Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data

Burhaneddin Yaman1,2 | Seyed Amir Hossein Hosseini1,2 | Steen Moeller2 | Jutta Ellermann2 | Kâmil Uğurbil | Mehmet Akçakaya1,2

1Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, Minnesota, USA
2Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, Minnesota, USA

Correspondence
Mehmet Akçakaya, Department of Electrical and Computer Engineering, University of Minnesota, 200 Union Street S.E., Minneapolis, MN 55455, USA. Email: akcakaya@umn.edu

Funding information
National Institutes of Health (U01EB025144 and P41EB027061) and the National Science Foundation (CAREER CCF-1651825)

Purpose: To develop a strategy for training a physics-guided MRI reconstruction neural network without a database of fully sampled data sets.

Methods: Self-supervised learning via data undersampling (SSDU) for physics-guided deep learning reconstruction partitions available measurements into two disjoint sets, one of which is used in the data consistency (DC) units in the unrolled network and the other is used to define the loss for training. The proposed training without fully sampled data is compared with fully supervised training with ground-truth data, as well as conventional compressed-sensing and parallel imaging methods using the publicly available fastMRI knee database. The same physics-guided neural network is used for both proposed SSDU and supervised training. The SSDU training is also applied to prospectively two-fold accelerated high-resolution brain data sets at different acceleration rates, and compared with parallel imaging.

Results: Results on five different knee sequences at an acceleration rate of 4 shows that the proposed self-supervised approach performs closely with supervised learning, while significantly outperforming conventional compressed-sensing and parallel imaging, as characterized by quantitative metrics and a clinical reader study. The results on prospectively subsampled brain data sets, in which supervised learning cannot be used due to lack of ground-truth reference, show that the proposed self-supervised approach successfully performs reconstruction at high acceleration rates (4, 6, and 8). Image readings indicate improved visual reconstruction quality with the proposed approach compared with parallel imaging at acquisition acceleration.

Conclusion: The proposed SSDU approach allows training of physics-guided deep learning MRI reconstruction without fully sampled data, while achieving comparable results with supervised deep learning MRI trained on fully sampled data.

KEYWORDS
accelerated imaging, convolutional neural networks, deep learning, image reconstruction, parallel imaging, self-supervised learning
1 | INTRODUCTION

Data acquisition in MRI is inherently slow, necessitating the use of accelerated imaging techniques. In these approaches, data are acquired at sub-Nyquist rates and reconstructed using additional information. Parallel imaging exploits the redundancies between receiver coils and is the most clinically used approach. Compressed sensing is another method that uses the compressibility of images based on linear sparsifying transforms for a regularized reconstruction, which can also be synergistically combined with multicoil acquisitions. At high acceleration rates, parallel imaging suffers from noise amplification, whereas compressed sensing may lead to residual artifacts. Furthermore, compressed-sensing reconstruction is computationally lengthy in nature and typically requires empirical fine-tuning of regularization parameters, although recent approaches using rapid self-tuning show promise for principled parameter selection.

Recently, deep learning (DL) has gained interest for high-quality accelerated MRI. Deep learning–based MRI reconstruction algorithms can be roughly divided into two categories: purely data-driven and physics-guided. In purely data-driven approaches, a mapping between the undersampled k-space/aliased image to full k-space/artifact-free image is learned. In the so-called physics-guided methods, the knowledge of the forward encoding operator, which contains the undersampling pattern and typically the coil sensitivities, is taken into account to solve an inverse problem based on a regularized least-squares objective function. Some other techniques have worked directly with multicoil data without explicitly including the coil sensitivities. These techniques unroll an iterative reconstruction algorithm for solving this objective method for a fixed number of iterations. The unrolled network alternates between data consistency (DC) and regularization, in which the regularization is implemented implicitly using a neural network. Subsequently, these unrolled networks are trained end-to-end with a loss function that characterizes similarity with a reference image obtained from fully sampled data. The parameters of the network can be different across the unrolled iterations or shared across them.

The aforementioned physics-guided methods have been trained in a supervised manner, in which fully sampled data are used as a reference during the training. However, in many practical imaging scenarios, it is infeasible to acquire fully sampled data sets. For instance, when imaging moving organs, such as the heart, there is often a short period of time during which the data need to be acquired. Example acquisitions include real-time imaging, myocardial perfusion, and numerous contrast-enhanced scans. Another hindrance for fully sampled acquisitions in some applications include the signal decay. This is pronounced in acquisitions such as diffusion MRI with EPI, in which the signal decays quickly with $T_2^*$, thus prohibiting use of fully sampled acquisitions—especially at high resolutions. In several other scenarios such as whole-heart coronary MRI or high-resolution anatomical brain imaging, it is impractical to acquire fully sampled data sets, as the scan time becomes extremely lengthy.

Furthermore, accelerated imaging methods are often used to improve acquisition resolution. When higher acceleration rates are achievable, these are not used solely for image time reduction, but rather a trade-off is made with improved resolution. However, this newer resolution may necessitate retraining of the DL reconstruction, as neural networks do not necessarily generalize across different resolutions, as depicted in Supporting Information Figure S1. Thus, if fully sampled data are required for training at higher resolutions, this may lead to excessive scan times, even for anatomical imaging protocols, making it difficult to make protocol changes to fully use the benefits of accelerated imaging.

In this study, we sought to develop a new self-supervised learning approach to train physics-guided DL-MRI reconstruction without fully sampled reference data. The proposed self-supervised approach that we refer to as Self-Supervision via Data Undersampling (SSDU) splits the acquired k-space indices into two disjoint sets. One of these is used in the DC unit for the network, while the other set is used to define the loss function in k-space. Hence, end-to-end training and evaluation of the network is done through only the acquired measurements, without making any other assumptions about image output or characteristics. We apply the proposed self-supervised training without fully sampled data on the fastMRI knee data sets and prospectively undersampled high-resolution brain MRI data sets. These are compared with parallel imaging, compressed sensing, and a supervised training of a DL-MRI network when fully sampled reference data are available. Our results indicate that the proposed self-supervised method performs similarly to the supervised approach trained on fully sampled data, although it is trained only on undersampled data.

2 | THEORY

2.1 | Physics-guided neural networks for MRI reconstruction

Let $x$ denote the image to be recovered and $y_\Omega$ represent acquired k-space measurements with undersampling pattern $\Omega$. The forward model for the acquisition is given as

$$y_\Omega = E_\Omega x + n,$$
where \( E : \mathbb{C}^{M_1 \times M_2} \rightarrow \mathbb{C}^P \) is the encoding operator including a partial Fourier matrix sampling the locations specified by \( \Omega \) and the coil sensitivities, and \( n \in \mathbb{C}^P \) is the measurement noise. The forward model presented in Equation (1) is usually ill-conditioned due to sub-Nyquist sampling; therefore, regularizers that induce prior information are incorporated into the objective function for the reconstruction. Possible choices for the regularizer include total variation\(^{10,45,46}\), \( \ell_1 \)-norm of wavelet coefficients\(^{4,8,47}\), sparsity in adaptive transform domains\(^{9,48}\), and more recently, neural networks\(^{27,28,33}\). The image recovery is then formulated as an optimization problem

\[
\text{arg min}_x \, \| y - E x \|^2_2 + R(x),
\]

where the first term represents DC with acquired measurements, and \( R(\cdot) \) is a regularization term. The optimization problem in Equation (2) can be solved in numerous ways, including proximal gradient descent, variable splitting with quadratic penalty, and alternating direction method of multipliers, among others\(^{27,30,32,49}\). In this study, we will consider the variable splitting with quadratic penalty approach\(^{50}\) for implementation, which has also been used in previous physics-guided DL-MRI approaches\(^{28,32}\). In this method, DC and regularization are decoupled as

\[
\text{arg min}_z \, \| y - E x \|^2_2 + \mu \| x - z \|^2_2 + R(z),
\]

where \( z \) is the auxiliary variable that is initially constrained to be equal to \( x \), and \( \mu \) is the parameter for the quadratic penalty for relaxing this intermediate constrained problem to an unconstrained one. The optimization problem in Equation (3) is then solved iteratively by alternating the minimization over the variables \( x \) and \( z \) as follows:

\[
z^{(i+1)} = \text{arg min}_z \, \mu \| x^{(i)} - z \|^2_2 + R(z),
\]

\[
x^{(i)} = \text{arg min}_x \, \| y - E x \|^2_2 + \mu \| x - z^{(i)} \|^2_2 + \mu \| x - z^{(i-1)} \|^2_2,
\]

where \( x^{(0)} \) is the initial image obtained from zero-filled undersampled k-space data; \( x^{(i)} \) is the network output at iteration \( i \); and \( z^{(i)} \) is an intermediate variable. In compressed-sensing methods, these problems are solved in an iterative manner by alternating between the regularizer and DC units until a stopping criterion is met, as shown in Figure 1A.

In physics-guided DL-MRI approaches, this iterative algorithm is unrolled for a fixed number of iterations, as depicted in Figure 1B. The regularization subproblem in Equation (4) is implicitly solved using a neural network. The DC subproblem in Equation (5) has a closed-form solution

\[
x^{(i)} = (E^H E + \mu I)^{-1} (E^H y + \mu z^{(i-1)}),
\]

where \( I \) is the identity operator and \( (\ast)^H \) is the conjugate transpose operator. Equation (6) can be solved using gradient descent or conjugate gradient, which itself is unrolled for a number of iterations\(^{28}\).
2.2 Supervised training with fully sampled reference data sets

Supervised learning performs end-to-end training using ground-truth images as the reference labels for the training loss function.\(^2\)\(^{27}\) Ground-truth images are obtained through SENSE-1 coil combination, which is the sum across the coil dimension of the product of the conjugate of the coil sensitivity maps with the corresponding coil images.\(^3\)\(^{31}\)\(^{32}\) Suppose that \(\mathbf{x}_{\text{ref}}\) is the ground-truth image for subject \(i\), and \(f(\mathbf{y}_{\Omega_i}, \mathbf{E}_{\Omega_i}; \mathbf{\theta})\) denotes the output of the unrolled network that is parametrized by \(\mathbf{\theta}\) for subsampled k-space data \(\mathbf{y}_{\Omega_i}\) and corresponding encoding matrix \(\mathbf{E}_{\Omega_i}\) of the same subject \(i\). The supervised training of a physics-guided DL-MRI method can be performed by minimizing the image domain loss as follows:

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

where \(N\) is the number of fully sampled training data in the database, and \(\mathcal{L}(\ldots)\) denotes the loss between the ground-truth and network output image.\(^27\)\(^{28}\)\(^{31}\) Alternatively, supervised training may be evaluated in k-space as

\[
\min_{\mathbf{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathbf{y}_{\text{ref}i}, \mathbf{E}_{\text{full}i} f(\mathbf{y}_{\Omega_i}, \mathbf{E}_{\Omega_i}; \mathbf{\theta})).
\]

where \(\mathbf{y}_{\text{ref}i}\) is the fully sampled reference k-space, and \(\mathbf{E}_{\text{full}i}\) is the fully sampled encoding operator that transforms network output to k-space across coils. Example loss functions include \(\ell_1\) norm, \(\ell_2\) norm, mixed norm, and perception-based loss.\(^25\)\(^{32}\)\(^{51}\)\(^{52}\)\(^{53}\) We note that the subsampling patterns \(\Omega\) used in this study are equispaced and the same for all subjects. However, subsampling pattern \(\Omega\) may vary per subject (ie, indexed by \(i\)) if random subsampling is used.

2.3 Proposed self-supervised training without fully sampled reference data

As discussed previously, acquiring fully sampled data are often difficult or impossible in many scenarios, due to constraints such as organ motion, signal decay, or lengthy scan times. Such cases pose an important challenge for the practicality of DL-MRI reconstruction methods that rely on supervised training, as ground-truth data are not available for training. To tackle this problem, we propose a self-supervised approach illustrated in Figure 2, where the acquired subsampled data indices, \(\Omega\), from each scan are divided into two sets, \(\Theta\) and \(\Lambda\), as follows:

\[
\Omega = \Theta \cup \Lambda.
\]

The set of k-space locations specified by \(\Theta\) are used within the network during training in the DC units, whereas the set of k-space points in \(\Lambda\) are used to define the loss function. Thus, to enable training without using fully sampled data, the following loss function is minimized:

\[
\min_{\mathbf{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathbf{y}_{\Lambda_i}, \mathbf{E}_{\Lambda_i} f(\mathbf{y}_{\Theta_i}, \mathbf{E}_{\Theta_i}; \mathbf{\theta})).
\]

In other words, the unrolled network output image \(f(\mathbf{y}_{\Theta_i}, \mathbf{E}_{\Theta_i}; \mathbf{\theta})\), which uses only the indices specified by \(\Theta\) for DC, is transformed to k-space using the encoding operator \(\mathbf{E}_{\Lambda_i}\) specified by the k-space indices in \(\Lambda\). Then the loss is calculated in k-space with respect to the acquired k-space data at these locations. In the proposed SSDU approach, \(\Theta\) was chosen as \(\Omega \setminus \Lambda\). Thus, in our self-supervised training methodology, the unrolled network only sees the acquired k-space data at locations \(\Theta = \Omega \setminus \Lambda\) to enforce DC.

The quality of the final reconstruction (ie, the network output image) is then checked by mapping to the individual coil k-spaces through \(\mathbf{E}_{\Lambda_i}\), and checking the discrepancy to these acquired measurements at these remaining locations \(\Lambda\). Thus, the network is trained to decrease the discrepancy between the network output transformed to all the coil k-spaces and the acquired measurements that it does not see within its unrolled DC units. After the network is trained with our proposed self-supervised approach, the reconstruction for unseen test data is performed using all available measurements at locations \(\Omega\).

Our proposed self-supervised approach shares similarities with the widely used concept of cross-validation. In machine learning, cross-validation is used to evaluate how accurately a model will perform with robustness to bias and overfitting issues. Cross-validation is performed by partitioning available data into two sets, one of which is used to train the model and the other for validation (ie, checking whether the trained model generalizes to unseen data). The key difference between our approach and cross-validation is that we perform partitioning per each slice in the data set, whereas in cross-validation the whole data set is partitioned only once. The key hyperparameter for success of cross-validation is the number of folds, which should be well-designed.\(^54\) Similarly, in our proposed self-supervised approach, subset-selection mechanisms for \(\Lambda\) and \(\Theta\) are critical, which are thoroughly studied in section 3.

3 METHODS

3.1 Network and training details

The network for solving sub-problems (5) and (6) was unrolled for 10 iterations. The DC in the unrolled network was implemented with conjugate gradient method for solving
Equation (6), which itself was unrolled for 10 iterations. The neural network for solving the subproblem 5 was implemented using a convolutional neural network (CNN) based on a ResNet structure, which has shown success in other regression problems. This CNN, shown in Figure 1C, consisted of a layer of input and output convolution layers, and 15 residual blocks with skip connections that facilitate information flow during network training. Each residual block consisted of two convolutional layers, in which the first layer is followed by a rectified linear unit, and the second layer is followed by a constant multiplication layer, with factor $C = 0.1$. All layers had a kernel size of $3 \times 3$ and 64 channels. This ResNet CNN had a total of 592,129 trainable parameters, which were shared across the unrolled iterations. Coil sensitivity maps were generated from the 24 × 24 center of k-space using ESPIRiT, with a kernel size of $6 \times 6$, as well as thresholds of 0.02 and 0.95 for calibration-matrix and eigenvalue decomposition.

A normalized $\ell_1-\ell_2$ loss, defined as

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

was used for both the supervised and the proposed self-supervised training. In the supervised setting, $u$ and $v$ correspond to the reference ground-truth image/fully sampled k-space and network output image/network output k-space obtained by transforming network output images to k-space by applying a fully sampled encoding operator, whereas for the proposed self-supervised training these correspond to the acquired k-space measurements at locations specified by $\Lambda$ and the k-space corresponding to the network output image at the same locations. For supervised training, k-space loss was used throughout the study, as it outperforms the image domain loss used in our preliminary results (Supporting Information Figure S2), while also matching our self-supervised framework. Before processing, maximum absolute value of the k-space data sets was normalized to 1 in all cases. The networks were trained using the Adam optimizer with a learning rate of $10^{-3}$ unless specified otherwise, by minimizing the corresponding loss function with a batch size of 1 over 100 epochs. All training was performed using TensorFlow in Python and processed on a workstation with an Intel E5-2640V3 CPU (2.6 GHz and 256 GB memory), and an NVIDIA Tesla V100 GPU with 32 GB memory.
3.2 | Choice of the loss mask

The proposed SSDU approach divides the acquired subsampled data into two disjoint sets Θ and Λ. Furthermore, in our implementation, Λ is allowed to vary for each different slice in the training database (i.e., they can be indexed as \( \{ \Lambda_i \}_{i=1}^{N} \)). The subset Λ is retrospectively selected from the acquired k-space points, Ω, to define the loss function. Hence, unlike the data acquisition process for sampling k-space locations Ω, which is affected by concerns about contrast changes or eddy current artifacts,9 selection of Λ is not limited by any physical constraints. This is because Λ is selected after data acquisition and amounts to the selection of an index set from all possible acquired k-space locations. Thus, distribution and size of Λ were the two hyperparameters that were studied. For the distribution of Λ, a uniformly random selection among elements of Ω, as well as a variable density selection based on Gaussian random weighting, were investigated. For its size, the ratio \( \rho = |\Lambda|/|\Omega| \) was varied among 0.05, 0.1, 0.2, ..., 0.8, 0.9, where \( |\cdot| \) is the cardinality of the index set. A five-fold cross-validation was also performed on training data for quantitative assessment of the distribution of Λ, as well as a subset of \( \rho \) values among 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6.

Additionally, the effect of the overlap between Θ and Λ on the reconstruction performance was also studied. The first scenario considered was the limiting case when \( \Omega = \Theta = \Lambda \). Subsequently, we created three different partial overlap scenarios for the best-performing \( \rho \) value as (1) The first case, referred to as disjoint sets, in which there is no overlap between Θ and Λ (as originally proposed); (2) The second case, referred to as 50% overlap, in which we included 50% of points from Λ in Θ as well (more formally, \( |\Lambda \cap \Theta| / |\Lambda| = 0.5 \)); (3) Finally, we have the 100% overlap case in which all points in Λ are included in Θ as well (in this case \( \Omega = \Theta \), but Λ is a subset of Ω).

3.3 | Fully sampled knee MRI

Knee data sets were obtained from the New York University fastMRI initiative database, which was curated with an approval from the New York University School of Medicine Institutional Review Board.58 Fully sampled raw data were acquired on a clinical 3T system (Magnetom Skyra; Siemens, Erlangen, Germany) with a 15-channel knee coil using 2D turbo spin-echo sequences. The imaging parameters used for the knee data acquisitions are provided in Supporting Information Table S1.

The fully sampled raw data were undersampled retrospectively for both training and testing using equispaced sampling patterns provided in the fastMRI database with an acceleration rate \( R = 4 \).27,58 The center of k-space was fully sampled with 24 lines of autocalibrated signal. The training set consisted of 300 slices from 15 subjects for coronal proton density (PD), coronal PD weighted with fat saturation (PD-FS), and 10 subjects for sagittal PD, sagittal T2, and axial T2. Testing was performed on all slices from 10 different subjects for all knee sequences. Ground-truth images for supervised training were generated with a SENSE-1 combination of the fully sampled data.31,32 The proposed self-supervised approach was compared with supervised DL-MRI trained on a fully sampled data set and conjugate-gradient SENSE (CG-SENSE).60 Additionally, comparison to a multicoil compressed-sensing reconstruction incorporating coil sensitivities with total generalized variation (TGV) as regularizer45 was carried out for illustration purposes. However, TGV was not performed on all test data sets, as it is computationally expensive, and a comparison between supervised DL-MRI and TGV was already performed in another study.27 For TGV, the MATLAB (MathWorks, Natick, MA) implementation provided by the authors was used.45 We note that TGV and CG-SENSE approaches are shown only for comparison purposes with more traditional methods, and are not considered as competitive baseline images, consistent with previously reported results in the literature.27

3.4 | Prospectively accelerated brain MRI

Brain imaging was performed on 19 healthy subjects using a 3T Siemens Magnetom Prisma (Siemens Healthcare) system with a 32-channel receiver head coil array. The imaging protocols were approved by the local institutional review board, and written informed consent was obtained from all participants before each examination for this Health Insurance Portability and Accountability Act–compliant study. Data acquisition was performed using a standard Siemens 3D-MPRAGE sequence with the following parameters: FOV = 224 × 224 × 157 mm³, resolution = 0.7 × 0.7 × 0.7 mm³, TR/TE = 2400 ms/2.2 ms, inversion time = 1000 ms, flip angle = 8°, bandwidth = 210 Hz/pixel, 3D matrix size = 320 × 320 × 224, prospective acceleration R = 2 (equispaced in \( k_z \)), autocalibrated signal lines = 32, and acquisition orientation = sagittal. The k-space data were inverse Fourier–transformed along the readout (foot–head) direction, and these axial slices were processed individually. The prospectively undersampled brain data sets were further retrospectively undersampled to \( R = 4, 6, \) and 8 using a sheared equispaced \( k_y-k_x \) undersampling pattern,61 with a 32 × 32 autocalibrated signal region in the \( k_y-k_x \) plane. Sampling masks are provided in Supporting Information Figure S3. We note that while in principle prospectively subsampled data can be acquired at all of these different rates, we chose to use further retrospective subsampling of prospectively accelerated data, as our focus is on the reconstruction quality and this approach avoids confounding factors between different scans, such as subject motion or variations from \( T_1 \) recovery. We also note that when the self-supervised approach
was used at one of these higher acceleration rates, it only had access to the k-space data corresponding to that acceleration rate, both during training and testing. The learning rate for training was set to $5 \cdot 10^{-4}$. The training set consisted of 300 slices from 10 subjects, formed by taking the central 30 slices from each subject. Testing was performed on all slices from 9 different subjects.

The proposed self-supervised DL-MRI results were compared with the CG-SENSE method. We note that a comparison to supervised DL-MRI was not possible in this setting, as there were no fully sampled ground-truth data.

3.5 | Image evaluation

Experimental results were quantitatively evaluated using normalized mean square error (NMSE) and structural similarity index (SSIM). Additionally, qualitative assessment of the image quality was performed by an experienced radiologist. For knee MRI, the proposed self-supervised DL-MRI approach was compared with ground-truth fully sampled images, supervised DL-MRI trained on fully sampled data, and CG-SENSE at the same acceleration $R = 4$. As noted previously, TGV was not included in the comparison due to its computational complexity and availability of a previous study comparing supervised DL-MRI and TGV.\textsuperscript{27} For brain MRI, proposed self-supervised DL-MRI reconstructions at acceleration $R = 4$, 6, and 8 were compared with the CG-SENSE approach at the acquisition acceleration $R = 2$. The reader was blinded to the reconstruction method, except for the knowledge of the reference image in knee MRI data sets. The order in which the methods were shown was also randomized. There were differences between the sequences used for the fastMRI database and our institutional sequences; therefore, this knowledge allowed the radiologist to assess the baseline image quality. All five knee MRI weightings and the brain data set were evaluated on a 4-point ordinal scale, adopted from an earlier study\textsuperscript{27} for blurring (1, no blurring; 2, mild blurring; 3, moderate blurring; 4, severe blurring), SNR (1, excellent; 2, good; 3, fair; 4, poor), aliasing artifacts (1, none; 2, mild; 3, moderate; 4, severe) and overall image quality (1, excellent; 2, good; 3, fair; 4, poor). Wilcoxon signed-rank test was used to evaluate the scores with a significance level of $P < 0.05$.

4 | RESULTS

4.1 | Choice of the loss mask

Figure 3 depicts the self-supervised network training using varying subsets across slices by uniformly random and variable-density Gaussian selection of $\Lambda \subset \Omega$ for $\rho = 0.1$. Uniformly random selection of $\Lambda$ suffers from visible residual artifacts, marked by red arrows. These artifacts are further suppressed in the Gaussian-based approach, and difference images align with these observations. The quantitative assessment from five-fold cross-validation are consistent with these qualitative assessments. The median and interquartile range of SSIM values were 0.9380 [0.9197, 0.9527], and 0.9457 [0.9293, 0.9575], and NMSE values were 0.0021 [0.0016, 0.0027] and 0.0019 [0.0015, 0.0023] using uniform random selection and Gaussian selection, respectively. Supporting Information Figure S4 shows additional reconstructions for uniform random and Gaussian selection for different $\rho$ values, which further highlights that Gaussian selection consistently outperforms uniform random selection across different $\rho$ values. Thus, a variable-density Gaussian selection was used for $\Lambda$ for the remainder of the study.

Figure 4 shows the effect of network training with varying $\rho$ ∈ {0.05, 0.1, 0.2, ..., 0.8, 0.9} using variable-density Gaussian selection. Red arrows show visible residual artifacts for low $\rho$ values of 0.05, 0.1, and 0.2. As cardinality of $\Lambda$ increases toward $\rho = 0.4$, residual artifacts decrease. At $\rho = 0.4$, visible artifacts seen at lower $\rho$ values are further suppressed. Residual artifacts start to reappear starting from $\rho = 0.5$, and these artifacts become more pronounced as $\rho$ increases. The quantitative assessment from five-fold cross-validation aligns with these qualitative assessments. The median and interquartile range of SSIM values were 0.9457 [0.9293, 0.9575], 0.9477 [0.9323, 0.9591], 0.9488 [0.9328, 0.9603], 0.9507 [0.9352, 0.9614], 0.9450 [0.9297, 0.9569], and 0.9391 [0.9225, 0.9524], and NMSE values were 0.0019 [0.0015, 0.0023], 0.0018 [0.0013, 0.0023], 0.0018 [0.0014, 0.0022], 0.0017 [0.0013, 0.0021], 0.0020 [0.0015, 0.0024], and 0.0022 [0.0016, 0.0028] using Gaussian selection for $\rho$ ∈ {0.1, 0.2, 0.3, 0.4, 0.5, 0.6} respectively. Hence, $\rho = 0.4$ was used for the remainder of the study.

Figure 5 shows the effect of different degrees of overlap between $\Lambda$ and $\Theta$ for $\rho = |\Lambda|/|\Omega| = 0.4$, as well as the limiting case that uses all available data for both DC and loss (ie, $\Omega = \Theta = \Lambda$). For the limiting case with $\Omega = \Theta = \Lambda$, the reconstruction results suffer from residual noise amplification. On the other hand, when $\Lambda$ and $\Theta$ were disjoint as proposed, such noise amplifications were significantly suppressed. Qualitative SSIM and NMSE evaluations of these methods over the data set are presented in Supporting Information Table S2, indicating that for different rates of overlap between $\Lambda$ and $\Theta$ with $\rho = 0.4$, the performance degrades as the amount of overlap increases. Thus, disjoint sets were used for the remainder of the study.

4.2 | Knee MRI

Figure 6 demonstrates the reconstruction results of coronal PD images using CG-SENSE, TGV, supervised DL-MRI,
and the proposed self-supervised DL-MRI approach along with the ground-truth reference, as well as difference images with respect to this reference. The CG-SENSE and TGV approaches suffer from visible residual artifacts, marked by red arrows, with the latter having fewer artifacts. The proposed self-supervised and supervised DL-MRI approaches successfully remove the residual artifacts, while achieving similar qualitative and quantitative performance. Quantitative metrics and difference images displayed in the figure are in agreement with these observations. Supporting Information Figure S5 shows the training loss curves for both approaches, in which loss decreases over epochs in a similar trend.

The same trends were observed for coronal PD-FS as depicted in Figure 7. Both the proposed and supervised DL-MRI approaches show similar performance, while improving the suppression of residual artifacts that are visible in the CG-SENSE and TGV methods. Quantitative evaluation and the residual artifacts apparent in the difference images also highlight these observations. Supporting Information Figure S6 shows the reconstruction results for axial T2, sagittal T2, and sagittal PD-weighted knee data sets, which align with the observation from coronal weighted knee data sets.

Figure 8 shows a box plot displaying the median and interquartile range (25th-75th percentile) of the quantitative metrics, SSIM and NMSE, across all test data sets for
each knee sequence. In all sequences, supervised and self-supervised DL-MRI approaches achieve similar quantitative performance for both SSIM and NMSE, while significantly outperforming the CG-SENSE approach. We note again that TGV was not included in these comparisons, as it is computationally expensive, and a comparison between supervised DL-MRI and TGV was already performed in an earlier study.\textsuperscript{27}

4.3 | Prospectively accelerated brain MRI

Figure 9 depicts a sagittal slice of the 3D-MPRAGE data set at acquisition acceleration $R = 2$ and further retrospective acceleration $R = 4$, 6, and 8 reconstructed with CG-SENSE, as well as $R = 4$, 6, and 8 reconstructed with the proposed self-supervised DL-MRI on a representative test subject, following reformatting to the original acquisition (sagittal) plane. The CG-SENSE method suffers from significant noise amplification at higher acceleration rates. Self-supervised DL-MRI successfully performs reconstruction at these higher acceleration rates, while achieving lower noise level and similar overall image quality with CG-SENSE at $R = 2$. Results from another subject are depicted in Supporting Information Figure S7 and show similar trends. The TGV method was not applied due to the high computational runtime across all axial slices, and supervised DL-MRI could not be applied in this setting due to the lack of fully sampled references.
4.4 Image evaluation scores

Figure 10 summarizes the results of the reader study for knee and brain data sets. For knee data sets, both supervised and self-supervised DL-MRI approaches get comparable scores to the reference image in terms of SNR, blurring, aliasing artifacts, and overall image quality. There was no statistical difference between reference and DL-MRI approaches in
terms of the evaluation criteria for all knee sequences, except for blurring between reference and DL-MRI approaches in coronal PD-FS. The CG-SENSE approach was significantly outperformed by both DL-MRI approaches, while showing statistically significant differences to the reference and both DL-MRI approaches for all knee sequences, except in blurring for coronal PD and PD-FS sequences. More comprehensive bar plots of the average scores including CG-SENSE and supervised training with image domain loss as in Equation (7) are presented in Supporting Information Figure S8.

For the 3D-MPRAGE data set, DL-MRI reconstructions trained using the proposed self-supervised approach at acceleration rates 4, 6, and 8 show similar statistical properties in terms of SNR and blurring with CG-SENSE at acquisition R = 2. However, in terms of aliasing artifacts and overall image quality, the proposed self-supervised approach at all three acceleration rates (R = 4, 6, and 8) outperforms CG-SENSE at R = 2. In terms of aliasing artifacts, proposed self-supervised approach for rates 4 and 6 shows similar quantitative behavior with each other, while significantly improving upon self-supervised
DL-MRI at R = 8 and CG-SENSE at R = 2, which perform statistically similar among themselves. The proposed self-supervised approach at R = 4 shows the best overall image quality and shows statistically significant differences with self-supervision at R = 6 and 8 and CG-SENSE at R = 2. As expected, the overall image quality decreases with higher acceleration rates using the proposed self-supervised DL-MRI approach, although these techniques still outperform CG-SENSE at R = 2.

5 | DISCUSSION

In this study, we developed a framework for self-supervised training of physics-based DL-MRI reconstruction without fully sampled data. The proposed approach split the acquired undersampled k-space indices into two disjoint sets, Θ and Λ, in which the former was used across the unrolled network to enforce DC, whereas the latter was used to define the loss function for the training. The results on retrospectively undersampled knee data sets showed that our SSDU approach achieves comparable results with a supervised DL-MRI approach using the same neural network architecture, while outperforming conventional CG-SENSE and TGV approaches. Results on prospectively undersampled brain data sets, for which supervised learning methods cannot be applied due to unavailability of fully sampled data, further confirmed the effectiveness of the proposed self-supervised training approach for DL-MRI reconstruction. These reconstructions at higher acceleration rates of 4, 6, and 8 visually outperformed CG-SENSE at R = 2 according to the reader study. We note that CG-SENSE was implemented without regularization, and its performance may be improved using Tikhonov regularization with the regularization parameter selected over a training set.27

Most DL-MRI approaches use supervised learning for network training, to provide improved accelerated MRI reconstruction.28,29,32,33,59 However, acquiring fully sampled data is challenging in many practical scenarios of interest. These may be due to constraints on timing, physiological constraints, signal decay, or long scan times.38-42 As an example, the fully sampled acquisition for the 3D-MPRAGE sequence with the resolution used in this study would be more than 15 minutes,41 which is impractical for large studies and may lead to patient discomfort. Furthermore, such long scan times increase susceptibility to motion artifacts, which would be more pronounced at these high resolutions. To further highlight the need for training data, we also performed experiments on prospectively subsampled snapshot cardiac MRI, in which it is infeasible to collect the ground-truth data. Results from these experiments are provided in Supporting Information Figure S9, which shows the applicability of our method in this setting as well. Thus, being able to
train DL-MRI reconstruction methods without fully sampled data is imperative to broaden their application to settings in which such data are challenging to acquire, where supervised training is no longer practical. Furthermore, this may also facilitate the integration of DL-MRI methods to many clinical scans that readily include a form of accelerated imaging, most commonly in the form of parallel imaging, by enabling the use of prospectively undersampled raw k-space data for training.

Given the importance of training without fully sampled data, there have been several studies that have tried to tackle this issue. For purely data-driven de-aliasing of single-coil data using image domain to image-domain mapping, a self-supervised approach has been proposed using a mixture of measurement and k-space losses. Unlike our approach, it uses all available data for training and loss (ie, identical sets). As a result, the reconstructions suffer from visible noise amplifications that also

FIGURE 10 The image reading results from the clinical reader study for knee and brain data sets. Bar plots show average reader scores and their standard deviation across the test subjects. Statistical testing was performed using one-sided Wilcoxon single-rank test, with * showing significant statistical difference with \( p < .05 \). For knee MRI, both supervised and self-supervised DL-MRI approaches get comparable scores to the reference image in terms of SNR, blurring, aliasing artifacts, and overall image quality. There was no statistical difference between reference and DL-MRI approaches in terms of the evaluation criterions for the knee data sets, except for blurring between reference and DL-MRI approaches in coronal PD-FS. For brain MRI, CG-SENSE at \( R = 2 \) and self-supervision at \( R = 4, 6, \) and 8 do not show any significant differences in terms of SNR and blurring. Self-supervision at all rates were evaluated to be significantly improved compared with CG-SENSE in terms of aliasing artifacts and overall image quality. Additionally, self-supervision at \( R = 6 \) and 8 were also significantly worse than self-supervision at \( R = 4 \) in terms of overall image quality.
align with our observation about the use of identical sets in Figure 5. An alternative approach, which assumes the same data are acquired with two separate acquisitions using different undersampling patterns, was also proposed extending on the Noise2Noise denoising framework. In the same image-domain reconstruction setting, a self-supervised learning scheme using cycleGANs with optimal transport cost minimization was proposed, although initial results exhibit blurring artifacts. Although purely data-driven image domain methods have been used for DL-MRI reconstruction, physics-guided DL-MRI techniques are more desirable, as they offer a degree of interpretability by incorporating domain knowledge on the MRI encoding mechanism.

In this physics-guided setting, an earlier work used the output of a regularized CG-SENSE algorithm based on compressed sensing as the reference for supervised training, showing that such training may overperform the compressed-sensing output, as some images are overregularized while others are underregularized. However, this approach assumes that the compressed-sensing algorithm output will be a reliable estimate of the image without residual aliasing artifacts, and is thereby limited by sampling strategies and acquisition acceleration rates, as high acceleration rates or equispaced sampling may lead to degradation in the compressed-sensing results. More recently, an unpaired learning approach using Wasserstein GANs was proposed, but this procedure still assumes the presence of high-quality images, although not requiring pairwise matching with undersampled data. Another approach uses the so-called unsupervised basis pursuit, in which the unrolled network consists of regularizer units followed by several consecutive DC units. This approach uses the current output of the DC unit as the training label, and iteratively updates both network parameters and this training label in a method reminiscent of semisupervised training. This method was investigated with random undersampling patterns, in which intermediate outputs tend to suffer from noise amplification but without significant residual artifacts. In this setting, this approach was able to reduce noise further, even though noise amplification was observed when compared with supervised training. However, this method was not investigated for equispaced undersampling, as is the focus of this study, in which intermediate DC outputs are both noisy and likely to have residual aliasing artifacts. Thus, the utility of this method in equispaced undersampling is unclear and warrants further investigation. In contrast, our SSDU approach uses physics-guided DL-MRI reconstruction, while not making any explicit assumptions about the final output in image space. In particular, we do not enforce the output of our network to align with a generative model or consider intermediate estimates as reference output for training. The training in SSDU only considers the acquired k-space data to evaluate the reconstruction quality, in effect using a physics-guided self-supervision approach. Furthermore, SSDU works for both equispaced undersampling patterns, as is the focus of the study, and random undersampling patterns (results not shown). Note that the former was considered to be more challenging for physics-guided DL-MRI reconstruction in previous studies, as networks trained with equispaced sampling were shown to generalize well to random sampling, but not vice versa.

Our training method is also reminiscent of the broader and fundamental concept of cross-validation in machine learning and statistics. When testing generalizability, the training database is partitioned into two sets of complementary data sets: one that is used for training the model (often called the training set), and one that is used to assess the performance in unseen data (often called the validation/testing set). In our approach, we do a similar partitioning of the acquired data to two sets we denote as Θ and Λ. The main difference to typical cross-validation is that our partitioning is done for each subject in the training set from the database. However, the intuition for partitioning within the network is similar, as the unrolled network only sees Θ for data consistency during training, whereas Λ is only used to establish the network loss. Indeed, as our experiments in Figure 5 show that when Θ and Λ are taken to be the same as Ω, such training leads to poor image quality with insufficient removal of aliasing artifacts and noise amplification, as the DC unit operating on the full Ω inherently matches well with the acquired data at these locations.

Selection of the loss mask, Λ, plays an important role in the performance of the proposed self-supervised training. One major design advantage is that because it only exists in post-processing, it can be chosen freely among all of the acquired measurements retrospectively, without physical constraints that are imposed during acquisition. Thus, even though 40% of the acquired indices in Ω were included in Λ, this is not the equivalent to training with an approximate eight-fold accelerated acquisition, especially for the 2D setting, as the points in Λ do not need to constitute fully sampled readout encoding lines along kₜ. This point is further illustrated in Supporting Information Figure S10, in the context of supervised training. This advantage is not as clear in the training for the 3D brain data set in this study, as the data had to be inverse Fourier-transformed along the foot–head readout direction and axial slices had to be processed due to memory issues in the GPUs. In this case, the sheared equispaced kₜ-kₖ undersampling pattern readily does not include any lines; thus, the selection of Λ may affect the DC units more substantially than in the 2D knee MRI experiments. Accordingly, the self-supervised approach is expected to show more gains and better reconstruction quality at higher acceleration rates for 3D imaging if 3D neural networks can be used. Thus, memory-efficient 3D neural network designs may warrant further investigation, although it is beyond the scope of the current study.

The data reduction arising from data splitting between Θ and Λ poses more challenges for training and reconstruction...
at higher acceleration rates, even for 2D acquisitions. This was further investigated to check how the performance of self-supervised and supervised training would change at higher acceleration rates when all of the training parameters and data sets are the same, as described previously. The results shown in Supporting Information Figure S11 indicate that both training methods perform similarly at R = 4 and 6 for knee MRI. However, at R = 8, in which the supervised training is able to suppress artifacts albeit at the cost of blurring artifacts, the self-supervised approach starts suffering from additional residual aliasing artifacts. Thus, at higher acceleration rates, in which reconstructions from the supervised training can operate without aliasing artifacts but with quality degradation, the self-supervised approach faces additional challenges including residual aliasing, due to the scarcity of data, especially after the splitting to two sets. The problem of data scarcity has been addressed by several important transfer learning methods when using supervised training with fully sampled data sets. 74,75 These approaches pretrain neural networks on fully sampled large data sets, and then fine-tune them on smaller data sets of interest. In such cases, if the smaller data set of interest is additionally not fully sampled, then the proposed self-supervised approach may be combined synergistically with transfer learning to tackle this challenging issue of both data scarcity and not having fully sampled data, although this was beyond the scope of this study. We also note that there are differences between the weights of the networks from supervised and self-supervised training approaches. However, a quantitative difference, such as NMSE, between learned weights of these two training approaches does not directly translate to reconstruction performance, as shown by our results. Nonetheless, it is noteworthy that two different trained networks with differences among their weights have similar reconstruction performance during the testing stage, further alluding to the complexity of the parameter space for the neural network.

All experiments in this study were based on Cartesian acquisitions. The proposed self-supervised approach can be extended to non-Cartesian acquisitions. In non-Cartesian acquisitions such as radial or spiral acquisitions, one can choose the subsets for training and loss mask from the acquired radial spokes and spirals, similar to Cartesian acquisitions used in this study, as this amounts to selecting a subset of individual k-space points on the spokes or spirals. We also note that for non-Cartesian acquisitions, the encoding operator also contains the gridding/degridding operation to account for nonuniform Fourier transforms. These extensions were not investigated, as it was beyond the scope of the current work.

In this study, we compared uniformly random selection with a variable-density approach based on Gaussian weighting for selecting Λ. In our experiments, the latter selection was favored, as it statistically outperformed and visibly improved upon the former. A self-supervised mask selection during the network training may further remove these hyperparameters and potentially lead to further improvements in reconstruction. However, this is a difficult problem, which warrants further investigation, and is beyond the scope of the current study. Using different distributions for selecting a number of distinct Θ and Λ pairs per subject may further improve performance, but currently these distributions would need to be chosen empirically. Due to the ad hoc nature of such a process and the wide range of available distributions, this was not explored in detail, but this idea also warrants more investigation in the context of self-supervised mask selection in future works. We also investigated the reconstruction performance using the same sets, Θ and Λ, across all training slices versus letting these vary across slices as Θi and Λi, as proposed. Although one can choose these sets to be the same for all slices, such an approach bears the risk of a suboptimal loss mask being used for all slices. Hence, having different sets for each slice in the data set may provide additional robustness. Supporting Information Figure S12 shows that having different loss and training sets for each slice shows slight improvement over using the same sets across the entire training data set. Finally, a heuristic choice was made to keep 4 × 4 central k-space lines in the Θ set, as the DC units did not work well without these high-energy components. In our experience, use of larger (8 × 8 or 16 × 16) or smaller (2 × 2) regions deteriorated the overall performance.

The same residual network structure for regularizer and unrolled conjugate gradient for DC units was used throughout the study. However, our approach is not restricted to these network and DC unit choices. Alternative approaches, such as a DenseNet, U-Net or variational neural network as a regularizer CNN,27,76,77 or gradient descent for the DC unit are also possible.27,33 However, these were not explored, as such network optimization was not the focus of our study. Instead we fixed one architecture and used this for both supervised and self-supervised training. In this study, we also shared the regularizer CNN parameters across the unrolled network28,33 to enable training with a smaller training data set. However, it is possible to use different parameters for each unrolled regularizer unit27,31 at the cost of a higher number of trainable parameters. A comparison between supervised training with shared and nonshared parameters in the unrolled network is provided in Supporting Information Figure S13. The results indicate that the two approaches perform similarly in terms of qualitative and quantitative assessments.

Selection of proper loss functions also plays a vital role for network training. The ℓ2 loss is a frequently used metric in DL-MRI with promising results,20,28 but it is sensitive to outliers. On the other hand, ℓ1 loss is more robust to outliers. Hence, we used a normalized ℓ1−ℓ2 loss to take advantage of the superior properties of each loss while minimizing their disadvantages.53 Other choices of losses such
as discriminative losses have also been popular for supervised training of DL-MRI methods.\(^\text{33,78}\) There have also been studies that incorporate the conventional loss functions such as \(\ell_1\) or \(\ell_2\) into adversarial losses.\(^\text{25,79-81}\) To the best of our knowledge, there are no works that use an adversarial loss in k-space, but such an extension may benefit the reconstruction quality when using the proposed self-supervision approach.

6 | CONCLUSIONS

The proposed training framework allows training of physics-guided DL-MRI reconstruction without requiring fully sampled data, while performing similar to conventional supervised DL-MRI approaches.

ACKNOWLEDGMENTS

Knee MRI data were obtained from the New York University (NYU) fastMRI initiative database.\(^\text{58}\) The NYU fastMRI database was acquired with the relevant institutional review board approvals as detailed in the original paper.\(^\text{58}\) The NYU fastMRI investigators provided data but did not participate in the analysis or writing of this report. A listing of NYU fastMRI investigators, subject to updates, can be found at fastmri.med.nyu.edu.

ORCID

Burhaneddin Yaman \(\text{https://orcid.org/0000-0003-0791-5900}\)

Seyed Amir Hossein Hosseini \(\text{https://orcid.org/0000-0002-0687-0782}\)

Steen Moeller \(\text{https://orcid.org/0000-0003-1698-7260}\)

Mehmet Akçakaya \(\text{https://orcid.org/0000-0001-6400-7736}\)

REFERENCES

1. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized auto-calibrating partially parallel acquisitions (GRAPPA). Magn Reson Med. 2002;47:1202-1210.
2. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: Sensitivity encoding for fast MRI. Magn Reson Med. 1999;42:952-962.
3. Lustig M, Pauly JM. SPIRiT: Iterative self-consistent parallel imaging from arbitrary k-space. Magn Reson Med. 2010;64:457-471.
4. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn Reson Med. 2007;58:1182-1195.
5. Haldar JP, Hernando D, Liang ZP. Compressed-sensing MRI with random encoding. IEEE Trans Med Imaging. 2011;30:893-903.
6. Trzasko J, Manduca A. Highly undersampled magnetic resonance image reconstruction via homotopic \(\ell_1\) -minimization. IEEE Trans Med Imaging. 2009;28:106-121.
7. Ye JC, Tak S, Han Y, Park HW. Projection reconstruction MR imaging using FOCUSS. Magn Reson Med. 2007;57:764-775.
8. Akcakaya M, Nam S, Hu P, et al. Compressed sensing with wavelet domain dependencies for coronary MRI: A retrospective study. IEEE Trans Med Imaging. 2011;30:1090-1099.
9. Akcakaya M, Basha TA, Goddu B, et al. Low-dimensional structure self-learning and thresholding: Regularization beyond compressed sensing for MRI Reconstruction. Magn Reson Med. 2011;66:756-767.
10. Block KT, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. Magn Reson Med. 2007;57:1086-1098.
11. Liang D, Liu B, Wang J, Ying L. Accelerating SENSE using compressed sensing. Magn Reson Med. 2009;62:1574-1584.
12. Otazo R, Kim D, Axel L, Sodickson DK. Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI. Magn Reson Med. 2010;64:767-776.
13. Robson PM, Grant AK, Madhuranthakam AJ, Lattanzi R, Sodickson DK, McKenzie CA. Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions. Magn Reson Med. 2008;60:895-907.
14. Chang Y, Liang D, Ying L. Nonlinear GRAPPA: A kernel approach to parallel MRI reconstruction. Magn Reson Med. 2012;68:730-740.
15. Madore B. UNFOLD-SENSE: A parallel MRI method with self-calibration and artifact suppression. Magn Reson Med. 2004;52:310-320.
16. Sung K, Hargreaves BA. High-frequency subband compressed sensing MRI using quadruplet sampling. Magn Reson Med. 2013;70:1306-1318.
17. Yang Y, Sun J, Li H, Xu Z. ADMM-CSNet: A deep learning approach for image compressive sensing. IEEE Trans Pattern Anal Mach Intell. 2018;42:521-538.
18. Shahbloo M, Ilicak E, Tofighi M, Saritas EU, Cetin AE, Cukur T. Projection onto epigraph sets for rapid self-tuning compressed sensing. IEEE Trans Med Imaging. 2019.38:1677-1689.
19. Ramani S, Liu Z, Rosen J, Nielsen JF, Fessler JA. Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV andSURE-based methods. IEEE Trans Image Process. 2012;21:3659-3672.
20. Liang D, Cheng J, Ke Z, Ying L. Deep magnetic resonance image reconstruction: Inverse problems meet neural networks. IEEE Sig Proc Mag. 2020;37:141-151. http://dx.doi.org/10.1109/msp.2019.2950557.
21. Wang S, Su Z, Ying L, et al. Accelerating magnetic resonance imaging via deep learning. In: Proceedings of the IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 2016. pp 514-517.
22. Lee D, Yoo J, Tak S, Ye JC. Deep residual learning for accelerated MRI using magnitude and phase networks. IEEE Trans Biomed Eng. 2018;65:1985-1995.
23. Akcakaya M, Moeller S, Weingartner S, Ugurbil K. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. Magn Reson Med. 2019;81:439-453.
24. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature. 2018;555:487-492.
25. Mardani M, Gong E, Cheng JY, et al. Deep generative adversarial neural networks for compressive sensing MRI. IEEE Trans Med Imaging. 2019;38:167-179.
26. Han Y, Sunwoo L, Ye JC. k-space deep learning for accelerated MRI. *IEEE Trans Med Imaging*. 2020;39:377-386.
27. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. 2018;79:3055-3071.
28. Aggarwal HK, Mani MP, Jacob M. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Trans Med Imaging*. 2019;38:394-405.
29. Zhang J, Ghanem B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, 2018. pp 1828-1837.
30. Yang Y, Sun J, Li H, Xu Z. Deep ADMM-Net for compressive sensing MRI. *Adv Neural Inf Process Syst*. 2016;29:10-18.
31. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. 2018;37:491-503.
32. Qin C, Schlemper J, Caballero J, Price AN, Hajnal JV, Rueckert D. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. 2019;38:280-290.
33. Mardani M, Sun Q, Donoho D, et al. Neural proximal gradient descent for compressive imaging. *Adv Neural Inf Process Syst*. 2018;9573-9583.
34. Gregor K, LeCun Y. Learning fast approximations of sparse coding. In: Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 2010. pp 399-406.
35. Hosseini SAH, Yaman B, Moeller S, Hong M, Akccakaya M. Dense recurrent neural networks for accelerated MRI: History-cognizant unrolling of optimization algorithms. *IEEE J Sel Top Sign Proc*. In press. https://doi.org/10.1109/JSTSP.2020.3003170.
36. Cheng JY, Pauly JM, Vasanawala SS. Multi-channel image reconstruction with latent coils and adversarial loss. In: Proceedings of the 27th Annual Meeting of ISMRM, Montréal, Canada, 2019. Abstract 1249.
37. Wang P, Chen EZ, Chen T, Patel VM, Sun S. Pyramid convolutional RNN for MRI reconstruction. arXiv:1912:00543; 2019.
38. Haji-Valizadeh H, Rahsepar AA, Collins JD, et al. Group COMwBaNCS. Validation of highly accelerated real-time cardiac cMRI with radial k-space sampling and compressed sensing in patients at 1.5T and 3T. *Magn Reson Med*. 2018;79:2745-2751.
39. Coelho-Filho OR, Rickers C, Kwong RY, Jerosch-Herold M. MR myocardial perfusion imaging. *Radiology*. 2013;266:701-715.
40. Kellman P, Hansen MS. T1-mapping in the heart: Accuracy and precision. *J Cardiovasc Magn Reson*. 2014;16:2.
41. Ugurbil K, Xu J, Auerbach EJ, et al. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage*. 2013;80:80-104.
42. Setsompop K, Kimmlingen R, Eberlein H, et al. Pushing the limits of in vivo diffusion MRI for the Human Connectome Project. *NeuroImage*. 2013;80:220-233.
43. Jung H, Sung K, Nayak KS, Kim EY, Ye JC. k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI. *Magn Reson Med*. 2009;61:103-116.
44. Gamper U, Boesiger P, Kozerke S. Compressed sensing in dynamic MRI. *Magn Reson Med*. 2008;59:365-373.
45. Knoll F, Bredies K, Pock T, Stollberger R. Second order total generalized variation (TGV) for MRI. *Magn Reson Med*. 2011;65:480-491.
46. Hu Y, Jacob M. Higher degree total variation (HDTV) regularization for image recovery. *IEEE Trans Image Process*. 2012;21:2559-2571.
47. Doneva M, Bornert P, Eggers H, Stehning C, Senegas J, Mertins A. Compressed sensing reconstruction for magnetic resonance parameter mapping. *Magn Reson Med*. 2010;64:1114-1120.
48. Ravishankar S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans Med Imaging*. 2011;30:1028-1041.
49. Fessler Jeffrey A. Optimization Methods for Magnetic Resonance Image Reconstruction: Key Models and Optimization Algorithms. *IEEE Signal Processing Magazine*. 2020;37 (1):33–40. http://dx.doi.org/10.1109/msp.2019.2943645.
50. Afonso MV, Bioucas-Dias JM, Figueiredo MA. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans Image Process*. 2010;19:2345-2356.
51. Hammernik K, Knoll F, Sodickson DK, Pock T. L2 or not L2: Impact of loss function design for deep learning MRI reconstruction. In: Proceedings of the 25th Annual Meeting of ISMRM, Honolulu, Hawaii, 2017. p 687.
52. Quan TM, Nguyen-Duc T, Jeong WK. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans Med Imaging*. 2018;37:1488-1497.
53. Knoll F, Hammernik K, Zhang C, et al. Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. *IEEE Signal Process Mag*. 2020:128-140.
54. Arlot S, Lerasle M. Choice of V for V-fold cross-validation in least-squares density estimation. *J Mach Learn Res*. 2016;17:7256-7305.
55. Timofte R, Agustsson E, Van Gool L, Yang M-H, Zhang L. Ntire 2017 challenge on single image super-resolution: Methods and results. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, Hawaii, 2017. pp 114-125.
56. Uecker M, Lai P, Murphy MJ, et al. ESPiRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magn Reson Med*. 2014;71:990-1001.
57. Yaman B, Hosseini SAH, Moeller S, Ellermann J, Ugurbil K, Akccakaya M. Self-supervised physics-based deep learning MRI reconstruction without fully-sampled data. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 2020. pp. 921-925.
58. Knoll F, Zbontar J, Sriman A, et al. fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR Image reconstruction using machine learning. *Radi Art Intel*. 2020;2:e190007. http://dx.doi.org/10.1148/ryai.2020190007.
59. Knoll F, Hammernik K, Kobler E, Pock T, Recht MP, Sodickson DK. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magn Reson Med*. 2019;81:116-128.
60. Pruessmann KP, Weiger M, Bornert P, Boesiger P. Advances in sensitivity encoding with arbitrary k-space trajectories. *Magn Reson Med*. 2001;46:638-651.
61. Breuer FA, Blaimer M, Mueller MF, et al. Controlled aliasing in volumetric parallel imaging (2D CAIPIRINA). *Magn Reson Med*. 2006;55:549-556.
62. Senouf O, Vedula S, Weiss T, Bronstein A, Michaelovich O, Zibulevsky M. Self-supervised learning of inverse problem solvers in medical imaging. In: *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*. DART 2019, MIL3ID 2019. Lecture Notes in Computer Science, vol 11795; 2019.
63. Huang P, Zhang C, Li H, et al. Deep MRI reconstruction without ground truth for training. In: Proceedings of 27th Annual Meeting of ISMRM, Montréal, Canada, 2019. Abstract 4668.

64. Liu J, Sun Y, Eldeniz C, Gan W, An H, Kamilov US. RARE: Image reconstruction using deep priors learned without ground truth. In Proceedings of the 35th International Conference on Machine Learning, PMLR, vol. 80; 2018:2965-2974.

65. Lehtinen J, Munkberg J, Hasselgren J, et al. Noise2noise: Learning image restoration without clean data. arXiv: 1803.04189; 2018.

66. Sim B, Oh G, Lim S, Ye JC. Optimal transport, CycleGAN, and penalized LS for unsupervised learning in inverse problems. arXiv: 1909.12116; 2019.

67. Chen F, Taviani V, Malkiel I, et al. Variable-density single-shot fast spin-echo MRI with deep learning reconstruction by using variational networks. Radiology. 2018;289:366-373.

68. Lei K, Mardani M, Pauly JM, Vasawanala SS. Wasserstein GANs for MR imaging: From paired to unpaired training. arXiv: 1910.07048; 2019.

69. Tamir JI, Stella XY, Lustig M. Unsupervised deep basis pursuit: Learning reconstruction without ground-truth data. In: Proceedings of the 27th Annual Meeting of ISMRM, Montréal, Canada, 2019. Abstract #660.

70. Tamir JI, Yu SX, Lustig M. Unsupervised deep basis pursuit: Learning inverse problems without ground-truth data. arXiv: 1910.13110; 2019.

71. Hammernik K, Knoll F, Sodickson D, Pock T. On the influence of sampling pattern design on deep learning-based MRI reconstruction. In: Proceedings of 25th Annual Meeting of ISMRM, Honolulu, Hawaii, 2017. p 644.

72. Browne MW. Cross-validation methods. J Math Psychol. 2000;44:108-132.

73. Kellman M, Zhang K, Tamir J, Bostan E, Lustig M, Waller L. Memory-efficient learning for large-scale computational imaging. arXiv: 2003.05551; 2020.

74. Dar SUH, Öz bey M, Çatlı AB, Çukur T. A transfer-learning approach for accelerated MRI using deep neural networks. Magn Reson Med. 2020;84:663-685.

75. Han Y, Yoo J, Kim HH, Shin HJ, Sung K, Ye JC. Deep learning with domain adaptation for accelerated projection-reconstruction MR. Magn Reson Med. 2018;80:1189-1205.

76. Hu Y, Shi X, Tian Q, et al. Reconstruction of multi-shot diffusion-weighted MRI using unrolled network with U-nets as priors. In: Proceedings of the 27th Annual Meeting of ISMRM, Montréal, Canada, 2019. Abstract 0465.

77. Yaman B, Hosseini SAH, Moeller S, Akcakaya M. Comparison of neural network architectures for physics-driven deep learning MRI reconstruction. In: Proceedings of IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, Canada, 2019. pp 155–159.

78. Sanchez I, Vilaplana V. Brain MRI super-resolution using 3D generative adversarial networks. In: Proceedings of the International Conference on Medical Imaging with Deep Learning, Amsterdam, Netherlands, 2018. Paper #114.

79. Lei L, Mardani M, Vasawanala SS, Pauly JM. Semi-supervised super-resolution GANs for MRI reconstruction. In: Proceedings of the 27th Annual Meeting of ISMRM, Montréal, Canada, 2019. Abstract 4648.

80. Yang G, Yu S, Dong H, et al. DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. IEEE Trans Med Imaging. 2018;37:1310-1321.

81. Dar SUH, Yurt M, Shahdloo M, Ildiz ME, Cukur T. Synergistic reconstruction and synthesis via generative adversarial networks for accelerated multi-contrast MRI. arXiv: 1805.10704; 2018.

**SUPPORTING INFORMATION**

Additional Supporting Information may be found online in the Supporting Information section.

**FIGURE S1** Reconstruction results for the generalization performance of supervised training across different image matrix sizes. The networks are trained by taking actual k-space, the central one-half of the k-space (ie, reducing the resolution by two-fold), and the central one-fourth of the k-space (ie, reducing the resolution by four-fold). All trained networks were then applied on actual-size data to test generalization. The generalization performance of convolutional neural networks on actual image size degrades as the training image size get smaller, with one-fourth k-space performing the worst

**FIGURE S2** Reconstruction results for supervised training with image domain (Eq. 7) and k-space (Eq. 8) losses. When using image domain loss, the reconstruction suffers from residual artifacts (red arrows), whereas using k-space loss suppresses these artifacts. Difference images also show that the supervised training with k-space loss has fewer residual artifacts. Across the data set, the two approaches perform quantitatively similar. The median and interquartile range for structural similarity index (SSIM) values across test data set were 0.967 [0.955, 0.978] and 0.966 [0.956, 0.977], and for normalized mean squared error (NMSE) values were 0.001 [0.001, 0.002] and 0.001 [0.001, 0.002] for supervised with image domain and k-space losses, respectively

**FIGURE S3** Subsampling masks used in the brain MRI study. Prospective subsampling was equipased with R = 2 in k_y and 32 autocalibrated signal (ACS) lines. Subsampling patterns for R = 4, 6, and 8 were obtained by sheared subsampling, while keeping the center 32 × 32 ACS region in the k_y-k_x plane

**FIGURE S4** Reconstruction results from self-supervised training with uniform random selection and variable-density Gaussian selection of Λ for ρ ∈ {0.1, 0.2, and 0.4}. Gaussian random selection consistently outperforms the uniform random selection at all ρ values in terms of reconstruction quality and suppression of residual artifacts, which is also highlighted in the difference images. For ρ ∈ {0.1 and 0.2}, both uniform and Gaussian random selection show visible residual artifacts, marked by red arrows, with the former showing more residual artifacts. For ρ = 0.4, uniform random selection still suffers from visible residual artifacts, whereas Gaussian selection further suppresses those artifacts and achieves artifact-free reconstruction. Difference images further confirm these observations
except in blurring for coronal PD and PD-FS sequences. Finally, we also note that the supervised training with k-space loss (Figure 10) outperforms supervised training with image domain loss in terms of reader scores for axial T₂, coronal PD-FS, and sagittal T₂ sequences.

**FIGURE S9** Reconstructed images from an eight-fold accelerated snapshot cardiac MRI data with $1.3 \times 1.3$ mm$^2$ in-plane resolution, acquired using a transient balanced SSFP sequence. These type of acquisitions are used commonly in cardiac parametric mapping, in which the image data for one contrast weighting need to be acquired within the diastolic quiescence of one heartbeat. A fully sampled acquisition at this higher resolution would take over 700 ms, which is impossible to fit in the diastolic quiescence of a single heartbeat. Training data were acquired on 14 subjects, and testing was performed on a different subject using the approach described in the manuscript. The proposed self-supervised approach achieves high-quality reconstruction, outperforming CG-SENSE, which suffers from residual artifacts and high noise.

**FIGURE S10** Reconstruction results for proposed self-supervised training at $R = 4$, supervised training at $R = 4$, $R = 4$ with $\rho = 0.4$, and $R = 8$. The amount of data used for self-supervised/supervised training at $R = 4$ (24 ACS lines) with $\rho = 0.4$ is 21 120 k-space points, which is approximately equivalent to training the network with an equispaced undersampling pattern of $R = 8$ (24 ACS lines) with 21 440 k-space points. The results show that supervised training at $R = 4$ with $\rho = 0.4$ is visibly similar with supervised and proposed self-supervised training at $R = 4$, and outperforms supervised training at $R = 8$. These results are visibly highlighted in difference images, which show supervised training at $R = 8$ suffering from residual artifacts, whereas other approaches show similar performance. Quantitative metrics on the test data set aligns with these qualitative assessments. The median and interquartile ranges for SSIM across the test data set were 0.961 [0.947, 0.972], 0.966 [0.956, 0.977], 0.966 [0.954, 0.976], and 0.929 [0.908, 0.950], and NMSEs were 0.002 [0.001, 0.002], 0.001 [0.001, 0.002], 0.002 [0.001, 0.002], and 0.004 [0.003, 0.005] for the proposed self-supervised at $R = 4$, supervised at $R = 4$, supervised at $R = 4$ with $\rho = 0.4$, and supervised at $R = 8$, respectively.

**FIGURE S11** Reconstruction results for the coronal PD-weighted data set at acceleration rates of 4, 6, and 8. For $R = 4$ and 6, the proposed self-supervised approach performs similarly with the supervised approach. However, at $R = 8$, the image quality degrades for both methods with more pronounced blurring, whereas the self-supervised approach further suffers from visible residual aliasing artifacts.

**FIGURE S12** Reconstruction results for the proposed self-supervised approach when using same or varying sets, $\Theta$ and $\Lambda$, across different training slices. The two approaches perform similarly, with the varying mask approach showing
slight improvement. The median and interquartile ranges for SSIM across the test data set were 0.959 [0.945, 0.970] and 0.960 [0.947, 0.971], and for NMSEs were 0.002 [0.001, 0.002] and 0.002 [0.001, 0.002] for varying mask and same mask scenarios, respectively.

**FIGURE S13** Reconstruction results for supervised training when using shared and distinct (nonshared) parameters across the unrolled network. The two approaches perform similarly both visually and quantitatively. The interquartile ranges of SSIM values across the test data set were 0.967 [0.955, 0.978] and 0.964 [0.953, 0.975], and NMSE values were 0.001 [0.001, 0.002] and 0.001 [0.001, 0.002], for the shared and nonshared scenarios, respectively. Note that the same training database was used for the two approaches. The nonshared approach has 10 times as many trainable parameters, and its generalization performance may benefit from a larger training database. This was not studied, as it was not the focus of our study.

**TABLE S1** Imaging parameters for the knee data sets

**TABLE S2** Median and interquartile ranges (25th-75th percentile) of the quantitative evaluation of SSIM and NMSE values for different overlap scenarios between Λ and Θ when ρ = 0.4. Overlap percentage, defined as |Λ∩Θ|/|Λ|, refers to the amount of data in the loss mask Λ that was also included in the training mask Θ. Performance of the self-supervised training degrades as the amount of overlap increases.

**How to cite this article:** Yaman B, Hosseini SAH, Moeller S, Ellermann J, Uğurbil K, Akçakaya M. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magn Reson Med*. 2020;84:3172–3191. [https://doi.org/10.1002/mrm.28378](https://doi.org/10.1002/mrm.28378)