Dense Recurrent Neural Networks for Inverse Problems: History-Cognizant Unrolling of Optimization Algorithms

Seyed Amir Hossein Hosseini, Student Member, IEEE, Burhaneddin Yaman, Student Member, IEEE, Steen Moeller, Mingyi Hong, Member, IEEE and Mehmet Akçakaya, Member, IEEE

Abstract—Inverse problems in medical imaging applications incorporate domain-specific knowledge about the forward encoding operator in a regularized reconstruction framework. Recently, physics-driven deep learning (DL) methods have been proposed to use neural networks for data-driven regularization. These methods unroll iterative optimization algorithms to solve the inverse problem objective function, by alternating between domain-specific data consistency and data-driven regularization via neural networks. The whole unrolled network is then trained end-to-end to learn the parameters of the network. Due to simplicity of data consistency updates with gradient descent steps, proximal gradient descent (PGD) is a common approach to unroll physics-driven DL reconstruction methods. However, PGD methods have slow convergence rates, necessitating a higher number of unrolled iterations, leading to memory issues in training and slower reconstruction times in testing. Inspired by efficient variants of PGD methods that use a history of the previous iterates, we propose a history-cognizant unrolling of the optimization algorithm with dense connections across iterations for improved performance. In our approach, the gradient descent steps are calculated at a trainable combination of the outputs of all the previous regularization units. We also apply this idea to unrolling variable splitting methods with quadratic relaxation. Our results in reconstruction of the fastMRI knee dataset show that the proposed history-cognizant approach reduces residual aliasing artifacts compared to its conventional unrolled counterpart without requiring extra computational power or increasing reconstruction time.

Index Terms—Inverse problems, unrolled optimization algorithms, physics-driven deep learning, neural networks, MRI reconstruction, recurrent neural networks.

I. INTRODUCTION

Inverse problems have been extensively utilized in many imaging modalities, such as magnetic resonance imaging (MRI) [1–7], computed tomography (CT) [8–13], and optical imaging [14–17]. These inverse problems are directly driven from the physics of data acquisition, known as the forward operator, which incorporates domain-specific knowledge. Such inverse problems are typically ill-conditioned and thus regularizers are incorporated into the respective objective functions.

This work was supported by NIH P41EB027061, NIH U01EB025144, NSF CAREER CCF-1651825, NSF CMMI-1727757, CIF-1910385, ARO grant 73202-CS.

Subsequently, iterative algorithms are employed to solve the regularized inverse problem by alternating between enforcing physics-driven data consistency and regularization using pre-selected priors.

Recently, deep learning (DL) has emerged as an alternative state-of-the-art technique for solving such inverse problems in imaging applications [18–50]. Several of the earlier works concentrated on training convolutional neural networks (CNN) to solve the problem without incorporating explicit knowledge of the acquisition physics, typically requiring re-training when acquisition parameters are modified [18–30]. Another line of work focuses on a physics-driven formulation, where the known forward model is incorporated into training, utilizing domain knowledge to solve the corresponding inverse problem [31–45, 51]. In physics-driven approaches, a conventional iterative optimization problem, which lands itself to a regularized least-squares formulation for modalities such as MRI and CT, is unrolled for a fixed number of iterations. This unrolled network alternates between data consistency and regularization steps similar to the conventional algorithm, where the regularizers are implemented via neural networks. The unrolled network is trained end-to-end to learn the network parameters, which predominantly characterize the regularization units.

There are numerous approaches for solving regularized least squares problems, which have been explored for algorithm unrolling in physics-driven DL reconstruction methods [31–45, 51]. These methods include gradient descent (GD) [31], proximal gradient descent (PGD) [32–36], variable-splitting (VS) [37–42] and primal-dual (PD) [43–44] methods. GD, PGD and PD unrolling typically use a gradient descent step for data consistency [31–44], while for VS-based methods, either a gradient descent [38] or conjugate gradient (CG) data consistency can be used, with the latter being utilized when the least squares problem involving the forward encoding operator becomes computationally expensive [37–39, 42]. Data consistency updates with gradient descent step is generally computationally inexpensive. However, these optimization methods have slower convergence, which necessitates higher number of unrolled iterations for improved performance, which in turn can lead to numerical issues such as gradient vanishing and memory constraints.

In this study, we seek to develop a new methodology for unrolling inverse problem optimization algorithms into dense recurrent neural networks (RNN) for improved reconstruction performance. Motivated by accelerated PGD approaches,
A. Forward Model and Inverse Problem

Let \(y\) denote a vector of noisy measurements from an imaging system and \(x\) be the corresponding underlying image. The forward model for this system is generally given as \(y = Ax + n\), where \(A\) is called a forward operator and \(n\) is measurement noise. In a number of medical imaging scenarios, such as MRI and CT reconstruction, the forward model is linear, leading to a simplified model

\[ y = Ax + n, \tag{1} \]

where \(A\) is a linear forward operator and \(n\) is measurement noise, which will be the focus of the rest of the study. The corresponding inverse problem is typically ill-conditioned. Thus, regularization is frequently used, leading to a regularized least-squares method:

\[ \arg \min_x ||y - Ax||_2^2 + \mathcal{R}(x), \tag{2} \]

where the first term enforces consistency with measurement data and \(\mathcal{R}(\cdot)\) denotes a regularizer. In classical reconstruction approaches, regularizers have included Tikhonov \cite{1,2}, edge-preserving functions \cite{62} or sparsity-promoting terms in predefined domains \cite{3,4,6,7}. A variety of algorithms can iteratively solve the optimization problem in (2) such as PGD \cite{63}, VS \cite{64,66} and PD \cite{67,68} methods.

B. Physics-Based DL Reconstruction by Unrolling PGD

Several physics-based DL methods have unrolled a PGD-based method \cite{32,36} for a fixed number of iterations. Such PGD-based unrolling alternates between a proximal operation that is implicitly defined by a learned neural network, and a gradient descent step to enforce data consistency, as follows:

\[ z^{(i)} = \text{Prox}_R(x^{(i-1)}) = \arg \min_z ||x^{(i-1)} - z||_2^2 + \mathcal{R}(z) \tag{3a} \]

\[ x^{(i)} = z^{(i)} + \mu_i A^H(y - Az^{(i)}) \tag{3b} \]

where \(\mu_i\) is the gradient descent step size at unrolled iteration \(i\), and the proximal operation in (3a) is performed by a neural network. This leads to the schematic in Figure 1a, where the unrolled RNN consists of multiple blocks of proximal (P) and data consistency (DC) units. The former is a neural network and the latter implements a gradient descent step based on the known forward encoding operator. The unrolled RNN is then trained end-to-end to learn the mapping from the acquired sensor-domain data to reference images for medical imaging reconstruction.

C. Proposed History-Cognizant PGD Unrolling with Dense Recurrent Neural Networks

One of the major appeals of the PGD unrolling scheme is its computational simplicity. However, PGD algorithm itself has a slow convergence rate \cite{62}, potentially requiring a large number of iterations (i.e., a large number of blocks in Figure 1a) when unrolled into a neural network for sufficient reconstruction quality. In turn, this increases memory requirements and impedes training complexity. On the other hand, several methods have been proposed to accelerate convergence rate of PGD \cite{52,58}. The underlying theme for these methods is to calculate the gradient step not only at the most recent proximal operation output, but at a linear combination of all past such outputs.

We propose to adapt this general accelerated PGD approach for network unrolling. In this setting, the accelerated PGD methods involve the following steps:

\[ z^{(i)} = \text{Prox}_R(x^{(i-1)}) \tag{4a} \]

\[ v^{(i)} = F(z^{(i)}, \ldots, z^{(i)}) \tag{4b} \]

\[ x^{(i)} = v^{(i)} + \mu_i A^H(y - A v^{(i)}) \tag{4c} \]

where \(F(\cdot)\) forms a linear combination of the previous proximal operator outputs. As a special case, Nesterov’s method uses the previous two outputs \cite{53}, while more general versions have also been studied, in which all the past iterations are used \cite{54,58}.
For our unrolled network, we propose to learn \( \mathbf{F}(\cdot) \) during training instead of specifying a pre-determined form for the linear combination of previous iterates \( \mathbf{F}(\cdot) \) as in Nesterov’s method [63]. This history-cognizant unrolling of the accelerated PGD algorithms leads to a dense recurrent neural network architecture (Dense-RNN) with skip connections across all unrolled iterations, as depicted in Figure 1b. In addition to the P and DC units to perform proximal operation (or regularization) and data consistency updates, each unrolled iteration also consists of a \( 1 \times 1 \) convolutional layer implementing \( \mathbf{F}(\cdot) \) on the concatenation (denoted by C) of all past proximal operator outputs prior to the DC unit. We refer to this method as history-cognizant PGD (HC-PGD) in the rest of this study.

### III. Multi-Coil MRI Reconstruction Experiments and Results

The proposed history-cognizant algorithm unrolling approach was compared to conventional algorithm unrolling in multi-coil MRI reconstruction. In a multi-coil MRI system with \( n_c \) coils, \( \mathbf{y} \) is the sub-sampled noisy data from all coils and \( \mathbf{x} \) is the underlying image to be recovered. The forward linear operator \( \mathbf{A} : \mathbb{C}^{M \times N} \rightarrow \mathbb{C}^{P} \) in this system includes sensitivities of the receiver coil array and a partial Fourier matrix for sub-sampling in k-space, i.e. spatial Fourier domain [1].

#### A. Knee MRI Datasets

Knee MRI datasets were obtained from the New York University (NYU) fastMRI initiative database [61] to investigate accelerated MRI reconstruction using conventional unrolled RNN and Dense-RNN architectures. Data were acquired without any acceleration on a clinical 3T system (Magnetom Skyra, Siemens, Germany) with a 15-channel knee coil. Data acquisition was performed using 2D turbo spin-echo sequences in coronal orientation with proton-density (Coronal PD) and proton-density weighted with fat suppression (Coronal PD-FS) weightings. Relevant imaging parameters were: resolution \( = 0.49 \times 0.44 \text{mm}^2 \), slice thickness = 3mm , matrix size = 320 \times 368 for both datasets.

All data were uniformly sub-sampled at an acceleration rate of 4 by keeping the central 24 lines using the mask provided in the fastMRI database [61] retrospectively. Uniform undersampling for 2D acquisitions is typically considered a more challenging setup for DL reconstruction than random undersampling [31]. A 24 \times 24 central window was used to generate coil sensitivity maps using ESPIRiT [69]. The training sets consisted of 300 slices from 15 subjects for each sequence. Testing was performed on 10 different subjects for both sequences.

#### B. Implementation Details

The optimization algorithms were unrolled for \( T = 10 \) iterations. The proposed HC-PGD algorithm was implemented using the Dense-RNN architecture in Figure 1b. Proximal operator unit outputs are first concatenated and then combined using convolutional layers with a \( 1 \times 1 \) kernel size and 2 output channels (real and imaginary components) before being inputted to the subsequent data consistency unit. On the other hand, the PGD algorithm was implemented using a conventional architecture that only includes P and DC units, without skip connections, concatenation of previous iterates or \( 1 \times 1 \) convolutions.

The same CNN was used to implement the proximal operation for both PGD and HC-PGD. A residual network (ResNet) that has been established in a different regression problem was utilized [70]. This CNN consisted of 15 sub-residual blocks each having two convolutional layers of kernel size = \( 3 \times 3 \) and output channels = 64. The first and second convolutional layers in each sub-residual block are followed by a ReLU activation function and a scaling factor of 0.1, respectively. In addition to sub-residual blocks, two input and output convolutional layers without any nonlinear activation function are to match number of desired channels. For both algorithm unrolling schemes, the proximal units shared parameters across unrolled iterations [39], leading to a total of 592,138 and 592,358 parameters for PGD and HC-PGD, respectively.

All networks were trained end-to-end, as summarized in Figure 2 by using Adam optimizer to minimize

\[
\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathbf{x}^i_{\text{ref}}, f(\mathbf{y}^i, \mathbf{A}^i; \theta)),
\]

where \( \mathbf{x}^i_{\text{ref}}, \mathbf{y}^i \) and \( \mathbf{A}^i \) are the reference SENSE-1 image, undersampled data and forward encoding operator for slice \( i \), respectively and \( N \) is the number of training slices. \( f(\mathbf{y}^i, \mathbf{A}^i; \theta) \) denotes the network output for slice \( i \) with the corresponding measurements and forward encoding operator, while \( \theta \) contains the trainable parameters of the networks. A normalized \( \ell_1 - \ell_2 \) loss [49] was used for \( \mathcal{L}(\cdot, \cdot) \):

\[
\mathcal{L}(\mathbf{u}, \mathbf{v}) = \frac{||\mathbf{u} - \mathbf{v}||_2}{||\mathbf{u}||_2} + \frac{||\mathbf{u} - \mathbf{v}||_1}{||\mathbf{u}||_1},
\]

with \( \mathbf{u} \) and \( \mathbf{v} \) representing reference fully-sampled and network output images, respectively. Training was performed over 100 epochs. All algorithms were implemented using
TensorFlow, and processed on a workstation with an Intel E5-2640V3 CPU (2.6GHz and 256GB memory), and two NVIDIA Tesla V100 GPUs with 32GB memory.

C. Performance of HC-PGD versus PGD Unrolling

Experiments were conducted on both Coronal PD and PD-FS datasets to compare the reconstruction quality of the conventional PGD unrolling and the proposed HC-PGD unrolling with the Dense-RNN architecture. A total of 380 slices from 10 subjects (excluded from training) were utilized for testing. Reconstructed testing data were evaluated quantitatively using normalized mean square error (NMSE) and structural similarity index (SSIM) metrics with respect to the fully-sampled reference image.

Figure 3 depicts a representative slice from the Coronal PD dataset reconstructed using DL-MRI with PGD and HC-PGD unrolling methods. The first row shows the reference fully-sampled image, as well as the zero-filled acquired image, which serves as the input $x^{(0)}$ to the network. Difference images with respect to the reference image, scaled by a factor of 10, are also provided to facilitate comparison by showing errors. DL-MRI based on the proposed HC-PGD unrolling was able to suppress the strong aliasing artifacts successfully, while residual aliasing is visible on the DL-MRI reconstructions based on conventional PGD unrolling. For both methods, there is a slight degree of blurring in the final images, consistent with previous reports on DL-MRI methods with similar training loss functions [31]. Quantitative evaluations, shown in the lower left corner of the images, also confirm the improvement using the proposed history-cognizant unrolling with Dense-RNN architecture compared to the conventional algorithm unrolling.

The reconstructed images for a representative slice from the coronal PD-FS dataset in addition to the difference images scaled by a factor of 10, as well as the reference and zero-filled images are shown in Figure 4. There are slight residual aliasing artifacts in the DL-MRI reconstruction with PGD unrolling. The proposed HC-PGD unrolling successfully removes these artifacts further. The quantitative metrics also show improvement for the reconstruction results in moving from PGD to HC-PGD unrolling.

Figure 5 summarizes the quantitative analysis of the DL-MRI reconstructions using NMSE and SSIM metrics for the PGD and HC-PGD unrollings for both the coronal PD and PD-FS datasets. The box plots depict the interquartile range with the center at the median of the metric for the corresponding reconstruction methods. Both metrics are visibly improved over the whole test sets when using the proposed history-cognizant algorithm unrolling over the conventional unrolling approach. Further statistical tests were performed using the Wilcoxon signed rank test with a significance level of $P < 0.05$, indicating that all differences are statistically significant based on both metrics and in both datasets.

D. Insensitivity of Dense-RNN Improvement to the Regularization CNN Choice

In order to verify that the improvements between PGD and HC-PGD unrolling are not limited to the choice of CNN used for the proximal operation, the two approaches in Figure 1 were also implemented using a U-Net architecture for the regularization CNN. This was based on the U-Net used in
Fig. 4: A representative slice from the Coronal PD-FS dataset with a 4-fold acceleration rate using a uniform undersampling pattern, reconstructed with conventional PGD and the proposed HC-PGD methods, along with the error images scaled by a factor of 10. The green rectangle in the reference image mark the zoom-up area. Proposed HC-PGD unrolling successfully removes some residual artifacts that are visible in the PGD. The quantitative metrics align with this observation.

E. Performance Evaluation for Random Undersampling

In order to evaluate performance of the proposed history-cognizant unrolling approach in random undersampling, we utilized an observation from a previous study [31, 71], showing that physics-based networks trained on uniformly undersampled data can be transferred to random undersampling patterns at the same rate directly, even though the opposite is not necessarily true. Thus, we employed the networks trained on uniformly undersampled data, as detailed in Section III-B, to reconstruct test slices which were randomly undersampled at a similar 4-fold rate. The corresponding results, shown in Figure 7 indicate that the improvement trend remains similar to the uniform sampling scenario with the proposed HC-PGD method outperforming the conventional PGD in suppressing residual artifacts as observed in both Coronal PD and PD-FS datasets, which also align with the quantitative evaluation metrics in the images.

F. Extensions to Other Optimization Algorithms

While PGD unrolling is commonly used for DL-MRI reconstruction [32–50], alternative algorithms have also been utilized [37–42]. One such method for solving the objective function in Equation (2) is VS with quadratic penalty, which decouples data consistency and regularization terms into two blocks of variables [62] by introducing auxiliary variables to formulate a constrained objective function. This is then relaxed to an unconstrained problem by enforcing similarity between these variables via a quadratic penalty:

$$\arg\min_{x, z} \|y - Ax\|_2^2 + R(z) + \beta\|x - z\|_2^2,$$

where $z$ is an auxiliary variable and $\beta$ is the quadratic penalty parameter. When this algorithm is unrolled as a neural network [38–42], reconstruction alternates between a proximal operation that is again learned by a neural network and data consistency, similar to PGD unrolling, as follows:

$$z^{(i)} = \text{Prox}_R(x^{(i-1)})$$

$$x^{(i)} = (A^H A + \mu I)^{-1}(A^H y + \beta z^{(i)})$$
Fig. 6: Results of changing the proximal operator CNN, with a slice from the Coronal PD dataset with 4-fold uniform undersampling, reconstructed with DL-MRI reconstruction using the PGD and HC-PGD methods unrolled in a network with a U-Net design for the proximal operator units, along with Error images ($\times 5$), with the green rectangle marking the zoom-up area. Similar to the ResNet experiments, the proposed HC-PGD outperforms PGD in suppressing visible residual artifacts.

where regularization update in (6a) is performed by a neural network and $\beta$ was absorbed into the regularizer for notational convenience. The data consistency update in (6b) is computationally challenging in many practical scenarios, such as multi-coil MRI reconstruction. Thus, this data consistency update can be solved iteratively via conjugate gradient (CG) to avoid matrix inversion [39].

As an extension to the HC-PGD approach, VS with quadratic relaxation can also be unrolled in a history-cognizant manner using the Dense-RNN in Figure 1 building on the optimization approaches in [56, 57]. This will be referred to as the history-cognizant VS method (HC-VS). We note that the main difference to HC-PGD is that the DC units in HC-VS now implement an unrolled CG algorithm for a different least squares problem instead of a gradient step. To test the applicability of the history-cognizant unrolling scheme to different baseline algorithms, both VS and HC-VS were also implemented on the coronal PD and PD-FS datasets, using the setup described in Section III-B. In each data consistency unit, CG was implemented with 10 internal iterations.

Figure 8 depicts representative slices from the Coronal PD and Coronal PD-FS datasets reconstructed using DL-MRI with the VS and HC-VS unrolling methods along with the error images scaled by a factor of 10 with respect to the reference fully-sampled image, as well as the reference and zero-filled images. The HC-VS approach successfully removes residual artifacts apparent in the VS results as shown by the yellow arrows. Improvements are also observed in the quantitative metrics.

Fig. 7: Results on random undersampling experiments from (a) the Coronal PD and (b) Coronal PD-FS datasets, at a 4-fold acceleration rate, reconstructed with DL-MRI reconstruction based on the conventional PGD and the proposed HC-PGD unrolling, as well as error images ($\times 10$), with the green rectangle showing the zoom-up area. Regardless of the undersampling pattern, the history-cognizant unrolling approach, HC-PGD suppresses the residual aliasing visible in the conventional PGD unrolling, as marked by the arrows. The quantitative metrics align with these observations.
In this study, we proposed a history-cognizant method for unrolling optimization problems to solve regularized inverse problems in medical imaging using physics-based DL reconstruction. Our unrolling method builds on optimization approaches that utilize a history of the previous iterates via their linear combinations in updating the current estimate [52–57]. While these optimization procedures have been traditionally proposed to improve convergence rates or provide stronger theoretical guarantees, in the context of algorithm unrolling for physics-based DL reconstruction, they are used to enhance reconstruction quality without incurring additional computational complexity. Our method was pre-dominantly motivated by accelerated PGD approaches [52–55] and was shown to improve upon conventional PGD unrolling for multi-coil MRI reconstruction for multiple scenarios. We also explored how the unrolling can be applied to VS with quadratic relaxation for solving a regularized least squares problem, and showed that it can improve over the conventional unrolling approach.

The proposed history-cognizant unrolling methodology leads to several skip connections across iterations, leading to a densely-connected unrolled neural network. Dense neural networks have been shown to improve training by facilitating information flow through the network in forward propagation and by relieving issues concerned with gradient vanishing in backward propagation [60]. One of the advantages of these dense connections in the proposed unrolling is that the only computational overhead compared to conventional unrolling is the calculation of the $1 \times 1$ convolutions from the previous iterations. Thus, once trained, the computational time is increased <1.5% compared to the conventional counterpart, as depicted in Table I.

One of the most commonly used accelerated PGD approaches is based on Nesterov acceleration [53]. In this case, the gradient descent is evaluated at a linear combination of the previous two iterates weighted by a step size $\mu$ and $1 - \mu$ respectively. Theoretically, this approach achieves the optimal convergence rate for first-order methods [53]. As such, Nesterov-type unrolling for DL reconstruction was readily explored in [46] for coded-illumination microscopy, where the unrolled accelerated PGD network used the history of the two previous outputs of the proximal units. However, the main focus of this work was developing an illumination pattern design algorithm for a fixed reconstruction algorithm. On the other hand, when focusing on the reconstruction aspect for inverse problems, we found that the Nesterov-type unrolling was outperformed by the full history-cognizant Dense-RNN architecture in our initial experiments (not shown). Furthermore, the full history-cognizant unrolling subsumes Nesterov-type as a special case. Thus, we focused on this full history-cognizant case throughout our study to maximize the performance benefits.

In our experiments, the history-cognizant unrolling visibly improved the performance of PGD-based methods compared to its conventional counterpart for all datasets. In the VS setting, especially for the coronal PD datasets, the improvement is less pronounced. Coronal PD datasets have higher SNR compared to PD-FS datasets [31]. In this higher SNR case, the conventional unrolling of the VS algorithm readily provides a high-quality reconstruction. Thus, additional improvement with HC-VS is less apparent than in the PGD setting.

Our experiments in multi-coil MRI reconstruction pre-
dominantly utilized uniform undersampling patterns. While random undersampling patterns are commonly used in literature for training and testing DL-based reconstruction, uniform undersampling remains the most clinically used approach, and was considered to be a harder problem compared to random undersampling in [31] [71], as detailed in Section III-E.

To the best of our knowledge, the Dense-RNN architecture that arises from the proposed history-cognizant unrolling have not been used for regression problems in general, and inverse problems in particular. However, Dense-RNNs have been independently proposed for different problems, such as scene parsing or labeling [72] where semantic information of an image segment is to be extracted. In such applications, RNNs are used for each image unit to receive the dependencies from other units through recurrent information forwarding between adjacent units, which decays for further units. In this setting, Dense-RNNs improve labeling by capturing long-range semantic dependencies among image units, highlighting a similar memory dependence as in this study.

V. CONCLUSION

We proposed a history-cognizant approach to unroll a given optimization algorithm for solving inverse problems in medical imaging using physics-driven deep learning. The proposed approach leads to an unrolled dense recurrent neural network architecture. From a network-design perspective, this consequent architecture enhances training and performance by facilitating information flow through the network. Accelerated MRI reconstruction experiments show that the proposed unrolling approach may considerably enhance reconstruction quality without substantially modifying computational complexity.

ACKNOWLEDGMENT

MRI data were obtained from the NYU fastMRI initiative database [61] which was acquired with the relevant institutional review board approvals as detailed in [61]. NYU fastMRI investigators, subject to updates, can be found at fastmri.med.nyu.edu.

REFERENCES

[1] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger, “Advances in sensitivity encoding with arbitrary k-space trajectories,” Magn Reson Med, vol. 46, no. 4, pp. 638–651, 2001.
[2] B. P. Sutton, D. C. Noll, and J. A. Fessler, “Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities,” IEEE Trans Med Imaging, vol. 22, no. 2, pp. 178–188, 2003.
[3] M. Lustig, D. Donoho, and J. Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” Magn Reson Med, vol. 58, pp. 1182–1195, 2007.
[4] K. T. Block, M. Uecker, and J. Frahm, “Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint,” Magn Reson Med, vol. 57, no. 6, pp. 1086–1098, 2007.
[5] J. A. Fessler, “Model-based image reconstruction for MRI,” IEEE Sig Proc Magazine, vol. 27, no. 4, pp. 81–89, 2010.
[6] M. Akçakaya, S. Nam, et al., “Compressed sensing with wavelet domain dependencies for coronary MRI: a retrospective study,” IEEE Trans Med Imaging, vol. 30, no. 5, pp. 1090–1099, 2010.
[7] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, “Second order total generalized variation (TGV) for MRI,” Magn Reson Med, vol. 65, no. 2, pp. 480–491, 2011.
[8] G. Wang, D. L. Snyder, J. A. O’Sullivan, and M. W. Vannier, “Iterative deblurring for ct metal artifact reduction,” IEEE Trans Med Imaging, vol. 15, pp. 657–664, 1996.
[9] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, “A three-dimensional statistical approach to improved image quality for multislice helical CT,” Medical physics, vol. 34, no. 11, pp. 4526–4544, 2007.
[10] P. Jin, C. A. Bouman, and K. D. Sauer, “A model-based image reconstruction algorithm with simultaneous beam hardening correction for X-ray CT,” IEEE Trans Comp Imaging, vol. 1, no. 3, pp. 200–216, 2015.
[11] L. Pfister and Y. Bresler, “Model-based iterative tomographic reconstruction with adaptive sparsifying transforms,” in Comp Imaging XII, 2014, vol. 9020, p. 90200H.
[12] Z. Chang, D. H. Ye, et al., “Prior-guided metal artifact reduction for iterative X-ray computed tomography,” IEEE Trans Med Imaging, vol. 38, pp. 1532–1542, 2018.
[13] A. Tuysuzoglu, Y. Khoo, and W. C. Karl, “Variable splitting techniques for discrete tomography,” in IEEE Int Conf Image Proc, 2016, pp. 1764–1768.
[14] L. Tian and L. Waller, “Quantitative differential phase contrast imaging in an LED array microscope,” Optics Express, vol. 23, no. 9, pp. 11394–11403, 2015.
[15] R. A. Claus, P. P. Naulleau, A. R. Neureuther, and L. Waller, “Quantitative phase retrieval with arbitrary pupil and illumination,” Optics Express, vol. 23, no. 20, pp. 26672–26682, 2015.
[16] Y.-Z. Lin, K.-Y. Huang, and Y. Luo, “Quantitative differential phase contrast imaging at high resolution with radially asymmetric illumination,” Optics letters, vol. 43, no. 12, pp. 2973–2976, 2018.
[17] S. Sreehari, S. V. Venkatakrishnan, et al., “Plug-and-play priors for bright field electron tomography and sparse interpolation,” IEEE Trans Comp Imaging, vol. 2, no. 4, pp. 408–423, 2016.
[18] S. Wang, Z. Su, et al., “Accelerating magnetic resonance imaging via deep learning,” in Proc IEEE ISBI, 2016.

|                | PGD | HC-PGD | VS | HC-VS |
|----------------|-----|--------|----|------|
| Coronal PD     | 151 ± 4.2 ms | 197 ± 7.2 ms | 240 ± 0.5 ms | 325 ± 6.1 ms |
| Coronal PD-FS  | 197 ± 6.6 ms | 198 ± 6.6 ms | 325 ± 6.1 ms | 325 ± 6.2 ms |

TABLE I: Average and standard deviation (in ms) of reconstruction time over all test slices for PGD, HC-PGD, VS and HC-VS in Coronal PD and Coronal PD-FS datasets. History cognizant unrolling has very slight impact on reconstruction time compared to its conventional counterpart (< 1.5%). Furthermore, VS methods, which utilize CG-based data consistency units, are considerably slower than PGD methods.
pp. 514–517.

[19] K. Kwon, D. Kim, and H. Park, “A parallel MR imaging method using multilayer perceptron,” *Medical Physics*, vol. 44, no. 12, pp. 6209–6224, 2017.

[20] D. Lee, J. Yoo, S. Tak, and J. C. Ye, “Deep residual learning for accelerated MRI using magnitude and phase networks,” *IEEE Trans Biomed Eng*, vol. 65, no. 9, pp. 1985–1995, 2018.

[21] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, “Image reconstruction is a new frontier of machine learning,” *IEEE Trans Med Imaging*, vol. 37, pp. 1289–1296, 2018.

[22] S. H. Dar and T. Çukur, “Transfer learning for reconstruction of accelerated MRI acquisitions via neural networks,” Paris, France, 2018, Proc ISMRM.

[23] H. Chen, Y. Zhang, et al., “Low-dose CT with a residual encoder-decoder convolutional neural network,” *IEEE Trans Med Imaging*, vol. 36, pp. 2524–2535, 2017.

[24] Y. Han, L. Sunwoo, and J. C. Ye, “k-space deep learning for accelerated MRI,” *IEEE Trans Med Imaging*, 2019.

[25] Y. Han, J. Yoo, et al., “Deep learning with domain adaptation for accelerated projection-reconstruction MR,” *Magn Reson Med*, vol. 80, pp. 1189–1205, 2018.

[26] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, “Efficient b-mode ultrasound image reconstruction from sub-sampled RF data using deep learning,” *IEEE Trans Med Imaging*, vol. 38, no. 2, pp. 325–336, 2018.

[27] M. U. Ghani and W. C. Karl, “Fast enhanced CT metal artifact reduction using data domain deep learning,” *IEEE Trans Comp Imaging*, 2019.

[28] Q. Yang, P. Yan, et al., “Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss,” *IEEE Trans Med Imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.

[29] E. Kang, W. Chang, J. Yoo, and J. C. Ye, “Deep convolutional framelet denosing for low-dose CT via wavelet residual network,” *IEEE Trans Med Imaging*, vol. 37, no. 6, pp. 1358–1369, 2018.

[30] Z. Wu, Y. Sun, et al., “Simba: Scalable inversion in optical tomography using deep denoising priors,” *preprint arXiv:1911.13241*, 2019.

[31] K. Hammernik, T. Klatzer, et al., “Learning a variational network for reconstruction of accelerated MRI data,” *Magn Reson Med*, vol. 79, pp. 3055–3071, 2018.

[32] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, “A deep cascade of convolutional neural networks for dynamic MR image reconstruction,” *IEEE Trans Med Imaging*, vol. 37, no. 2, pp. 491–503, 2017.

[33] M. Mardani, H. Monajemi, et al., “Recurrent generative adversarial networks for proximal learning and automated compressive image recovery,” *preprint arXiv:1711.10046*, 2017.

[34] J. Zhang and B. Ghanem, “ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing,” in *Proc IEEE CVPR*, 2018, pp. 1828–1837.

[35] E. Kobler, M. Muckley, et al., “Variational deep learning for low-dose computed tomography,” in *Proc IEEE ICASSP*, 2018, pp. 6687–6691.

[36] M. Kellman, E. Bostan, N. Repina, and L. Waller, “Physics-based learned design: Optimized coded-illumination for quantitative phase imaging,” *IEEE Trans Comp Imaging*, 2019.

[37] Y. Yang, J. Sun, H. Li, and Z. Xu, “Deep ADMM-Net for compressive sensing MRI,” in *Advances in neural information processing systems*, 2016, pp. 10–18.

[38] D. H. Ye, S. Srivastava, J.-B. Thibault, K. Sauer, and C. Bouman, “Deep residual learning for model-based iterative CT reconstruction using plug-and-play framework,” in *Proc IEEE ICASSP*, 2018, pp. 6668–6672.

[39] H. K. Aggarwal, M. P. Mani, and M. Jacob, “Modl: Model-based deep leaning architecture for inverse problems,” *IEEE Trans Med Imaging*, vol. 38, no. 2, pp. 394–405, 2018.

[40] C. Qin, J. Schlemper, et al., “Convolutional recurrent neural networks for dynamic MR image reconstruction,” *IEEE Trans Med Imaging*, vol. 38, pp. 280–290, 2018.

[41] J. Duan, J. Schlemper, et al., “VS-Net: Variable splitting network for accelerated parallel MRI reconstruction,” in *Proc MICCAI*, 2019, pp. 713–722.

[42] B. Yaman, S. A. H. Hosseini, et al., “Self-supervised physics-based deep learning MRI reconstruction without fully-sampled data,” *preprint arXiv:1910.09116*, 2019.

[43] J. Adler and O. Öktem, “Learned Primal-Dual Reconstruction,” *IEEE Trans Med Imaging*, vol. 37, no. 6, pp. 1322–1332, 06 2018.

[44] J. Cheng, H. Wang, L. Ying, and D. Liang, “Model learning: Primal dual networks for fast MR imaging,” in *Proc MICCAI*, 2019, pp. 21–29.

[45] E. Bostan, U. S. Kamilov, and L. Waller, “Learning-based image reconstruction via parallel proximal algorithm,” *IEEE Sig Proc Letters*, vol. 25, no. 7, pp. 989–993, 2018.

[46] M. Akçakaya, S. Moeller, S. Weingärtner, and K. Üğurbil, “Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging,” *Magn Reson Med*, vol. 81, no. 1, pp. 439–453, 2019.

[47] S. A. H. Hosseini, C. Zhang, et al., “Accelerated coronary MRI with sRAKI: A database-free self-consistent neural network k-space reconstruction for arbitrary undersampling,” *preprint arXiv:1907.08137*, 2019.

[48] S. A. H. Hosseini, C. Zhang, K. Üğurbil, S. Moeller, and M. Akçakaya, “sRAKI-RNN: accelerated MRI with scan-specific recurrent neural networks using densely connected blocks,” in *SPIE Wavelets and Sparsity XVIII*, 2019, p. 111381B.

[49] F. Knoll, K. Hammernik, et al., “Deep learning methods for parallel magnetic resonance image reconstruction,” *preprint arXiv:1904.01112*, 2019.

[50] G. Wang, “A perspective on deep imaging,” *IEEE Access*, vol. 4, pp. 8914–8924, 2016.

[51] U. S. Kamilov and H. Mansour, “Learning optimal nonlinearities for iterative thresholding algorithms,” *IEEE Sig Proc Letters*, vol. 23, no. 5, pp. 747–751, 2016.

[52] Y. Nesterov, “Smooth minimization of non-smooth functions,” *Mathematical programming*, vol. 103, no. 1,
[53] Y. Nesterov, “Gradient methods for minimizing composite functions,” *Mathematical Programming*, vol. 140, no. 1, pp. 125–161, 2013.

[54] Y. Nesterov and M. Florea, “Gradient methods with memory,” Tech. Rep., 2019.

[55] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J Imag Sci*, vol. 2, no. 1, pp. 183–202, 2009.

[56] A. Chambolle and T. Pock, “A remark on accelerated block coordinate descent for computing the proximity operators of a sum of convex functions,” 2015, technical report.

[57] M. Hong, X. Wang, M. Razaviyayn, and Z.-Q. Luo, “Iteration complexity analysis of block coordinate descent methods,” *Mathematical Programming Series A*, 2016.

[58] P. Tseng, “On accelerated proximal gradient methods for convex-concave optimization,” 2008, preprint.

[59] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[60] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proc IEEE CVPR*, 2017, pp. 4700–4708.

[61] J. Zbontar, F. Knoll, et al., “fastMRI: An open dataset and benchmarks for accelerated MRI,” preprint arXiv:1811.08839, 2018.

[62] J. A. Fessler, “Optimization methods for MR image reconstruction,” preprint arXiv:1903.03510, 2019.

[63] P. L. Combettes and J.-C. Pesquet, “Proximal splitting methods in signal processing,” in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

[64] T. Goldstein and S. Osher, “The split bregman method for L1-regularized problems,” *SIAM journal on imaging sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[65] S. Ramani and J. A. Fessler, “Parallel MR image reconstruction using augmented lagrangian methods,” *IEEE Trans Med Imaging*, vol. 30, no. 3, pp. 694–706, 2010.

[66] S. Boyd, N. Parikh, et al., “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[67] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journal of mathematical imaging and vision*, vol. 40, no. 1, pp. 120–145, 2011.

[68] T. Valkonen, “A primal–dual hybrid gradient method for nonlinear operators with applications to MRI,” *Inverse Problems*, vol. 30, no. 5, pp. 055012, 2014.

[69] M. Uecker, P. Lai, et al., “ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA,” *Magn Reson Med*, vol. 71, no. 3, pp. 990–1001, 2014.

[70] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang, “Nüre 2017 challenge on single image super-resolution: Methods and results,” in *Proc IEEE CVPR*, 2017, pp. 114–125.

[71] F. Knoll, K. Hammernik, et al., “Assessment of the generalization of learned image reconstruction and the potential for transfer learning,” *Magn Reson Med*, vol. 81, pp. 116–128, 2019.

[72] H. Fan, P. Chu, L. J. Latecki, and H. Ling, “Scene parsing via dense recurrent neural networks with attentional selection,” in *Proc IEEE WACV*, 2019, pp. 1816–1825.