trainable architecture, demonstrates a linear increase in memory with the growing number of alterations. Note that the case of "0 alterations" corresponds to CNN pretraining and is not a UN in the strict sense. Due to memory limits of our GPU, it was only possible to perform up to 3 alterations for Alt-UN. Despite the same trainable component of IMUNNE and CU-Net (21.7K trainable parameters), the former requires the additional storage of raw k-space data and the sensitivity maps for the data-consistency component, which leads a minor increase in memory. The initial gap between the both unrolled networks is cause by the difference in the number of trainable parameters (Alt-UN: 129K; IMUNNE: 21.7K).

C. Performance Comparison

Here, we demostrate the reconstruction results of the fully-trained architectures. For IMUNNE, we used the hyperameters obtained through a grid search described above. Other relevant network and training parameters are described in Sections II.F and II.G.

Fig. 5 shows a representative example (see also supplementary Video S1 for dynamic display). Both unrolled networks clearly outperform the standalone CU-Net, even in case when CU-Net was trained with an extended training dataset (80% of all available data). In particular, CU-Net fails

TABLE I
QUALITATIVE EVALUATION METRICS, TRAINING TIMES, AND EVALUATION TIMES.
*CU-NET TRAINED ON THE EXTENDED TRAINING DATASET (80%) AND EVALUATED ON A LIMITED TESTING SET (20%).
**TRAINING INVOLVES PRETRAINING THE CNN BLOCK AND FINE-TUNING THE END-TO-END NETWORK [32].
+#P > 0.05 CORRESPONDS TO NON-SIGNIFICANT DIFFERENCE IN PAIR.

	CS Reference	Zero-filled NUFFT	CU-Net (80%)*	CU-Net	Alt-UN	IMUNNE
SSIM		0.223 ± 0.055	0.912 ± 0.033	0.880 ± 0.045	0.925 ± 0.029	**0.950 ± 0.021**
NRMSE		0.200 ± 0.042	0.021 ± 0.009	0.032 ± 0.012	0.026 ± 0.009	**0.019 ± 0.007**
PSNR (dB)		14.16 ± 1.84	34.10 ± 2.85	30.19 ± 2.94	32.34 ± 2.86	**34.88 ± 3.20**
2D Blur	0.362 ± 0.062	0.313 ± 0.048	0.398 ± 0.046	0.418 ± 0.061	0.392 ± 0.062+	**0.385 ± 0.060+**
FFT Blur	0.223 ± 0.031	0.250 ± 0.031#	0.220 ± 0.027	0.277 ± 0.029	0.249 ± 0.029#	**0.241 ± 0.030**
ST-RRED		1.798 ± 0.837	0.286 ± 0.147	0.412 ± 0.221	0.311 ± 0.152	**0.215 ± 0.125**
Training time	–	–	1d 3h 500 epochs	**7.5h** 500 epochs	5d 4h 500+150 ep.**	3d 8h 500 epochs
Inference time	20 s ± 82.6 ms	607 ms ± 6.2 ms	653 ms ± 2.8 ms	653 ms ± 2.8 ms	9.53 s ± 40.4 ms	2.04 s ± 8.1 ms

to remove the radial streaking artifacts. Compared to Alt-UN, IMUNNE showed slightly better image quality and significantly outperformed Alt-UN in all reported metrics except the spatial blur metric. Furthermore, IMUNNE was able to produce these results while reducing the training time from 124 hours to 80 hours (> 35% reduction); since our approach uses less unrolled iterations, the inference time was also reduced from 9.53 seconds to 2.04 seconds (> 75% reduction). The training and inference times and the image quality metrics are summarized in Table I.

D. Robustness Analysis

In this subsection, we demonstrate the comparison results under the presense of complex Gaussian noise. The k-space measurements were noise-corrupted as described in Section II.G. Subsequently, the architectures were evaluated using the noisy k-space data as input.

Fig. 6 (see also supplementary Video S2) shows the same representative example as in Fig. 5. IMUNNE maintains its superior performance under presence of noise and successfully removes streaking artifacts. The evaluation metrics are summarized in the supplementary Table S.I which shows a similar trend as in the base case (uncorrupted k-space input).

IV. DISCUSSION

In this paper, we introduced a novel implicit training strategy for unrolled networks (IMUNNE) to address the challenges associated with memory-intensive end-to-end training of unrolled networks for multi-dimensional MRI reconstruction. Our approach leverages implicit network theory to approximate the gradient for weight updates during training, enabling us to overcome the computational bottleneck and efficiently apply unrolled networks to multi-dimensional MRI datasets.

Our experiments with 2D multi-coil real-time cardiac cine MRI dataset obtained using radial k-space sampling demonstrated the efficacy of IMUNNE. The results indicated that IMUNNE outperforms a purely data-driven approach (CU-Net reconstruction) and is competitive with a state-of-the-art end-to-end unrolled network (Alt-UN). While achieving similar image quality, IMUNNE is able to cut down training time significantly since – in contrast to end-to-end training – backpropagation is performed only for the last unrolled iteration. Given the free GPU memory, it is possible to further reduce the training time for IMUNNE by increasing the batch size (Fig. 4). However, for the sake of simplicity and a fair comparison with the baseline methods, we fixed the batch size to 1. Additionally, the 2D blur metric indicated that IMUNNE preserves fine details slightly better, enhancing the resulting spatial resolution. The significantly better FFT blur metric points to improved temporal resolution achieved by IMUNNE.

In the following, we present some theoretical considerations, discuss advantages and limitations of our proposed approach relative to other works.

A. Limitations

In order to calculate the fixed point of an implicit equation to a high degree of precision, implicit networks generally need to compute a large number of iterations [33], [34], [35], [52]. However, as we illustrated in Section III.A, the implicit network theory was only heuristically applied for IMUNNE; that is, the architecture could be trained for only a small number of unrolls due to practical constraints and lengthy training times. Despite this deviation from original implicit networks, we were able to successfully conduct a hyperparameter search to find the best values for the aforementioned parameters. As a result, IMUNNE showed promising results, outperforming CU-Net in terms of image quality and Alt-UN in terms of training and inference times.

Furthermore, we only evaluated IMUNNE in a supervised training setting [32], where ground truth data was obtained through a priori CS reconstruction. This is a limitation since, in this case, IMUNNE can never learn to outperform CS [66], [67]. However, there is nothing inherent in our proposed method that would preclude its application in an unsupervised training scenario. We leave this interesting research direction for future investigations.

B. Relation to Similar Works

Although extensive research has been conducted to accelerate MRI reconstruction using unrolled networks [30],

[32], [44], [45], very few strategies apply unrolled networks to high-dimensional MRI with non-Cartesian k-space sampling patterns. Beside Alt-UN [32], a similar approach was proposed for 3D cardiac cine MRI reconstuction employing a cascade of backpropagatable data consistency layers interleaved with U-Net-based regularizers [30]. However, this work focuses on Cartesian k-space data which does not pose the additional challenge of NUFFT operators. Other works have addressed non-Cartesian k-space sampling patterns but mostly in the context of non-dynamic reconstruction [68], [69].

Since fully sampled imaging data for training DL networks is notoriously challenging to obtain, there has been a growing research interest to develop unsupervised training techniques for reconstruction networks [70], [71]. For example, the dual-domain self-supervised training [45] divides the k-space measurement into disjoint partitions and compares the reconstructions of each of the partitions in the image and the k-space domains, exploiting self-similarity in both domains in the loss function. This strategy achieves competitive performance relative to supervised training for low acceleration factors $R = 2{,}4$. However, for higher undersampling ratios, the given k-space coverage is much lower, which makes the partitioning strategy unsuitable because it would further reduce the available input information from the measurement.

Although a true end-to-end training of the unrolled network is impossible due to GPU memory limits, it would be conceivable to compare IMUNNE with a pseudo end-to-end training approach using gradient checkpointing [72]–[74]. This strategy reduces the memory cost by storing only a part of the intermediate values required for backpropagation. Then, after the error for the stored part has been evaluated, another forward pass is required to further calculate backpropagation throughout the architecture. In the most extreme case, this evaluation can be repeated for each unrolled block, thereby reducing the memory cost to $O(1)$. This, however, comes at the expense of additional computational time to recalculate the forward pass n times, leading to a computational time increase from $O(n)$ to $O(n^2)$. Even though this approach calculates the gradient precisely, we decided not to perform this experiment since it would take several weeks to train the network in this manner for one single hyperparameter setting.

Finally, implicit networks have been applied to undersampled MRI reconstruction [36], [37]. In one of the most recent works [36] the authors developed a self-supervised implicit model that approaches the performance of supervised training. However, the method has only been tested on multi-coil 2D brain MRI using Cartesian k-space sampling with relatively moderate acceleration factors (≤ 8). Since the forward pass uses 100 iterations for the fixed-point estimation, it cannot be applied to dynamic non-Cartesian k-space data directly, especially when the model includes the computationally costly NUFFT operator.

In principle, it is expected that many of the available techniques can be combined with IMUNNE to further improve its performance. For example, we performed ablations with instance and layer normalization layers in the regularizer CU-Net but did not observe improvements in performance. We did not test batch normalization due to its poor performance for small batch sizes [75]. Furthermore, IMUNNE can be integrated with the self-supervised training approaches presented above. In contrast to these techniques, whose core objective is to develop a specialized loss term to achieve self-supervision, our strategy is primarily focused on a special backpropagation to update the network weights within the architecture itself. Nevertheless, a rigorous evaluation of a possible compatibility is necessary which we leave for future work.

V. Conclusion

In conclusion, our novel IMUNNE training strategy successfully applies unrolled networks to non-Cartesian, real-time cine MRI reconstruction, outperforming purely data-driven approaches. IMUNNE efficiently approximates the gradient for weight updates during training, significantly reducing memory usage and training time compared to fully unrolled networks. This work has the potential to facilitate a more widespread adoption of CMR by reducing training time, inference time and memory cost of state-of-the-art reconstruction methods, thereby lowering the clinical hardware requirements and energy consumption necessary for state-of-art reconstruction methods.

References

[1] D. S. Fieno *et al.*, "TrueFISP: assessment of accuracy for measurement of left ventricular mass in an animal model," *J. Magn. Reson. Imag.*, vol. 15, no. 5, pp. 526–531, Apr. 2002, doi: 10.1002/jmri.10107.

[2] F. Grothues *et al.*, "Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy," *Am. J. of Cardiol.*, vol. 90, no. 1, pp. 29–34, Jul. 2002, doi: 10.1016/S0002-9149(02)02381-0.

[3] J. C. Carr *et al.*, "Cine MR angiography of the heart with segmented true fast imaging with steady-state precession," *Radiology*, vol. 219, no. 3, pp. 828–834, Jun. 2001, doi: 10.1148/radiology.219.3.r01jn44828.

[4] J. N. Morelli *et al.*, "An image-based approach to understanding the physics of MR artifacts," *Radiograph.*, vol. 31, no. 3, pp. 849–856, May 2011, doi:10.1148/rg.313105115.

[5] P. Kellman *et al.*, "Adaptive sensitivity encoding incorporating temporal filtering (TSENSE)," *Magn. Reson. Med.*, vol. 45, no. 5, pp. 846–852, Apr. 2001, doi: 10.1002/mrm.1113.

[6] F. A. Breuer *et al.*, "Dynamic autocalibrated parallel imaging using temporal GRAPPA (TGRAPPA)," *Magn. Reson. Med.*, vol. 53, no. 4, pp. 981–985, Apr. 2005, doi: 10.1002/mrm.20430.

[7] M. Lustig *et al.*, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007, doi: 10.1002/mrm.21391.

[8] L. Feng *et al.*, "Highly accelerated real-time cardiac cine MRI using k-t SPARSE-SENSE," *Magn. Reson. Med.*, vol. 70, no. 1, pp. 64–74, Jul. 2013, doi: 10.1002/mrm.24440.

[9] K. Hong *et al.*, "Accelerated wideband myocardial perfusion pulse sequence with compressed sensing reconstruction for myocardial blood flow quantification in patients with a cardiac implantable electronic device," *Radiol. Cardiothorac. Imag.*, vol. 2, no. 2, Apr. 2020, doi: 10.1148/ryct.2020190114.

[10] H. Haji-Valizadeh *et al.*, "Validation of highly accelerated real-time cardiac cine MRI with radial k-space sampling and compressed sensing in patients at 1.5T and 3T," *Magn. Reson. Med.*, vol. 79, no. 5, pp. 2745–2751, May 2018, doi: 10.1002/mrm.26918.

[11] J. Frahm *et al.*, "Real-Time Magnetic Resonance Imaging: Radial Gradient-Echo Sequences With Nonlinear Inverse Reconstruction," *Invest. Radiol.*, vol. 54, no. 12, pp. 757–766, Dec. 2019, doi: 10.1097/RLI.0000000000000584.

[12] D. Voit *et al.*, "Real-time cardiovascular magnetic resonance at 1.5 T using balanced SSFP and 40 ms resolution," *J. Cardiovasc. Magn. Reson.*, vol. 15, no. 1, Sep. 2013, doi: 10.1186/1532-429X-15-79.

[13] S. Wundrak et al., "Golden ratio sparse MRI using tiny golden angles," *Magn. Reson. Med.*, vol. 75, no. 6, pp. 2372–2378, Jun. 2016, doi: 10.1002/mrm.25831.

[14] D. Shen *et al.*, "Rapid reconstruction of highly undersampled, non-Cartesian real-time cine k-space data using a perceptual complex neural network (PCNN)," *NMR Biomed.*, vol. 34, no. 1, Jan. 2021, doi: 10.1002/nbm.4405.

[15] L. Fan *et al.*, "Rapid dealiasing of undersampled, non-Cartesian cardiac perfusion images using U-net," *NMR Biomed.*, vol. 33, no. 5, May 2020, doi: 10.1002/nbm.4239.

[16] H. Haji-Valizadeh *et al.*, "Rapid Reconstruction of Four-dimensional MR Angiography of the Thoracic Aorta Using a Convolutional Neural Network," *Radiol. Cardiothorac. Imag.*, vol. 2, no. 3, Jun. 2020, doi: 10.1148/ryct.2020190205.

[17] P. P. Angelov *et al.*, "Explainable artificial intelligence: an analytical review," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 11, no. 5, e1424, Sep. 2021, doi: 10.1002/widm.1424.

[18] T. B. Brown *et al.*, "Adversarial Patch," 2017, *arXiv:1712.09665v2*.

[19] I. J. Goodfellow *et al.*, "Explaining and Harnessing Adversarial Examples," in *Proc. ICLR*, May 2015.

[20] F. Knoll et al., "Deep-Learning Methods for Parallel Magnetic Resonance Imaging Reconstruction: A Survey of the Current Approaches, Trends, and Issues," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 128-140, Jan. 2020, doi: 10.1109/MSP.2019.2950640.

[21] K. Hammernik et al., "Physics-Driven Deep Learning for Computational Magnetic Resonance Imaging: Combining physics and machine learning for improved medical imaging," *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 98-114, Jan. 2023, doi: 10.1109/MSP.2022.3215288.

[22] V. Antun *et al.*, "On instabilities of deep learning in image reconstruction and the potential costs of AI," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 48, pp. 30088–30095, Dec. 2020, doi: 10.1073/pnas.1907377117.

[23] A. K. Maier *et al.*, "Learning with Known Operators reduces Maximum Training Error Bounds," *Nat. Mach. Intell.*, vol. 1, no. 8, pp. 373-380, Aug. 2019, doi: 10.1038/s42256-019-0077-5.

[24] D. Gilton *et al.*, "Neumann Networks for Linear Inverse Problems in Imaging," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 328–343, Oct. 2019, doi: 10.1109/TCI.2019.2948732.

[25] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Mach. Learning*, 2011, doi: 10.1561/2200000016.

[26] I. Daubechies *et al.*, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004, doi: 10.1002/cpa.20042.

[27] H. Su *et al.*, "ADMM-Net: A Deep Learning Approach for Parameter Estimation of Chirp Signals under Sub-Nyquist Sampling," *IEEE Access*, vol. 8, pp. 75714–75727, Apr. 2020, doi: 10.1109/ACCESS.2020.2989507.

[28] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 1828–1837, Dec. 2018, doi: 10.1109/CVPR.2018.00196.

[29] D. You *et al.*, "ISTA-NET++: Flexible Deep Unfolding Network for Compressive Sensing," in *Proc. IEEE/ICME*, Mar. 2021, doi: 10.1109/ICME51207.2021.9428249.

[30] T. Küstner *et al.*, "CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-70551-8.

[31] C. Qin *et al.*, "Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 280–290, Jan. 2019, doi: 10.1109/TMI.2018.2863670.

[32] A. Kofler *et al.*, "An end-to-end-trainable iterative network architecture for accelerated radial multi-coil 2D cine MR image reconstruction," *Med. Phys.*, vol. 48, no. 5, pp. 2412–2425, May 2021, doi: 10.1002/mp.14809.

[33] S. W. Fung *et al.*, "JFB: Jacobian-Free Backpropagation for Implicit Networks," in *Proc. AAAI*, 2022, pp. 6648-6656, doi: 10.1609/aaai.v36i6.20619.

[34] S. Bai *et al.*, "Deep Equilibrium Models," in *Proc. NeurIPS*, vol. 32, 2019.

[35] E. Winston and J. Z. Kolter, "Monotone operator equilibrium networks," in *Proc. NeurIPS*, vol. 33, pp. 10718–10728, 2020.

[36] W. Gan *et al.*, "Self-Supervised Deep Equilibrium Models for Inverse Problems with Theoretical Guarantees," *IEEE Trans. Comput. Imag.,* vol. 9, pp. 796-807, Aug. 2023, doi: 10.1109/TCI.2023.3304475.

[37] A. Pramanik and M. Jacob, "Improved Model Based Deep Learning Using Monotone Operator Learning (Mol)," in *Proc. IEEE/ISBI*, pp. 1-4, Apr. 2022, doi: 10.1109/ISBI52829.2022.9761520.

[38] Y. Zhao *et al.*, "Deep Equilibrium Models for Snapshot Compressive Imaging," in *Proc. AAAI*, vol. 37, no. 3, pp. 3642–3650, Jun. 2023, doi: 10.1609/aaai.v37i3.25475.

[39] P. Qu *et al.*, "Convergence behavior of iterative SENSE reconstruction with non-Cartesian trajectories," *Magn. Reson. Med.*, vol. 54, no. 4, pp. 1040–1045, Oct. 2005, doi: 10.1002/mrm.20648.

[40] R. D. Peters and M. L. Wood, "Multilevel wavelet-transform encoding in MRI," *J. Magn. Reson. Imag.*, vol. 6, no. 3, pp. 529–540, May 1996, doi: 10.1002/jmri.1880060317.

[41] E. Hot and P. Sekulić, "Compressed sensing MRI using masked DCT and DFT measurements," in *Proc. MECO*, pp. 323–326, Aug. 2015, doi: 10.1109/MECO.2015.7181934.

[42] B. Zhang *et al.*, "Research on image processing with compressed sensing algorithm: Base on the improved layered discrete cosine transform," in *Proc. IEEE/ICC*, pp. 6357–6361, Nov. 2012, doi: 10.1109/ICC.2012.6364661.

[43] A. Godino-Moya *et al.*, "Space-time variant weighted regularization in compressed sensing cardiac cine MRI," *Magn. Reson. Imag.*, vol. 58, pp. 44–55, May 2019, doi: 10.1016/j.mri.2019.01.005.

[44] K. Hammernik *et al.*, "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, Jun. 2018, doi: 10.1002/mrm.26977.

[45] B. Zhou *et al.*, "Dual-domain self-supervised learning for accelerated non-Cartesian MRI reconstruction," *Med. Image Anal.*, vol. 81, Oct. 2022, doi: 10.1016/j.media.2022.102538.

[46] D. Rumelhart *et al.*, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.

[47] V. Sitzmann *et al.*, "Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations," in *Proc. NeurIPS*, vol. 32, 2019.

[48] S. M. A. Eslami *et al.*, "Neural scene representation and rendering," *Science*, vol. 360, no. 6394, pp. 1204-1210, Jun. 2018, doi: 10.1126/science.aar6170.

[49] L. Shen *et al.*, "NeRP: Implicit Neural Representation Learning With Prior Embedding for Sparsely Sampled Image Reconstruction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 770-782, Jan. 2024, doi: 10.1109/TNNLS.2022.3177134.

[50] W. Huang *et al.*, "Neural Implicit k-Space for Binning-Free Non-Cartesian Cardiac MR Imaging," in *Proc. IPMI*, vol. 13939, pp. 548-560, Jun. 2023, doi: 10.1007/978-3-031-34048-2_42.

[51] J. Feng *et al.*, "Spatiotemporal implicit neural representation for unsupervised dynamic MRI reconstruction," Dec. 2022, *arXiv:2301.00127v2*.

[52] S. Banach, "Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales," *Fund. Math.*, vol. 3, no. 1, pp. 133–181, 1922.

[53] C. Trabelsi *et al.*, "Deep Complex Networks," in *Proc. ICLR*, 2018.

[54] D. P. Kingma and J. Ba, "ADAM: A Method for Stochastic Optimization" in *Proc. ICLR*, May 2015.

[55] L. Feng *et al.*, "Golden-angle radial sparse parallel MRI: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI," *Magn. Reson. Med.*, vol. 72, no. 3, pp. 707–717, Sep. 2014, doi: 10.1002/mrm.24980.

[56] D. O. Walsh *et al.*, "Adaptive reconstruction of phased array MR imagery," *Magn. Reson. Med.*, vol. 43, no. 5, pp. 682-690, May 2000, doi:10.1002/(SICI)1522-2594(200005)43:5<682::AID-MRM10>3.0.CO;2-G.

[57] F. Huang *et al.*, "A software channel compression technique for faster reconstruction with many channels," *J. Magn. Reson. Imag.*, vol. 26, no. 1, pp. 133–141, Jan. 2008, doi: 10.1016/J.MRI.2007.04.010.

[58] S. Wundrak *et al.*, "Golden ratio sparse MRI using tiny golden angles," *Magn. Reson. Med.*, vol. 75, no. 6, pp. 2372–2378, Jun. 2016, doi: 10.1002/mrm.25831.

[59] N. R. Goodman, "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152-177, Mar. 1963, doi: 10.1214/AOMS/1177704250.

[60] Z. Wang *et al.*, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[61] F. Crete *et al.*, "The blur effect: perception and estimation with a new no-reference perceptual blur metric," in *Proc. SPIE*, vol. 6492, , Feb. 2007, doi: 10.1117/12.702790.

[62] L. Fan *et al.*, "Ultra-rapid, Free-breathing, Real-time Cardiac Cine MRI Using GRASP Amplified with View Sharing and KWIC Filtering," *Radiol. Cardiothorac. Imag.*, vol. 6, no. 1, Feb. 2024, doi: 10.1148/ryct.230107.

[63] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012, doi: 10.1109/TIP.2012.2197011.

[64] R. Soundararajan and A. C. Bovik, "Video Quality Assessment by Reduced Reference Spatio-Temporal Entropic Differencing," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 23, no. 4, pp. 684–694, Apr. 2013, doi: 10.1109/TCSVT.2012.2214933.

[65] M. J. Muckley *et al.*, "TorchKbNufft: A High-Level, Hardware-Agnostic Non-Uniform Fast Fourier Transform," in *ISMRM Workshop on Data Sampling & Image Reconstruction*, 2020.

[66] A. Dosovitskiy *et al.*, "Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1734–1747, Sep. 2016, doi: 10.1109/TPAMI.2015.2496141.

[67] Y. Zhou and Y. Wu, "Analyses on Influence of Training Data Set to Neural Network Supervised Learning Performance," in *Proc. AINSC*, vol. 106, pp. 19–25, 2011, doi: 10.1007/978-3-642-23753-9_4.

[68] M. O. Malavé *et al.*, "Reconstruction of undersampled 3D non-Cartesian image-based navigators for coronary MRA using an unrolled deep learning model," *Magn. Reson. Med.*, vol. 84, no. 2, pp. 800-812, Aug. 2020, doi: 10.1002/mrm.28177.

[69] J. Schlemper *et al.*, "Nonuniform Variational Network: Deep Learning for Accelerated Nonuniform MR Image Reconstruction," in *Proc. MICCAI*, pp. 57–64, Oct. 2019, doi: 10.1007/978-3-030-32248-9_7.

[70] M. Akcakaya *et al.*, "Unsupervised Deep Learning Methods for Biological Image Reconstruction and Enhancement: An overview from a signal processing perspective," *IEEE Signal Process. Mag.*, vol. 39, no. 2, pp. 28-44, Mar. 2022, doi: 10.1109/MSP.2021.3119273.

[71] G. Zeng *et al.*, "A review on deep learning MRI reconstruction without fully sampled k-space," *BMC Med. Imag.*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/S12880-021-00727-9.

[72] K. Wang *et al.*, "Memory-Efficient Learning for High-Dimensional MRI Reconstruction," in *Proc. MICCAI*, vol. 12906, pp. 461–470, Sep. 2021, doi: 10.1007/978-3-030-87231-1_45.

[73] M. Kellman *et al.*, "Memory-Efficient Learning for Large-Scale Computational Imaging," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1403–1414, Sep. 2020, doi: 10.1109/TCI.2020.3025735.

[74] Z. Cheng *et al.*, "Memory-Efficient Network for Large-scale Video Compressive Sensing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 16241–16250, Jun. 2021, doi: 10.1109/CVPR46437.2021.01598.

[75] Z. Yao *et al.*, "Cross-Iteration Batch Normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 12326-12335, Jun. 2021, doi: 10.1109/CVPR46437.2021.01215.