BRIEF REPORT

# Revisiting $\ell_1$-wavelet compressed sensing MRI in the era of deep learning

**Hongyi Gu**[a,b], **Burhaneddin Yaman**[a,b], **Steen Moeller**[b], **Jutta Ellermann**[b], **Kamil Ugurbil**[b], **and Mehmet Akçakaya**[a,b,1]

[a]Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, 55455; [b]Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, MN, 55455

This manuscript was compiled on June 22, 2022

**Following their success in numerous imaging and computer vision applications, deep learning (DL) techniques have emerged as one of the most prominent strategies for accelerated magnetic resonance imaging (MRI) reconstruction. These methods have been shown to outperform conventional regularized methods based on compressed sensing (CS). However, in most comparisons, CS is implemented with two or three hand-tuned parameters, while DL methods enjoy a plethora of advanced data science tools. In this work, we revisit $\ell_1$-wavelet CS reconstruction using these modern tools. Using ideas such as algorithm unrolling and advanced optimization methods over large databases that DL algorithms utilize, along with conventional insights from wavelet representations and CS theory, we show that $\ell_1$-wavelet CS can be fine-tuned to a level close to DL reconstruction for accelerated MRI. The optimized $\ell_1$-wavelet CS method uses only 128 parameters compared to $> 500,000$ for DL, employs a convex reconstruction at inference time, and performs within $< 1\%$ of a DL approach that has been used in multiple studies in terms of quantitative quality metrics.**

deep learning | AI | compressed sensing | MRI reconstruction | inverse problems

Lengthy data acquisition remains an impediment for MRI, requiring the use of accelerated imaging techniques. Recently, deep learning (DL) methods have emerged as a powerful strategy for accelerated MRI (1–3), with many studies showing substantial improvement over conventional methods, such as compressed sensing (CS) (4). Among DL methods, physics-guided (PG-DL) approaches that incorporate the forward MRI encoding operator have received increased attention (1, 2). These methods use a non-linear representation for regularization, implicitly learned through neural networks, as opposed to the linear transform based representations of images in CS. DL reconstruction methods are trained on large databases, include hundreds of thousands tunable parameters, use advanced optimization techniques and loss functions (3). When CS methods are implemented for comparison, they typically use two or three hand-tuned parameters, and do not leverage the sophisticated tools from the DL era.

In this study, we use these advanced data science tools to revisit $\ell_1$-wavelet CS for accelerated MRI. To this end, we leverage ideas that are often used for DL reconstructions, such as algorithm unrolling and end-to-end training over large databases, as well as conventional insights from CS methodology, such as wavelet sub-band processing (5) and reweighted $\ell_1$ minimization (6). We show that an optimized learned $\ell_1$-wavelet CS strategy with a mere 128 tunable parameters performs close to a PG-DL method with $> 500,000$ parameters that has been used in previous studies (7), quantitatively in terms of peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and blur metric, and qualitatively in terms of expert reader scores.

## Approach

**Regularized Reconstruction for Accelerated MRI.** The inverse problem for accelerated MRI involves solving the objective function:

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

where $\mathbf{E} : \mathbb{C}^N \rightarrow \mathbb{C}^M$ is the forward multi-coil encoding operator used in parallel imaging that contains coil sensitivity maps and partial Fourier matrix for under-sampling in k-space (3, 8), $\mathbf{y} \in \mathbb{C}^M$ is the undersampled k-space data from all coils, and $\mathbf{x} \in \mathbb{C}^N$ is the image to be reconstructed. Note the quadratic term in Eq. 1 enforces data consistency (DC), while $\mathcal{R}(\cdot)$ is a regularizer.

In conventional CS MRI reconstruction, the regularizer is typically a weighted $\ell_1$ norm of transform coefficients, i.e. $\mathcal{R}(\mathbf{x}) = \sum_{l=1}^{L} \lambda_l \|\mathbf{W}_l\mathbf{x}\|_1$, where $\mathbf{W}_1$ is a pre-specified linear transform, such as a discrete wavelet transform (DWT) (4). This leads to a convex objective function, which is solved using iterative optimization algorithms (8). For instance the alternating direction method of multipliers (ADMM) algorithm leads to three tunable parameters, one for DC, one for dual update, and one for transform domain soft-thresholding, which are often hand-tuned in practice. Such algorithms are run until a stopping criterion is met, which further makes parameter tuning difficult.

On the other hand, in PG-DL methods, the problem in Eq. 1 is solved using the idea of algorithm unrolling, which unrolls an iterative algorithm for this problem, such as ADMM, for a fixed number of iterations (3). In this case, the DC units are implemented with conventional methods, such as gradient descent or conjugate gradient (7), while the proximal operation related to the regularizer unit is solved implicitly using neural networks. This unrolled network is then trained end-to-end over a large database using a loss function:

$$\min_{\boldsymbol{\theta}} \frac{1}{N}\sum_{n=1}^{N} \mathcal{L}\Big(\mathbf{y}_{\text{ref}}^n,\ \mathbf{E}_{\text{full}}^n\big(f(\mathbf{y}^n, \mathbf{E}^n; \boldsymbol{\theta})\big)\Big), \qquad [2]$$

where $\mathbf{y}_{\text{ref}}^n$ denotes the fully-sampled reference k-space of the $n^{\text{th}}$ subject, $\mathbf{E}_{\text{full}}^n$ is the fully sampled multi-coil encoding operator of the $n^{\text{th}}$ subject, $N$ is the number of datasets in the training database, $f(\mathbf{y}^n, \mathbf{E}^n; \boldsymbol{\theta})$ denotes the network output of the unrolled network with parameters $\boldsymbol{\theta}$ of the $n^{\text{th}}$ subject, and $\mathcal{L}(\cdot,\cdot)$ is a loss function, such as the $\ell_2$ norm or $\ell_1$ norm (3).

**Proposed Learning of Optimized $\ell_1$-wavelet CS MRI Reconstruction.** For the optimized $\ell_1$-wavelet CS method, we propose to use both the aforementioned algorithm unrolling and end-to-end training strategies. To this end, we unroll the ADMM algorithm for $T = 10$ iterations. While a single DWT, with Daubechies4 wavelets

Author contributions: H. G., B. Y., and M. A. designed research; H. G. and B. Y. performed research; J. E. performed expert radiologist reading of data; H. G. and B. Y. analyzed data; H. G., B. Y., S. M., J.E., K. U. and M. A. wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. E-mail: akcakayaumn.edu

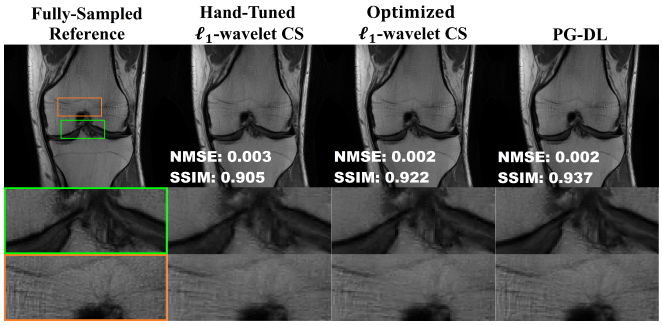


**Fig. 1.** A representative slice from coronal PD knee MRI, reconstructed using hand-tuned $\ell_l$-wavelet CS, optimized $\ell_1$-wavelet CS, and PG-DL. The proposed optimized $\ell_1$-wavelet CS outperforms hand-tuned $\ell_l$-wavelet CS, and it has comparable performance to PG-DL.

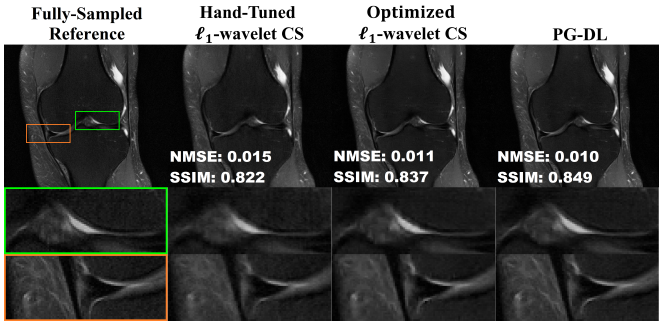


**Fig. 2.** A representative coronal PD-FS knee slice, reconstructed using hand-tuned $\ell_l$-wavelet CS, optimized $\ell_1$-wavelet CS, and PG-DL. The proposed optimized $\ell_1$-wavelet CS performs closely to PG-DL, and has better reconstruction quality than hand-tuned $\ell_1$-wavelet CS method.

being popular, is commonly used in conventional CS MRI reconstruction (4), a level of redundancy will further benefit the reconstruction. In order to show the versatility of the original CS formulation, we use $L = 4$ DWTs, Daubechies1-4 wavelets, to form a simple overcomplete representation (9). Noting that for a given DWT, ADMM has three tunable parameters, for data consistency, dual update and transform domain soft-thresholding (8), this leads to a total of $3 \cdot L = 12$ = 12 learnable parameters, all of which can be learned using the end-to-end training formulation described in Eq. 2.

Building on this basic model, we augment the $\ell_1$-wavelet CS method with two strategies derived using the characteristics of DWTs and $\ell_1$ minimization. First, due to the substantial signal scaling changes between different sub-bands of a DWT, we propose to use a different soft-thresholding parameter for each wavelet sub-band. For $S$ sub-bands, this leads to $L \cdot (S+2) = 4(S+2)$ learnable parameters. In our experiments, $S = 14$ leads to 64 learnable parameters. Finally, especially in lower signal-to-noise ratio regimes, it has been shown that reweighted $\ell_1$ minimization (6) may help enhance the recovery of small coefficients, which in turn may approve blurring artifacts associated with CS reconstruction. In our setup, the output of the sub-band processed optimized $\ell_1$-wavelet CS reconstruction is used to define the new weights for reweighted $\ell_1$ regularizer. Once these signal-dependent weights are incorporated into the objective function, the sub-band thresholding weights are relearned. Note there are still $L \cdot (S+2)$ learnable parameters for this stage, and the total number of learnable parameters across the two stages are $2 \cdot L \cdot (S+2) = 128$. This proposed approach is also trained using the formulation of Eq. 2, and will be referred to as the optimized $\ell_1$-wavelet CS reconstruction.

## Results

Experiments were carried out using fully-sampled coronal proton density (PD) and PD with fat-suppression (PD-FS) knee data obtained from the NYU-fastMRI database (10). The datasets were retrospectively undersampled with a random mask ($R$ = 4 with 24 central k-space autocalibration signal lines). Training and testing were performed on PD and PD-FS data separately. For comparison, a PG-DL approach was implemented using the same ADMM unrolling except the use of a ResNet-based regularizer unit, with a total of 592,130 learnable parameters. This ResNet was originally adapted from the winner of a super-resolution challenge, and has been used in recent MRI studies successfully (7). Note the only difference between this PG-DL and the proposed $\ell_1$-wavelet CS methods is the $\mathcal{R}(\cdot)$ term, where the former employs a neural network for implicit regularization, while the latter uses weighted $\ell_1$ norm of wavelets for solving a convex problem. Finally, for baseline comparison, a conventional implementation of $\ell_1$-wavelet CS reconstruction was implemented using D4 wavelets for regularization (4), which was solved using ADMM (8). Here, the parameters of ADMM were hand-tuned empirically, and this method is referred to as the hand-tuned $\ell_1$-wavelet CS reconstruction.

Figures 1 and 2 show representative slices from coronal PD and PD-FS knee MRI, respectively. In both cases, the proposed optimized $\ell_1$-wavelet CS has visibly comparable image quality to PG-DL, while both methods yield sharper images compared to the conventional hand-tuned $\ell_1$-wavelet CS.

Table 1 summarizes the assessments over all test datasets, including quantitative measures of SSIM and PSNR with respect to the reference image, and a referenceless blur metric (11), as well as qualitative image reading scores for SNR and aliasing artifacts, which were evaluated on a 4-point ordinal scale (1: excellent, 2: good, 3: fair, 4: poor) (2). Both SSIM and PSNR show that the proposed optimized $\ell_l$-wavelet CS outperforms the conventional hand-tuned $\ell_l$-wavelet CS method, with a performance close to PG-DL. The referenceless quantitative blur metrics show the same trend, with the optimized $\ell_l$-wavelet CS comfortably outperforming hand-tuned $\ell_l$-wavelet CS, while having close metrics to PG-DL. These results suggest that the difference in performance on individual patient/scan reconstructions between

**Table 1. Summary of results over coronal PD and PD-FS test datasets. We used a total of 786 slices of coronal PD and PD-FS from 10 subjects for testing. SSIM, NMSE and blur metrics were calculated individually for each of these slices. The first and second row show the median and the interquartile range [25$^{th}$, 75$^{th}$ percentile] of the PSNR and SSIM metrics for all methods. The third row shows the median and the interquartile range [25$^{th}$, 75$^{th}$ percentile] of the blur metric of the reference and all methods. Qualitative image readings were also performed by an expert radiologist, where one score was given for each of the PD and PD-FS datasets per subject. The fourth and the fifth row show the mean and the +/- standard deviation of image readings for SNR and Aliasing Artifacts respectively, for the reference and all methods.**

| | Reference | Hand-Tuned $\ell_1$-wavelet CS | Optimized $\ell_1$-wavelet CS | PG-DL |
|---|---|---|---|---|
| PSNR | - | 36.5230 [34.0447, 38.7672] | 37.9107 [35.1328, 40.2464] | 38.7396 [35.5185, 41.2859] |
| SSIM | - | 0.9190 [0.8510, 0.9499] | 0.9317 [0.8618, 0.9619] | 0.9423 [0.8737, 0.9683] |
| Blur Metric | 0.2777 [0.2183, 0.3207] | 0.3329 [0.2683, 0.3876] | 0.3278 [0.2804, 0.3742] | 0.3202 [0.2652, 0.3664] |
| Perceived SNR Reading | 2.50 +/- 0.5130 | 2.50 +/- 0.5130 | 2.25 +/- 0.4435 | 2.20 +/- 0.7194 |
| Aliasing Artifact Reading | 1.75 +/- 0.7018 | 2.80 +/- 0.5187 | 2.60 +/- 0.4756 | 2.25 +/- 0.4051 |

PG-DL and optimized $\ell_l$-wavelet CS approaches is typically smaller than the inter-patient/scan variability of either approach. In terms of qualitative image readings, all methods performed similarly in terms of perceived SNR. Interestingly, the reference had the worst score due to the higher level of noise from the acquisitions. For aliasing artifacts readings, the trend was the same with PG-DL outperforming both $\ell_l$-wavelet CS methods, though the difference to the proposed optimized $\ell_l$-wavelet CS method was not statistically significant.

## Discussion

In this study, we revisited $\ell_l$-wavelet CS for accelerated MRI using powerful data science tools that have been developed in the DL era for fine tuning. While the highly tunable PG-DL outperformed our optimized $\ell_l$-wavelet CS approach as expected, the performance gap was smaller than previously reported in the literature. This is interesting for a number of reasons. First, during regularization, PG-DL implicitly uses a sophisticated non-linear representation for the underlying images with a large number of learnable parameters. On the other hand, our $\ell_1$-wavelet approach uses linear representations, involve only a small number of parameters, and enable an explainable convex optimization procedure at inference time. Interestingly, there is $<$0.01 difference in SSIM between our proposed learned $\ell_1$-wavelet method that uses 128 parameters and the PG-DL approach that uses $>$500,000 parameters. Second, while PG-DL may potentially be further improved using more sophisticated neural networks and training strategies (12), it is worth noting that our $\ell_1$-wavelet CS approach used a very simple linear model described by four fixed orthogonal DWTs, and did not involve any learning of such linear representations.

Though the difference between optimized $\ell_1$-wavelet CS and PG-DL was smaller than previously reported quantitatively and not significant qualitatively, PG-DL did have the highest metrics and was the preferred reconstruction method of the expert reader for all subjects, attesting to its ability to retain subtle details beyond the 4-point ordinal scale. Although this is not entirely surprising given the complexity of the PG-DL model, both the improved quantitative metrics and the statistically similar image readings are encouraging for traditional CS methods. Going beyond the simple 128-parameter used here with pre-defined Daubechies wavelet transforms, use of more advanced wavelets, such as symmlets, or CS with linear transform/tight frame learning (13, 14) may further close this gap. Given the promising results from the current model, such improvements warrant further investigation, since the $\ell_1$-norm based CS reconstruction uses convex sparse image reconstruction with linear representation at inference time, which may be beneficial for characterizing robustness, generalizability and stability (15).

## Materials and Methods

**Training and Testing Details.** For all methods, ADMM was unrolled for $T = 10$ iterations. In all cases, the input to the network was the zerofilled image, $\mathbf{x}^{(0)} = \mathbf{E}^H\mathbf{y}$. The coil sensitivites in $\mathbf{E}$ were estimated using ESPIRiT (16). DC subproblem was solved using conjugate gradient (3, 7) with 5 iterations and warm-start. All tunable parameters were shared across iterations, consistent with the ADMM solution for Eq. 1 The parameters were randomly initialized during training. Adam optimizer with learning rate $5 \cdot 10^{-3}$ 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 (3, 7), using TensorFlow in Python. The reference images for supervised training were generated using optimal coil combination, where the fully-sampled coil images were elementwise multiplied by the complex conjugate of the coil sensitivities, and summed across the coil dimension.Training was performed on 300 slices from 10 subjects for coronal PD and PD-FS datasets. Testing was performed on all 786 coronal PD and PD-FS slices from 10 different subjects. For each method, SSIM, NMSE and blur metrics were calculated individually for each of these slices.

**Implementation Details for $\ell_1$-wavelet CS Method.** For the learning of $\ell_1$-wavelet CS reconstruction, note that the regularizer in Eq. 1 scales with $||\mathbf{x}||_\infty$, while the DC term scales with $||\mathbf{x}||_\infty^2$. Thus, the soft-thresholding parameter for the $s^{\text{th}}$ sub-band of the $l^{\text{th}}$ DWT was represented as $\gamma_{l,s}||\mathbf{D}_l^s\mathbf{W}_l\mathbf{x}^{(0)}||_\infty$, where $\mathbf{D}_l^s$ is an operator that selects the $s^{\text{th}}$ sub-band of the $l^{\text{th}}$ DWT $\mathbf{W}_l$. During training, these scaling-invariant parameters $\{\gamma_{l,s}\}_{l,s=1}^{L,S}$ were learned.

Similarly, for the reweighted $\ell_1$ case, the regularization term is unaffected if $\mathbf{x}$ is scaled by a constant $\alpha$, while the DC term in Eq. 1 scales with $\alpha^2$. Thus, a scaling-invariant thresholding factor was implemented as $\gamma_{l,s}^r||\mathbf{D}_l^s\mathbf{W}_l\mathbf{x}^{(0)}||_\infty^2$. During end-to-end training, $\{\gamma_{l,s}^r\}$ were learned for $l \in \{1,\dots,L\}$ and $s \in \{1,\dots,S\}$. Finally while reweighting, a small constant is used to avoid numerical issues when dividing by zero (6). This was set to $10^{-9}$ in our experiments.

**Image Readings and Statistical Analysis.** Qualitative assessment of the image quality from the three different reconstruction methods (PG-DL, hand-tuned $\ell_1$-wavelet CS and proposed optimized $\ell_1$-wavelet CS) was performed by an experienced radiologist. The reader was blinded to the reconstruction methods, which were shown in a randomized order to avoid bias, except for the knowledge of the reference image. There were differences between the sequences used for the fastMRI database and our institutional sequences, thus this knowledge allowed the radiologist to assess the baseline image quality. Evaluations were based on a 4-point ordinal scale, adopted from (2) for perceived SNR (1: excellent, 2: good, 3: fair, 4: poor) and aliasing artifacts (1: none, 2: mild, 3: moderate, 4: severe), where one score was used per subject. Wilcoxon signed-rank test was used to evaluate the scores with a significance level of $P < 0.05$. Additionally, instead of scoring for overall image quality as in (2), the radiologist was asked to identify their preferred reconstruction method for each subject. This was done to capture subtle differences and preferences that would not be captured as differences in the 4-point ordinal scale.

**Reproducibility and Data Availability.** All source codes for training and testing, and the weights of the pretrained networks are available through https://imagine.umn.edu/research/software. The raw MRI data used for this study is available in the fastMRI database, https://fastmri.org/ (10).

**ACKNOWLEDGMENTS.** This work was partially supported by NIH R01HL153146, P41EB027061, U01EB025144; NSF CAREER CCF-1651825. Preliminary results of this work were presented at the Annual Meeting of International Society of Magnetic Resonance in Medicine (ISMRM) (17).

### References.

1. J Schlemper, et al., A deep cascade of convolutional neural networks for dynamic MRI reconstruction. *IEEE Trans Med Imag* **37**, 491–503 (2018).
2. K Hammernik, et al., Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson. Med* **79**, 3055–3071 (2018).
3. F Knoll, et al., Deep-learning methods for parallel magnetic resonance imaging reconstruction. *IEEE Sig Proc Mag* **37**, 128–140 (2020).
4. M Lustig, D Donoho, J Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson. Med* **58**, 1182–1195 (2007).
5. M Vetterli, J Kovačevic, *Wavelets and Subband Coding.* (Prentice-Hall, Inc., USA), (1995).
6. E Candès, M Wakin, S Boyd, Enhancing sparsity by reweighted l1 minimization. *J. Fourier Analysis Appl.* **14**, 877–905 (2007).
7. B Yaman, et al., Self-supervised learning of physics-guided reconstruction neural networks without fully-sampled reference data. *Magn Reson. Med* **84**, 3172–3191 (2020).
8. JA Fessler, Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms. *IEEE Signal Process. Mag.* **37**, 33–40 (2020).
9. RR Coifman, DL Donoho, *Translation-Invariant De-Noising*, eds. A Antoniadis, G Oppenheim. (Springer New York, New York, NY), pp. 125–150 (1995).
10. F Knoll, et al., fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MRI reconstruction using machine learning. *Radiol. AI* **2**, e190007 (2020).
11. P Marziliano, F Dufaux, S Winkler, T Ebrahimi, A no-reference perceptual blur metric in *Proceedings. International Conference on Image Processing.* Vol. 3, pp. III–III (2002).
12. MJ Muckley, et al., Results of the 2020 fastMRI challenge for machine learning MR image reconstruction. *IEEE Trans Med Imaging* **in press**, doi:10.1109/TMI.2021.3075856 (2021).
13. B Wen, S Ravishankar, L Pfister, Y Bresler, Transform learning for magnetic resonance image reconstruction. *IEEE Sig Proc Mag* **37**, 41–53 (2020).
14. IY Chun, JA Fessler, Convolutional Analysis Operator Learning: Acceleration and Convergence. *IEEE Trans Image Proc* **29**, 2108–2122 (2020).
15. V Antun, et al., On instabilities of deep learning in image reconstruction and the potential costs of ai. *Proc. Natl. Acad. Sci.* **117**, 30088–30095 (2020).
16. M Uecker, et al., ESPIRiT–an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magn Reson. Med* **71**, 990–1001 (2014).
17. H Gu, et al., Compressed sensing MRI revisited: Optimizing $\ell_1$-wavelet reconstruction with modern data science tools in *Proc. ISMRM.* p. 274 (2021).