OPEN ACCESS

Revisiting ℓ_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]

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 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 ℓ_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 ℓ_1-wavelet CS can be fine-tuned to a level close to DL reconstruction for accelerated MRI. The optimized ℓ_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 | compressed sensing | MRI reconstruction | AI | 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 DL (PG-DL) approaches that incorporate the forward MRI encoding operator have received increased attention (1, 2). These methods use a nonlinear 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 of tunable parameters, and 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 ℓ_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 subband processing (5) and reweighted ℓ_1 minimization (6). We show that an optimized learned ℓ_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 $\mathbf{E} : \mathbb{C}^N \to \mathbb{C}^M$ is the forward multicoil encoding operator used in parallel imaging that contains coil sensitivity maps and partial Fourier matrix for undersampling 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 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 prespecified 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 $\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}\left(\mathbf{y}_{\text{ref}}^n, \mathbf{E}_{\text{full}}^n\left(f(\mathbf{y}^n, \mathbf{E}^n; \boldsymbol{\theta})\right)\right)$, where $\mathbf{y}_{\text{ref}}^n$ denotes the fully sampled reference k space of the nth subject; $\mathbf{E}_{\text{full}}^n$ is the fully sampled multicoil encoding operator of the nth 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 nth subject; and $\mathcal{L}(\cdot, \cdot)$ is a loss function, such as the ℓ_2 norm or ℓ_1 norm (3).

Proposed Learning of Optimized ℓ_1-Wavelet CS MRI Reconstruction. For the optimized ℓ_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 being popular, is commonly used in conventional CS MRI reconstruction (4), a level of redundancy will further benefit the reconstruction. To show the versatility of the original CS formulation, we use $L = 4$ DWTs, Daubechies1-to-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 earlier.

Building on this basic model, we augment the ℓ_1-wavelet CS method with two strategies derived using the characteristics of DWTs and ℓ_1 minimization. First, due to the substantial signal scaling changes between different subbands of a DWT, we propose to use a different soft-thresholding parameter for each wavelet subband. For S subbands, 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 ℓ_1 minimization (6) may help enhance the recovery of small coefficients, which in turn may improve blurring artifacts associated with CS reconstruction. In our setup, the output of the subband processed optimized ℓ_1-wavelet CS reconstruction is used to define the new weights for the reweighted ℓ_1 regularizer. Once these signal-dependent weights are incorporated into the objective function, the subband 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 is $2 \cdot L \cdot (S + 2) = 128$. This proposed approach is also trained using the earlier end-to-end formulation and is referred to as the optimized ℓ_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 New York University (NYU)-fastMRI database (10). The datasets were retrospectively undersampled with a random mask ($R = 4$ with 24 central k-space autocalibration signal
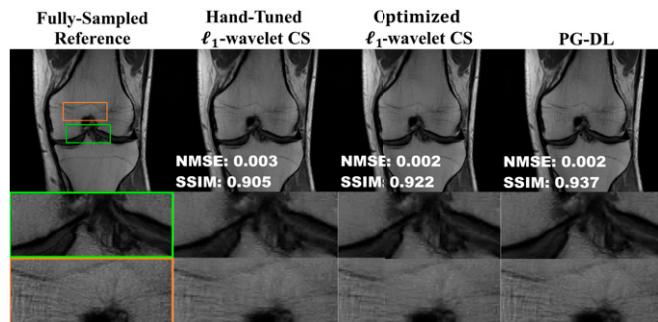

Fig. 1. A representative slice from coronal PD knee MRI, reconstructed using hand-tuned ℓ_l-wavelet CS, optimized ℓ_l-wavelet CS, and PG-DL. The proposed optimized ℓ_1-wavelet CS outperforms hand-tuned ℓ_l-wavelet CS, and it has comparable performance to that of PG-DL.

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 residual network (ResNet) regularizer unit, with a total of 592,130 learnable parameters. This ResNet was originally adapted from the winner of a superresolution challenge and has been used in recent MRI studies successfully (7). Note the only difference between this PG-DL and the proposed ℓ_1-wavelet CS methods is the $\mathcal{R}(\cdot)$ term, where the former employs a neural network for implicit regularization, while the latter uses a weighted ℓ_1 norm of wavelets for solving a convex problem. Finally, for baseline comparison, a conventional implementation of ℓ_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 ℓ_1-wavelet CS reconstruction.

Figs. 1 and 2 show representative slices from coronal PD and PD-FS knee MRI, respectively. In both cases, the proposed optimized ℓ_1-wavelet CS has visibly comparable image quality to PG-DL, while both methods yield sharper images compared to the conventional hand-tuned ℓ_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 four-point ordinal scale (1, excellent; 2, good; 3, fair; 4, poor) (2). Both SSIM and PSNR show that the proposed optimized ℓ_1-wavelet CS outperforms the conventional hand-tuned ℓ_1-wavelet CS method, with a performance close to that of PG-DL. The referenceless quantitative blur metrics show the same trend, with the optimized ℓ_l-wavelet CS comfortably outperforming hand-tuned ℓ_l-wavelet CS, while having close metrics to those of PG-DL. These results suggest that the difference in
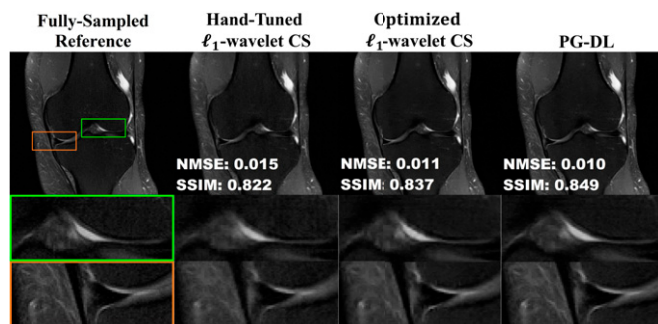

Fig. 2. A representative coronal PD-FS knee slice, reconstructed using hand-tuned ℓ_l-wavelet CS, optimized ℓ_l-wavelet CS, and PG-DL. The proposed optimized ℓ_1-wavelet CS performs closely to PG-DL and has better reconstruction quality than the hand-tuned ℓ_1-wavelet CS method.

Table 1. Summary of results over coronal PD and PD-FS test datasets

	Reference	Hand-tuned ℓ_1-wavelet CS	Optimized ℓ_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	2.500 ± 0.5130	2.500 ± 0.5130	2.250 ± 0.4435	2.200 ± 0.7194
Aliasing artifact	1.750 ± 0.7018	2.800 ± 0.5187	2.600 ± 0.4756	2.250 ± 0.4051

786 slices of coronal PD and PD-FS from 10 subjects were used for testing. SSIM, NMSE, and blur metrics were calculated individually for each of these slices. The first and second rows show the median and the interquartile range [25th, 75th percentiles] of the PSNR and SSIM metrics. The third row shows the median and the interquartile range [25th, 75th percentiles] of the blur metric. 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 row and the fifth row show the mean and ±SD of image readings for SNR and aliasing artifacts, respectively, for the reference and all methods.

performance on individual patient/scan reconstructions between PG-DL and optimized ℓ_1-wavelet CS approaches is typically smaller than the interpatient/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 ℓ_1-wavelet CS methods, although the difference from the proposed optimized ℓ_1-wavelet CS method was not statistically significant.

Discussion

In this study, we revisited ℓ_1-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 ℓ_1-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 nonlinear representation for the underlying images with a large number of learnable parameters. On the other hand, our ℓ_1-wavelet approach uses linear representations, involves only a small number of parameters, and enables an explainable convex optimization procedure at inference time. Interestingly, there is <0.01 difference in SSIM between our proposed learned ℓ_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 ℓ_1-wavelet CS approach used a very simple linear model described by four fixed orthogonal DWTs and did not involve learning of linear representations.

Although the difference between optimized ℓ_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 four-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 parameters used here with predefined 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 ℓ_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

For all methods, ADMM was unrolled for $T = 10$ iterations. In all cases, the input to the network was the zero-filled image, $\mathbf{x}^{(0)} = \mathbf{E}^H \mathbf{y}$. The coil sensitivities in $\mathbf{E}$ were estimated using ESPIRiT (16). The DC subproblem was solved using conjugate gradient (3, 7) with five iterations and warm start. All tunable parameters were shared across iterations, consistent with the ADMM solution for Eq. **1**. Supervised training was performed with a normalized ℓ_1-ℓ_2 loss in k space (3, 7), using TensorFlow in Python. 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 (17). For each method, SSIM, PSNR, and blur metrics were calculated individually for each of these slices. Further implementation details are provided in *SI Appendix*. For image readings, Wilcoxon signed-rank test was used to evaluate the scores with a significance level of $P < 0.05$. Further details on image readings are also available in *SI Appendix*.

1. J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, D. Rueckert, A deep cascade of convolutional neural networks for dynamic MRI reconstruction. *IEEE Trans. Med. Imaging* **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 Signal Process. Mag.* **37**, 128–140 (2020).
4. M. Lustig, D. Donoho, J. M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58**, 1182–1195 (2007).
5. M. Vetterli, J. Kovačević, *Wavelets and Subband Coding* (Prentice-Hall, Inc., 1995).
6. E. Candès, M. Wakin, S. Boyd, Enhancing sparsity by reweighted l1 minimization. *J. Fourier Anal. 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. J. A. Fessler, Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms. *IEEE Signal Process. Mag.* **37**, 33–40 (2020).
9. R. R. Coifman, D. L. Donoho, "Translation-Invariant De-Noising" in *Wavelets and Statistics*, A. Antoniadis, G. Oppenheim, Eds. (Springer, New York, NY, 1995). pp. 125–150.
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. Artif. Intell.* **2**, e190007 (2020).
11. P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, "A no-reference perceptual blur metric" in *Proceedings of the International Conference on Image Processing* (IEEE, 2002), vol. 3, pp. 57–60.
12. M. J. Muckley *et al.*, Results of the 2020 fastMRI challenge for machine learning MR image reconstruction. *IEEE Trans. Med. Imaging* **40**, 2306–2317 (2021).
13. B. Wen, S. Ravishankar, L. Pfister, Y. Bresler, Transform learning for magnetic resonance image reconstruction. *IEEE Signal Process. Mag.* **37**, 41–53 (2020).
14. I. Y. Chun, J. A. Fessler, Convolutional analysis operator learning: Acceleration and convergence. *IEEE Trans. Image Process.* **29**, 2108–2122 (2020).
15. V. Antun, F. Renna, C. Poon, B. Adcock, A. C. Hansen, On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc. Natl. Acad. Sci. U.S.A.* **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 ℓ_1-wavelet reconstruction with modern data science tools" in *Proceedings of the ISMRM* (ISMRM, 2021), p. 274.
18. H. Gu, hongygu/Optimized_l1_wavelet_CS. Zenodo. https://zenodo.org/record/6808387. Deposited 7 July 2022.
19. H. Gu *et al.*, Software for "Revisiting l1-wavelet compressed sensing MRI in the era of deep learning." University of Minnesota, Imagine Lab. https://imagine.umn.edu/research/software. Accessed 4 August 2022.