2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Compressed Sensing MRI with ℓ_1-Wavelet Reconstruction Revisited Using Modern Data Science Tools

Hongyi Gu[1,2], Burhaneddin Yaman[1,2], Kamil Ugurbil[2], Steen Moeller[2] and Mehmet Akçakaya[1,2]

***Abstract*— Deep learning (DL) has emerged as a powerful tool for improving the reconstruction quality of accelerated MRI. These methods usually show enhanced performance compared to conventional methods, such as compressed sensing (CS) and parallel imaging. However, in most scenarios, CS is implemented with two or three empirically-tuned hyperparameters, while a plethora of advanced data science tools are used in DL. In this work, we revisit ℓ_1-wavelet CS for accelerated MRI using modern data science tools. By using tools like algorithm unrolling and end-to-end training with stochastic gradient descent over large databases that DL algorithms utilize, and combining these with conventional concepts like wavelet subband processing and reweighted ℓ_1 minimization, we show that ℓ_1-wavelet CS can be fine-tuned to a level comparable to DL methods. While DL uses hundreds of thousands of parameters, the proposed optimized ℓ_1-wavelet CS with sub-band training and reweighting uses only 128 parameters, and employs a fully-explainable convex reconstruction model.**

I. INTRODUCTION

Slow data acquisition remains a challenge for MRI, requiring accelerated imaging strategies. Conventional methods, such as parallel imaging [1], [2] and compressed sensing (CS) [3] are used clinically, but typically their acceleration rates are limited by noise amplification and residual aliasing artifacts in reconstructed images. Recently, deep learning (DL) methods for accelerated MRI [4]–[8] have emerged as a powerful tool for MRI reconstruction, with improved performance over conventional methods in many studies. Among DL methods, physics-guided DL (PG-DL) methods that unroll conventional optimization algorithms that incorporate the encoding operator have received attention [6]–[9]. While CS uses a linear transform-based representation of images for regularization, PG-DL methods utilize a non-linear representation for regularization, which is implicitly learned through neural networks.

DL reconstruction methods are trained using large databases, include a large number (usually more than hundreds of thousands or millions [6], [10], [11]) of parameters, incorporate sophisticated optimization algorithms for training [12], and utilize state-of-the-art loss functions [13], [14]. On the other hand, when CS reconstruction methods are implemented for comparison, they typically use two or three parameters, which are frequently hand-tuned using a simple grid search. Although some automatic tuning methods have been proposed [15], [16], these have not leveraged the widely-available and popular modern data science tools from the DL era.

In this work, we use these data science tools to revisit ℓ_1-wavelet CS for accelerated MRI. Similar to PG-DL methods, we unroll an ADMM algorithm and train it end-to-end, while using only 4 orthogonal wavelet bases for the regularizing transforms for a total of 12 tunable parameters. Building on this naive model, we further incorporate processing of each individual wavelet subband [17], and reweighted ℓ_1 minimization [18], leading to 64 and 128 tunable parameters respectively. Results show that even the naive model closes the gap in reconstruction performance to advanced PG-DL methods, while the incorporation of subband and reweighting further improves the quality of reconstruction to a level comparable to PG-DL methods. All the models proposed here enable a linear representation for *interpretable and convex* sparse image reconstruction at inference time.

II. MATERIALS AND METHODS

A. Inverse Problem for Accelerated MRI

The forward model for accelerated MRI is given as

$$\mathbf{y} = \mathbf{E}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^N$ is the image to be reconstructed, $\mathbf{y} \in \mathbb{C}^M$ is the undersampled k-space data from all coils, $\mathbf{E} : \mathbb{C}^N \to \mathbb{C}^M$ is the linear forward encoding operator containing coil sensitivity maps and partial Fourier matrix for undersampling in k-space [19], and $\mathbf{n}$ is the measurement noise. The inverse problem involves solving the objective function:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{E}\mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}) \tag{2}$$

where $\|\mathbf{y} - \mathbf{E}\mathbf{x}\|_2^2$ enforces data consistency (DC) and $\mathcal{R}(\mathbf{x})$ is a regularizer.

In conventional CS MRI reconstruction, the form of $\mathcal{R}(\mathbf{x})$ is often a weighted ℓ_1-norm of transform coefficients, i.e. $\mathcal{R}(\mathbf{x}) = \sum_{l=1}^{L} \lambda_l \|\mathbf{W}_l\mathbf{x}\|_1$, where $\mathbf{W}_l$ is a pre-specified linear (often orthogonal) transform, such as a discrete wavelet transform (DWT) [3], and L is the number of linear transforms used for regularization. The resulting convex objective function is solved via an iterative optimization algorithm [20]. These algorithms are conventionally run until a stopping criterion is met, making hyperparameter tuning difficult.

On the other hand, in PG-DL reconstruction, the inverse problem is usually solved by unrolling an iterative optimization algorithm for a fixed number of iterations [21], [22]. Typically, the solutions are decoupled to a series of regularizer and DC units. The regularizer in PG-DL is

[1]Center for Magnetic Resonance Research and [2]Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA. e-mails: {gu003818, yaman013, ugurb001, moell018, akcakaya}@umn.edu

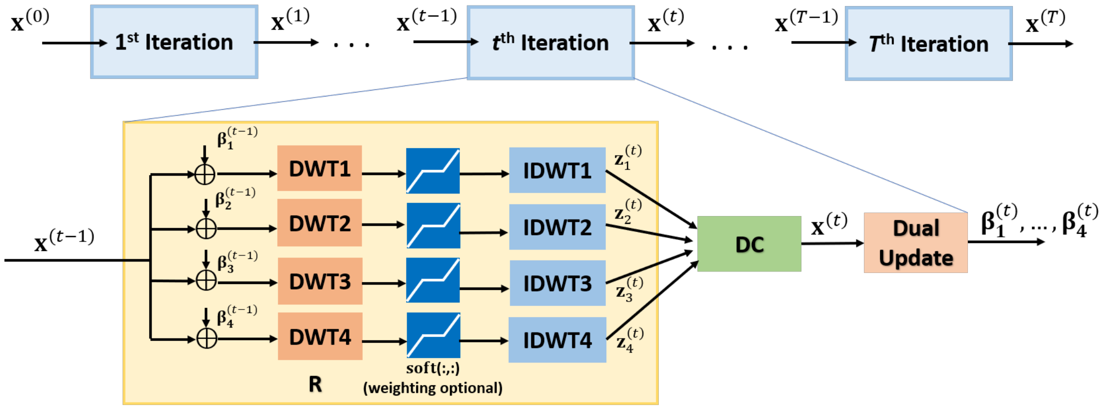

Fig. 1: Schematic of the unrolled ADMM for ℓ_1-wavelet compressed sensing (CS). One unrolled iteration of ADMM with ℓ_1-wavelet regularizers consists of regularizer (R), data consistency (DC) and dual update (DWT: Discrete wavelet transform). In accordance with the ADMM framework, learnable parameters are shared across different unrolled iterations. In its simplest form, this leads to 3 trainable parameters per wavelet. Further enhancements, such as separate thresholds for wavelet subbands and reweighted ℓ_1 minimization increase this number to 16 trainable parameters per wavelet.

implemented implicitly via convolutional neural networks (CNNs), while the DC unit is solved by linear methods, such as gradient descent or conjugate gradient [7]. The network is trained end-to-end as:

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}\Big(\mathbf{y}_{\text{ref}}^n,\ \mathbf{E}_{\text{full}}^n\big(f(\mathbf{y}^n, \mathbf{E}^n; \boldsymbol{\theta})\big)\Big), \tag{3}$$

where $\mathbf{y}_{\text{ref}}^n$ denotes the fully-sampled reference k-space of the n^{th} subject, $f(\mathbf{y}^n, \mathbf{E}^n; \boldsymbol{\theta})$ denotes network output of the unrolled network with parameters $\boldsymbol{\theta}$ of the n^{th} subject, $\mathbf{E}_{\text{full}}^n$ is the fully sampled multi-coil encoding operator of the n^{th} subject, N is the number of datasets in the training database, and $\mathcal{L}(\cdot,\cdot)$ is a loss function between the network output and the reference. Common choices for $\mathcal{L}(\cdot,\cdot)$ include ℓ_2 norm, ℓ_1 norm, mixed norms and perception-based loss [8], [14].

B. Proposed Learning of ℓ_1-Wavelet CS Reconstruction

We optimize conventional ℓ_1-Wavelet CS reconstruction by using the data science tools utilized in PG-DL techniques. First, we utilize ADMM to set

$$\mathbf{x}^{(t+1)} = \left(\mathbf{E}^H\mathbf{E} + \sum_{l=1}^{L} \rho_l \mathbf{I}\right)^{-1} \left(\mathbf{E}^H\mathbf{y} + \sum_{l=1}^{L} \rho_l \mathbf{W}_l^H \left(\mathbf{z}_l^{(t)} - \boldsymbol{\beta}_l^{(t)}\right)\right) \tag{4a}$$

$$\mathbf{z}_l^{(t+1)} = \text{soft}\left(\mathbf{W}_l\mathbf{x}^{(t+1)} + \boldsymbol{\beta}_l^{(t)}; \lambda_l/\rho_l\right) \tag{4b}$$

$$\boldsymbol{\beta}_l^{(t+1)} = \boldsymbol{\beta}_l^{(t)} + \eta_l^{(t+1)}\left(\mathbf{W}_l\mathbf{x}^{(t+1)} - \mathbf{z}_l^{(t+1)}\right) \tag{4c}$$

where $\mathbf{z}_l$ are auxiliary variables in wavelet domain, $\boldsymbol{\beta}_l$ are dual variables, $\text{soft}(\cdot; \lambda_l/\rho_l)$ is the ℓ_1 soft-thresholding operator parameterized by λ_l/ρ_l, and t denotes the iteration count. The algorithm is unrolled for T iterations, as depicted in Figure 1.

The learnable parameters in this algorithm are ρ_l, λ_l/ρ_l and η_l, which correspond to parameters for augmented Lagrangian relaxation, ℓ_1 soft-thresholding and the dual update per each wavelet transform. We note that these are shared across all the unrolled iterations to ensure that the objective function in (2) remains unchanged throughout the iterations, and interpretability of the algorithm can be maintained. Thus, there are $3 \cdot L$ learnable parameters for the whole algorithm when using L orthogonal DWTs.

This approach serves as the foundation for all our proposed models, and is subsequently referred to as the learned naive ℓ_1-Wavelet reconstruction. In all our models, the input to the network is the zerofilled image, $\mathbf{x}^{(0)} = \mathbf{E}^H\mathbf{y}$. Furthermore, since the regularizer in (2) scales with $||\mathbf{x}||_\infty$, while the DC term in (2) scales with $||\mathbf{x}||_\infty^2$, λ_l/ρ_l is parametrized as $\gamma_l \left\|\mathbf{W}_l\mathbf{x}^{(0)}\right\|_\infty$, and the scaling-invariant parameter γ_l is learned. Overall, the learned parameters are $\{\rho_l, \gamma_l, \eta_l\}_{l=1}^L$.

C. Further Enhancements for Optimized ℓ_1-Wavelet CS Reconstruction

The naive optimized ℓ_1-Wavelet approach can further be enhanced using our understanding of wavelet representations [17] and of ℓ_1 minimization problems [18]. In particular, we use the fact that signal scaling changes severely between different wavelet subbands for the former, and that reweighted ℓ_1 minimization helps recover finer details for the latter.

Learning ℓ_1-wavelet reconstruction with subband processing: For different subbands of a wavelet transform, the soft-thresholding parameters may be different. To this end, let $\mathbf{D}_l^s$ be an operator that select the s^{th} subband of the l^{th} wavelet transform. We propose to use the regularizer

$$\mathcal{R}(\mathbf{x}) = \sum_{l=1}^{L}\sum_{s=1}^{S} \lambda_{l,s} \left\|\mathbf{D}_l^s\mathbf{W}_l\mathbf{x}\right\|_1, \tag{5}$$

which also lead to the following modified update in (4b).

$$\mathbf{D}_l^s\mathbf{z}_l^{(t+1)} = \text{soft}\Big(\mathbf{D}_l^s\big(\mathbf{W}_l\mathbf{x}^{(t+1)} + \boldsymbol{\beta}_l^{(t)}\big); \frac{\lambda_{l,s}}{\rho_l}\Big) \tag{6}$$

for all $s \in \{1, \ldots, S\}$. Thus, the learnable soft-thresholding parameters are $\lambda_{l,s}/\rho_l$ for the s^{th} subband of the l^{th} wavelet transform. During end-to-end training, this parameter is again implemented in a scaling-invariant manner by defining

Fig. 2: A representative slice from coronal PD knee MRI, reconstructed using PG-DL, learned naive ℓ_1-wavelet, learned ℓ_1-wavelet with subbands, and learned reweighted ℓ_1-wavelet with subbands. The proposed optimized ℓ_1-wavelet reconstructions perform closely to PG-DL. Learned reweighted ℓ_1-wavelet with subbands performs the best among these ℓ_1-wavelet CS variants, resulting in sharp images with similar quantitative metrics to PG-DL.

$\gamma_{l,s} = (\lambda_{l,s}/\rho_l)/\left\|\mathbf{D}_l^s \mathbf{W}_l \mathbf{x}^{(0)}\right\|_\infty$ and learning $\{\gamma_{l,s}\}$ for $l \in \{1, \ldots, L\}$ and $s \in \{1, \ldots, S\}$. Thus this approach leads to a total of $L \cdot (S+2)$ learnable parameters when using S subbands and L orthogonal DWTs.

Learning reweighted ℓ_1-wavelet reconstruction with subband processing: A further improvement in performance, especially in the lower SNR regimes, may be achieved using reweighted ℓ_1 minimization, which has been shown to improve recovery of small coefficients [18]. To this end, let $\hat{\mathbf{x}}_{\text{sb}}$ denote the output of the learned subband ℓ_1-wavelet reconstruction. We define a diagonal weight matrix $\mathbf{U}_l$ whose $(k,k)^{\text{th}}$ entry is given as

$$(\mathbf{U}_l)_{(k,k)} = \frac{1}{|(\mathbf{W}_l \hat{\mathbf{x}}_{\text{sb}})_k| + \epsilon}, \tag{7}$$

where $(\cdot)_k$ denotes the k^{th} coefficient of the vector $(\cdot)$, and ϵ is a small constant to avoid numerical issues when dividing by zero. This weight matrix is used to define the reweighted ℓ_1 regularizer with subband processing as:

$$\mathcal{R}(\mathbf{x}) = \sum_{l=1}^{L} \sum_{s=1}^{S} \lambda_{l,s} \left\|\mathbf{D}_l^s \mathbf{U}_l \mathbf{W}_l \mathbf{x}\right\|_1 . \tag{8}$$

This leads to the following modified update in (4b).

$$\mathbf{D}_l^s \mathbf{z}_l^{(t+1)} = \text{soft}\Big(\mathbf{D}_l^s \big(\mathbf{W}_l \mathbf{x}^{(t+1)} + \boldsymbol{\beta}_l^{(t)}\big); \frac{\lambda_{l,s}}{\rho_l} \mathbf{D}_l^s \text{diag}(\mathbf{U}_l)\Big). \tag{9}$$

for all $s \in \{1, \ldots, S\}$. We note that the regularizer in (8) does not change if $\mathbf{x}$ is scaled by a constant α, while the DC term in (2) still scales with $||\mathbf{x}||_\infty^2$. Thus, we define a scaling-invariant thresholding factor $\gamma_{l,s}^r = (\lambda_{l,s}/\rho_l)/\left\|\mathbf{D}_l^s \mathbf{W}_l \mathbf{x}^{(0)}\right\|_\infty^2$. During end-to-end training, $\{\gamma_{l,s}^r\}$ is learned for $l \in \{1, \ldots, L\}$ and $s \in \{1, \ldots, S\}$ in addition to $\{\rho_l, \eta_l\}_{l=1}^L$.

This approach still has $L \cdot (S+2)$ learnable parameters during the reweighting stage, even though signal-dependent weights are incorporated via (7). Including the learned subband reconstruction, which is used to determine the weights in (7) leads to a total of $2L \cdot (S+2)$ learnable parameters for the whole reconstruction pipeline. We also note that once these parameters are learned, they can be applied for multiple reweightings, k_{rew}, during testing, since the scaling of $\{\gamma_{l,s}^r\}$ remains on the same order.

D. Imaging Data

Fully-sampled coronal proton density (PD), and PD with fat-suppression (PD-FS) knee data obtained from the NYU-fastMRI database [23] were used throughout the experiments. Relevant imaging parameters were: matrix size = 320×368, in-plane resolution = 0.49×0.44 mm^2, slice thickness = 3 mm. The datasets were retrospectively under-sampled with a random mask ($R = 4$ with 24 ACS lines). Training was performed on 300 slices from 10 different subjects. Testing was performed on all slices from 10 different subjects. Coil sensitivity maps were generated using ESPIRiT [24].

E. Implementation Details

For all models, $L = 4$ wavelets were used, corresponding to Daubechies1-4 orthogonal wavelets with 14 subbands for each. Thus, the total number of learnable parameters were 12, 64 and 128 for the learned naive ℓ_1-wavelet, ℓ_1-wavelet with subbands, and reweighted ℓ_1-wavelet with subbands reconstructions, respectively. For the last method, $k_{rew} = 2$ was used in testing, similar to [18].

ADMM algorithm was unrolled for $T = 10$ for all models. ϵ was set to 10^{-9} in (7). DC subproblem was solved using conjugate gradient [7] with 5 iterations and warm-start. All tunable parameters were randomly initialized. Adam optimizer with learning rate 5×10^{-3} was used for training over 100 epochs, with a batch size of 1. Supervised training was performed with a normalized $\ell_1 - \ell_2$ loss in k-space [8], [10], using TensorFlow in Python.

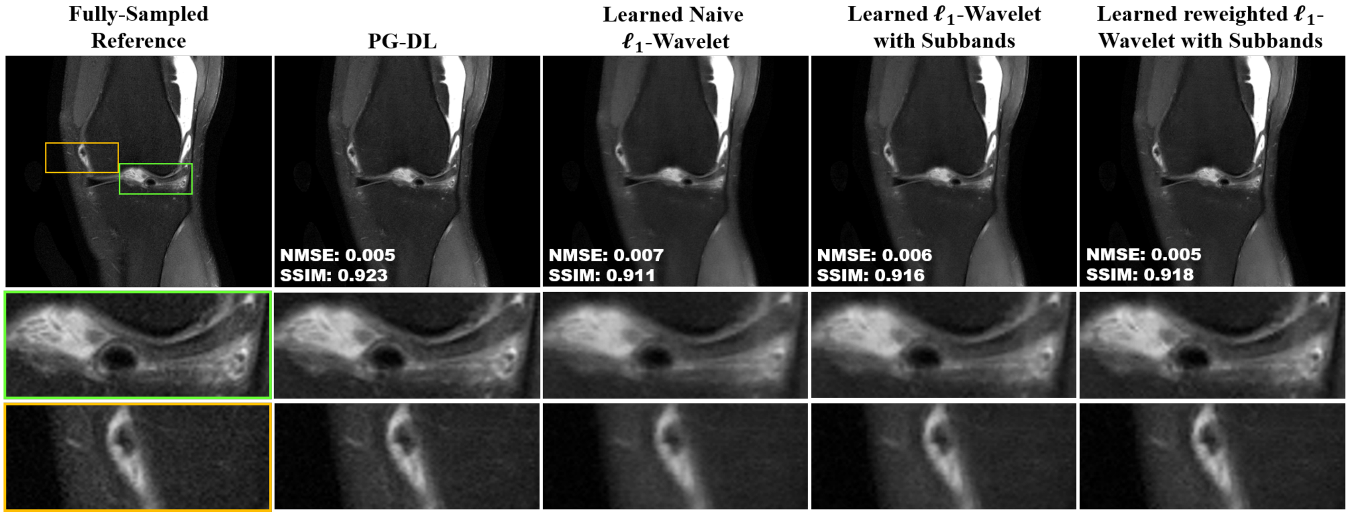

Fig. 3: A representative slice from coronal PD–FS knee MRI, reconstructed using PG-DL, learned naive ℓ_1-wavelet, learned ℓ_1-wavelet with subbands, and learned reweighted ℓ_1-wavelet with subbands. For this lower SNR acquisition, the learned naive ℓ_1-wavelet suffers from visible blurring artifacts. These are improved using subband processing, while the sharpness is only fully recovered using the learned reweighted ℓ_1-wavelet with subbands method, which leads to a visibly similar reconstruction to PG-DL.

For comparison, a PG-DL approach was implemented using the same ADMM unrolling except for using a ResNet-based regularizer unit. The ResNet was originally adapted from the winner of a super-resolution challenge [25], and has been used in multiple recent MRI studies successfully [9], [10]. The PG-DL approach has a total of 592,130 learnable parameters. Note this constitutes a head-to-head comparison, with the only difference being in the $\mathcal{R}(\mathbf{x})$ term, where our approaches employ ℓ_1-norm of wavelets for solving a convex problem, while PG-DL uses a CNN for implicit regularization. All results were quantitatively compared using SSIM and NMSE.

III. Results

Figure 2 shows a representative slice from coronal PD knee MRI, reconstructed using PG-DL and the three ℓ_l-wavelet CS approaches optimized with modern data science tools, respectively. For this high SNR acquisition, even the learned naive ℓ_1-wavelet results in a high-quality reconstruction, while the learned ℓ_1-wavelet with subbands leads to a slightly sharper reconstruction. Learned reweighted ℓ_1-wavelet with subbands performs the best among the three optimized ℓ_l-wavelet CS approaches, resulting in a sharp reconstruction, showing comparable visual quality and quantitative metrics to PG-DL.

Figure 3 depicts a representative coronal PD-FS knee MRI slice, reconstructed using PG-DL and the three optimized ℓ_l-wavelet CS approaches. The PD-FS dataset has an inherently lower SNR compared to PD. As such, the learned naive ℓ_1-wavelet results in a blurry image due to the difficulty of fine-tuning a single thresholding parameter. This is improved with subband processing, but sharpness is fully recovered only with the learned reweighted ℓ_1-wavelet with subbands approach, which results in a visibly similar reconstruction to the PG-DL method.

Table I summarizes quantitative results from knee MRI. While PG-DL has the best metrics, the gap between proposed optimized ℓ_l-wavelet CS and PG-DL is small, with < 0.01 for SSIM and < 0.0008 for NMSE.

IV. Discussion and Conclusion

In this study, we revisited ℓ_l-wavelet CS for accelerated MRI using modern data science tools for fine tuning. As expected, PG-DL outperformed our three optimized ℓ_l-wavelet CS approaches, but the performance gap was smaller than previously published literature. This is interesting for a number of reasons. First, PG-DL used a sophisticated non-linear representation for the underlying images during regularization with a large number of learnable parameters. On the other hand, the wavelet-based representations we

	Coronal PD		Coronal PD-FS	
	NMSE	SSIM	NMSE	SSIM
PG-DL	.0013 [.0010, .0017]	0.9673 [0.9547, 0.9775]	.0074 [.0049, 0.0107]	0.8795 [0.8322, 0.9146]
Learned Naive ℓ_1-wavelet	.0019 [.0014, .0024]	0.9574 [0.9404, 0.9692]	.0090 [.0062, 0.0125]	0.8598 [0.8121, 0.9023]
Learned ℓ_1-wavelet with Subbands	.0019 [.0014, .0024]	0.9588 [0.9423, 0.9699]	.0081 [.0055, .0121]	0.8662 [0.8202, 0.9057]
Learned reweighted ℓ_1-wavelet with Subbands	.0017 [.0013, .0022]	0.9602 [0.9444, 0.9709]	.0082 [.0053, .0116]	0.8694 [0.8211, 0.9068]

TABLE I: The median and the interquartile range [25^{th}, 75^{th} percentile] of the NMSE and SSIM metrics on test slices from 10 subjects for coronal PD and PD-FS datasets. While PG-DL has the best metrics as expected, the gap in SSIM between PG-DL and the learned reweighted ℓ_1-wavelet with subbands is < 0.01.

used were linear, involved only a small number of parameters, and allowed for convex optimization. Interestingly, there was <0.01 difference in SSIM between our proposed learned reweighted ℓ_1-wavelet with subbands that used 128 parameters and the PG-DL approach that used $>500{,}000$ parameters. Second, while PG-DL can be further improved with more advanced neural networks and training strategies [26], our CS approach used one of the simplest linear models described by fixed orthogonal wavelets, and did not involve learning of the representation. Our results also showed that the performance gap decreased as we proceeded from learned naive ℓ_1-wavelet to learned reweighted ℓ_1-wavelet with subbands, demonstrating subband training and reweighting help improve reconstruction sharpness and quantitative metrics to a level comparable to PG-DL. Further gains for CS may be possible via learning linear representations/frames [27], [28], which warrants investigation.

V. ACKNOWLEDGEMENTS

This work was partially supported by NIH R01HL153146, NIH P41EB027061, NIH U01EB025144; NSF CAREER CCF-1651825.

REFERENCES

[1] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn Reson Med*, vol. 42, pp. 952–962, 1999.

[2] M. A. Griswold, P. M. Jakob, et al., "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magn Reson Med*, vol. 47, pp. 1202–1210, 2002.

[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] S. Wang, Z. Su, et al., "Accelerating magnetic resonance imaging via deep learning," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*. IEEE, 2016, pp. 514–517.

[5] 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, pp. 491–503, 2018.

[6] 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.

[7] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-Based Deep Learning Architecture for Inverse Problems," *IEEE Trans Med Imaging*, vol. 38, pp. 394–405, 2019.

[8] F. Knoll, K. Hammernik, et al., "Deep-learning methods for parallel magnetic resonance imaging reconstruction," *IEEE Sig Proc Mag*, vol. 37, no. 1, pp. 128–140, 2020.

[9] S. A. H. Hosseini, B. Yaman, S. Moeller, M. Hong, and M. Akçakaya, "Dense recurrent neural networks for accelerated MRI: History-cognizant unrolling of optimization algorithms," *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 6, pp. 1280–1291, 2020.

[10] B. Yaman, S. A. H. Hosseini, et al., "Self-supervised learning of physics-guided reconstruction neural networks without fully-sampled reference data," *Magn Reson Med*, vol. 84, pp. 3172–3191, Dec 2020.

[11] A. Sriram, J. Zbontar, et al., "End-to-end variational networks for accelerated mri reconstruction," 2020.

[12] F. Zou, L. Shen, Z. Jie, W. Zhang, and W. Liu, "A sufficient condition for convergences of adam and rmsprop," in *Proc. IEEE/CVF Conf. Comp. Vis. and Patt. Rec. (CVPR)*, June 2019.

[13] M. Mardani, E. Gong, et al., "Deep Generative Adversarial Neural Networks for Compressive Sensing MRI," *IEEE Trans Med Imaging*, vol. 38, pp. 167–179, 2019.

[14] M. Seitzer, G. Yang, et al., "Adversarial and perceptual refinement for compressed sensing MRI reconstruction," in *Proc. MICCAI*, 2018, pp. 232–240.

[15] M. Shahdloo, E. Ilicak, et al., "Projection onto epigraph sets for rapid self-tuning compressed sensing mri," *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1677–1689, 2019.

[16] S. Ramani, Z. Liu, J. Rosen, J. Nielsen, and J. A. Fessler, "Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods," *IEEE Trans Image Proc*, vol. 21, no. 8, pp. 3659–3672, 2012.

[17] M. Vetterli and J. Kovačevic, *Wavelets and Subband Coding*, Prentice-Hall, Inc., USA, 1995.

[18] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 11 2007.

[19] K. P. Pruessmann, M. Weiger, P. Bornert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magn Reson Med*, vol. 46, pp. 638–651, 2001.

[20] J. A. Fessler, "Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms," *IEEE Sig Proc Mag*, vol. 37, no. 1, pp. 33–40, 2020.

[21] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. Int. Conf. Mach. Learn*, 2010, pp. 399–406.

[22] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Sig Proc Mag*, vol. 38, no. 2, pp. 18–44, 2021.

[23] F. Knoll, J. Zbontar, et al., "fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning," *Radiol AI*, p. e190007, 2020.

[24] 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, Mar 2014.

[25] R. Timofte, E. Agustsson, L. Van Gool, M. H. Yang, and L. Zhang, "Ntire 2017 challenge on single image super-resolution: Methods and results," in *Proc IEEE CVPR*, 2017.

[26] M. J. Muckley, B. Riemenschneider, et al., "State-of-the-art machine learning MRI reconstruction in 2020: Results of the second fastMRI challenge," 2020.

[27] B. Wen, S. Ravishankar, L. Pfister, and Y. Bresler, "Transform learning for magnetic resonance image reconstruction," *IEEE Sig Proc Mag*, vol. 37, no. 1, pp. 41–53, 2020.

[28] M. Akcakaya and V. Tarokh, "A frame construction and a universal distortion bound for sparse representations," *IEEE Trans Sig Proc*, vol. 56, no. 6, pp. 2443–2450, 2008.