RESEARCH ARTICLE

NMR IN BIOMEDICINE WILEY

Multi-mask self-supervised learning for physics-guided neural networks in highly accelerated magnetic resonance imaging

Burhaneddin Yaman[1,2] | Hongyi Gu[1,2] | Seyed Amir Hossein Hosseini[1,2] |
Omer Burak Demirel[1,2] | Steen Moeller[2] | Jutta Ellermann[2] | Kâmil Uğurbil[2] |
Mehmet Akçakaya[1,2]

[1]Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, Minnesota, USA

[2]Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, Minnesota, USA

Correspondence
Mehmet Akçakaya, University of Minnesota, 200 Union Street South East, Minneapolis, MN 55455, USA.
Email: akcakaya@umn.edu

Self-supervised learning has shown great promise because of its ability to train deep learning (DL) magnetic resonance imaging (MRI) reconstruction methods without fully sampled data. Current self-supervised learning methods for physics-guided reconstruction networks split acquired undersampled data into two disjoint sets, where one is used for data consistency (DC) in the unrolled network, while the other is used to define the training loss. In this study, we propose an improved self-supervised learning strategy that more efficiently uses the acquired data to train a physics-guided reconstruction network without a database of fully sampled data. The proposed multi-mask self-supervised learning via data undersampling (SSDU) applies a holdout masking operation on the acquired measurements to split them into multiple pairs of disjoint sets for each training sample, while using one of these pairs for DC units and the other for defining loss, thereby more efficiently using the undersampled data. Multi-mask SSDU is applied on fully sampled 3D knee and prospectively undersampled 3D brain MRI datasets, for various acceleration rates and patterns, and compared with the parallel imaging method, CG-SENSE, and single-mask SSDU DL-MRI, as well as supervised DL-MRI when fully sampled data are available. The results on knee MRI show that the proposed multi-mask SSDU outperforms SSDU and performs as well as supervised DL-MRI. A clinical reader study further ranks the multi-mask SSDU higher than supervised DL-MRI in terms of signal-to-noise ratio and aliasing artifacts. Results on brain MRI show that multi-mask SSDU achieves better reconstruction quality compared with SSDU. The reader study demonstrates that multi-mask SSDU at R = 8 significantly improves reconstruction compared with single-mask SSDU at R = 8, as well as CG-SENSE at R = 2.

KEYWORDS
convolutional neural networks, data augmentation, deep learning, image reconstruction, parallel imaging, self-supervised learning

---

1 | INTRODUCTION

Data acquisition is lengthy in many magnetic resonance imaging (MRI) examinations, creating challenges for improving resolution and coverage, hence making accelerated MRI reconstruction an ongoing research topic. Parallel imaging[1–3] and compressed sensing[4–9] are two commonly used acceleration methods, with the former being the clinical gold standard for fast MRI, and the latter providing additional acceleration in a number of scenarios. However, acceleration rates remain limited as reconstructed images may suffer from noise amplification[10] or residual artifacts[11,12] in parallel imaging and compressed sensing, respectively. Recently, deep learning (DL) has emerged as an alternative to accelerated MRI because of its improved reconstruction quality compared with conventional approaches, especially at higher acceleration rates.[13–18]

Among DL methods, physics-guided DL approaches, which incorporate the MRI-encoding operator to solve a regularized inverse problem, have attracted interest because of their robustness.[18,19] Physics-guided DL approaches unroll an iterative process that alternates between data consistency (DC) and regularization for a certain number of iterations. They are trained end-to-end, typically in a supervised manner, by minimizing the difference between network output and a ground truth reference obtained from fully sampled data.[14,20–22] However, the acquisition of fully sampled data, especially on large patient populations, is either challenging or impossible in many practical scenarios.[23–27]

As supervised training becomes inoperative in the absence of fully sampled data, several methods have been proposed to train networks without fully sampled data.[28–33] Among these approaches, self-supervision via data undersampling (SSDU) trains physics-guided neural networks by utilizing only the acquired subsampled measurements.[31] In SSDU, the available measurements are split into two disjoint sets by a masking operation, which reduces the sensitivity to overfitting and is central for reliable performance. One of these sets is used in the DC units of the network, and the other is used to define the loss function in k-space. For moderately high acceleration rates, the networks trained using SSDU match the performance of those from supervised learning. While SSDU demonstrated that the splitting of acquired points into two sets was sufficient for training a neural network for reconstruction from undersampled data, a strategy that augments the use of the subsampled data to improve the performance is essential for higher acceleration rates.

In this study, we sought to improve the performance of SSDU with multiple masks. The proposed multi-mask SSDU splits acquired measurements into multiple pairs of disjoint sets for each training slice, while using one of these sets in a pair for DC units and the other for defining loss, similar to the original SSDU. The proposed multi-mask SSDU approach is applied on fully sampled 3D knee MRI datasets from mridata.org,[34] as well as a prospectively undersampled high-resolution 3D brain MRI dataset, and is compared with parallel imaging, SSDU with a single mask,[31] and supervised DL-MRI when fully sampled data are available. The results show that the proposed multi-mask SSDU approach at high acceleration rates significantly improves upon SSDU and performs as well as supervised DL-MRI, while the reader studies indicate that the proposed multi-mask approach also outperforms the supervised DL-MRI approach in terms of signal-to-noise ratio (SNR) improvement and aliasing artifact reduction.

2 | METHODS

2.1 | Supervised training of physics-guided DL-MRI reconstruction

Let $\mathbf{y}_\Omega$ be the acquired subsampled measurements, with Ω denoting the subsampling pattern and $\mathbf{x}$ the image to be recovered. The forward model for encoding is

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

where $\mathbf{E}_\Omega : \mathbb{C}^M \to \mathbb{C}^P$ is the encoding operator including the coil sensitivities and a partial Fourier matrix sampling the locations specified by Ω, and $\mathbf{n} \in \mathbb{C}^P$ is measurement noise. For sub-Nyquist sampling at high rates, the forward model may be ill-conditioned, necessitating the use of regularization, leading to an inverse problem for image reconstruction:

$$\arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega\mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}), \tag{2}$$

where the first term represents DC and the second term, $\mathcal{R}(\cdot)$, is the regularizer. Several approaches may be used to iteratively solve the above optimization problem.[35] In this work, we use variable splitting via the quadratic penalty method,[21,31,35] which decouples DC and regularizer operations:

$$\mathbf{z}^{(i-1)} = \arg\min_{\mathbf{z}} \mu\|\mathbf{x}^{(i-1)} - \mathbf{z}\|_2^2 + \mathcal{R}(\mathbf{z}), \tag{3}$$

$$\mathbf{x}^{(i)} = \arg\min_{\mathbf{x}} \|\mathbf{y}_\Omega - \mathbf{E}_\Omega\mathbf{x}\|_2^2 + \mu\|\mathbf{x} - \mathbf{z}^{(i-1)}\|_2^2, \tag{4}$$

where μ is the quadratic penalty parameter, $\mathbf{x}^{(i)}$ is the network output at iteration i, $\mathbf{z}^{(i)}$ is an intermediate variable, and $\mathbf{x}^{(0)}$ is the initial image obtained from zero-filled undersampled k-space data. In physics-guided DL, this iterative optimization is unrolled for a fixed number of iterations. Equation (3) corresponds to a regularizer, which is implicitly solved by a neural network, whereas Equation (4) has a closed form solution[31] that can be solved by gradient descent methods such as conjugate gradient.[20]

In traditional DL-MRI approaches, training datasets contain pairs of undersampled k-space/image and fully sampled k-space/ground truth image.[14,20,21,36] Let $\mathbf{y}_{ref}^i$ be the fully sampled reference k-space data for subject i, and $f\left(\mathbf{y}_{\Omega}^i, \mathbf{E}_{\Omega}^i; \boldsymbol{\theta}\right)$ denote the output of the unrolled network that is parametrized by $\boldsymbol{\theta}$ for subsampled k-space data $\mathbf{y}_{\Omega}^i$ and the corresponding encoding matrix $\mathbf{E}_{\Omega}^i$ of the same subject i. The supervised PG-DL training is performed by defining the loss function in image domain or k-space.[31] Training can be performed by minimizing a k-space loss function as

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\mathbf{y}_{ref}^i, \mathbf{E}_{full}^i\left(f\left(\mathbf{y}_{\Omega}^i, \mathbf{E}_{\Omega}^i; \boldsymbol{\theta}\right)\right)\right), \tag{5}$$

where N is the number of fully sampled training data in the database, $\mathbf{E}_{full}^i$ is the fully sampled encoding operator that transforms network output to k-space, and $\mathcal{L}(.,.)$ is the loss between the fully sampled and reconstructed k-spaces. The sampling locations, Ω, may vary per subject in a more general setup (i.e., indexed by i). However, this was not included for simplicity of notation.

2.2 | Self-supervision via data undersampling

To enable training without fully sampled datasets, SSDU has been proposed,[31] where the acquired subsampled data indices, Ω, from each scan are divided into two disjoint sets, Θ and Λ. Θ is used in DC units in the unrolled network and Λ is used to define the loss function, and the following self-supervised loss function is minimized:

$$\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(\mathbf{y}_{\Lambda}^i, \mathbf{E}_{\Lambda}^i\left(f\left(\mathbf{y}_{\Theta}^i, \mathbf{E}_{\Theta}^i; \boldsymbol{\theta}\right)\right)\right). \tag{6}$$

Unlike the supervised approach, only a subset of measurements, Θ, are used as the input to the unrolled network. The network output is transformed to k-space, where the loss is calculated only at unseen k-space indices, Λ. After training is completed, testing is performed on unseen data using all available measurements, Ω.

2.3 | Proposed multi-mask SSDU

SSDU reconstruction quality degrades at very high acceleration rates due to higher data scarcity, arising from the splitting into Θ and Λ. To tackle this issue, we propose a multi-mask SSDU approach, which retrospectively splits acquired indices Ω into disjoint sets Θ and Λ multiple times, as shown in Figure 1. Formally, we split available measurements multiple times for each subject i such that for each partition $\Omega = \Theta_j \cup \Lambda_j$, for $j = 1, ..., K$, denoting the number of partitions for each scan. Similar to SSDU, each pair of sets in each scan was disjoint (i.e., $\Lambda_j = \Omega \setminus \Theta_j$ for $j \in 1, ..., K$). In other words, Ω is retrieved by the union of each pair of Λ_j and Θ_j for any $j \in 1, ..., K$. Hence, the loss function to minimize during training becomes

$$\min_{\boldsymbol{\theta}} \frac{1}{N \cdot K} \sum_{i=1}^{N} \sum_{j=1}^{K} \mathcal{L}\left(\mathbf{y}_{\Lambda_j}^i, \mathbf{E}_{\Lambda_j}^i\left(f\left(\mathbf{y}_{\Theta_j}^i, \mathbf{E}_{\Theta_j}^i; \boldsymbol{\theta}\right)\right)\right). \tag{7}$$

The proposed multi-mask approach enables efficient use of available data by ensuring a higher fraction of low and high frequency components are utilized in the training and loss masks. Such utilization was inherently limited in the original SSDU approach, because each acquired k-space point was either used in training or loss masks only once.

2.4 | 3D imaging datasets

Fully sampled 3D knee datasets were obtained from mridata.org,[34] and were acquired with approval from the local institutional review board on a 3-T GE Discovery MR 750 system with an eight-channel knee coil array using a fast spin-echo sequence. The relevant imaging parameters were: field of view (FOV) = $160 \times 160 \times 154\ mm^3$, resolution = $0.5 \times 0.5 \times 0.6\ mm^3$, and matrix size = $320 \times 320 \times 256$.
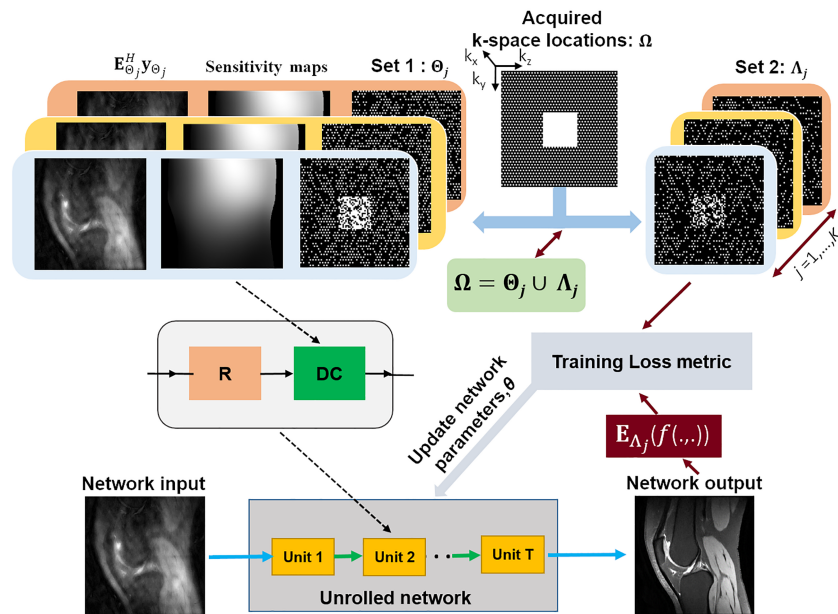
FIGURE 1 The multi-mask self-supervised learning scheme to train physics-guided deep learning without fully sampled data. The acquired subsampled k-space measurements for each scan, Ω, are split into multiple disjoint sets, Θ_j and Λ_j, in which $\Omega = \Theta_j \cup \Lambda_j$, for $j \in 1, ..., K$. The first set of indices, Θ_j, is used in the data consistency (DC) unit of the unrolled network, while the latter set, Λ_j, is used to define the loss function for training. During training, the output of the network is transformed to k-space, and the available subset of measurements at Λ_j are compared with the corresponding reconstructed k-space values. Based on this training loss, the network parameters are subsequently updated

Brain imaging was performed on healthy subjects using a standard Siemens 3D magnetization-prepared rapid gradient echo imaging (MPRAGE) sequence on a 3-T Siemens Magnetom Prisma (Siemens Healthcare, Erlangen, Germany) system using a 32-channel receiver head coil-array.[31] 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 HIPAA-compliant study. The relevant imaging parameters were: FOV = $224 \times 224 \times 157$ mm^3, resolution = $0.7 \times 0.7 \times 0.7$ mm^3, matrix size = $320 \times 320 \times 224$, prospective acceleration R = 2 (uniform in k_y) and autocalibration signal (ACS) lines = 32.[31]

The 3D k-space datasets were inverse Fourier-transformed along the read-out direction, and these slices were processed individually. The knee and brain datasets were retrospectively undersampled to R = 8 using a uniform sheared 2D undersampling pattern.[37] Additionally, for the knee datasets, where a fully sampled reference was available, further undersampling was performed at R = 8 using uniform 1D and 2D (k_y-k_z) random, and 1D and 2D and Poisson undersampling masks. The undersampling masks are provided in Figure S1. Finally, knee datasets were also undersampled to R = 12 using 2D random and Poisson undersampling masks. A 24×24 and 32×32 ACS region in the k_y-k_z plane was kept fully sampled for the knee and brain datasets, respectively.[31] The training sets for both knee and brain datasets consisted of 300 slices from 10 subjects, formed by taking 30 slices from each subject. For knee MRI, two different subjects with 200 slices were used for validation in multi-mask hyperparameter tuning, and eight other different subjects were used for the testing of the final method. For the brain dataset, the testing was performed on nine different subjects. The proposed multi-mask SSDU approach was compared with: the parallel imaging method, CG-SENSE; the blind compressed sensing method (Blind-CS)[38]; the zero-shot learning approach, deep image prior (DIP)[39,40]; the state-of-the-art self-supervised learning approach (SSDU)[31]; and the state-of-the-art supervised DL method when applicable.[20] For Blind-CS and DIP, we used publicly available implementations from github.com/wenbihan/FRIST_ivp2017 and github.com/MLI-lab/ConvDecoder, respectively. We tuned the number of epochs for DIP for maximal performance.

2.5 | Choice of multi-mask hyperparameters

There are several tunable hyperparameters in multi-mask SSDU, including the number of partitions, K in Equation (7), as well as the distribution and size of Λ as in SSDU. A variable-density Gaussian distribution was used for Λ in[31] for a single mask. In this study, we used a uniformly random distribution for the proposed approach, as the benefits of a variable density distribution diminish with multiple masks (Figure S2). In,[31] the size of Λ was optimized to $\rho = |\Lambda|/|\Omega| = 0.4$, which is also the optimal choice for the distribution considered here (Figure S3). After these two

hyperparameters were set, the number of partitions of each scan, K, was varied among 3, 5, 6, 7, 8, and 10 to optimize the remaining distinct hyperparameter of the multi-mask SSDU.

2.6 | Network and training details

The iterative optimization problem in Equations (3) and (4) was unrolled for $T = 10$ iterations. Conjugate gradient descent was used in DC units of the unrolled network.[20,31] The proximal operator corresponding to the solution of Equation (3) employs the ResNet structure used in SSDU.[31] It comprises input and output convolution layers and 15 residual blocks (RBs) each containing two convolutional layers, where the first layer is followed by a rectified linear unit (ReLU) and the second layer is followed by a constant multiplication layer. All layers had a kernel size of 3×3, 64 channels. The unrolled network, which shares parameters across the unrolled iterations, had a total of 592,129 trainable parameters. As in SSDU, a ResNet structure was used for the regularizer in Equation (3), where the network parameters were shared across the unrolled network.[31] Coil sensitivity maps were generated from 24×24 center of k-space using ESPIRiT.[41]

As a preprocessing step, the maximum absolute value of the k-space for each slice in the datasets was normalized to 1 in all cases. The real and imaginary parts of the complex MRI dataset were concatenated as two channels prior to inputting into the network. Separate networks for knee and brain datasets were trained using the Adam optimizer with a learning rate of 5×10^{-4}, by minimizing a normalized $\ell_1 - \ell_2$ loss function defined as $\mathcal{L}(\boldsymbol{u}, \boldsymbol{v}) = \frac{\|\boldsymbol{u} - \boldsymbol{v}\|_2}{\|\boldsymbol{u}\|_2} + \frac{\|\boldsymbol{u} - \boldsymbol{v}\|_1}{\|\boldsymbol{u}\|_1}$,[31] with a batch size of 1 over 100 epochs. All training was performed using Tensorflow in Python, and was processed on a workstation with an NVIDIA Tesla V100 GPU with 32 GB of memory. The implementation of this method will be provided online (https://github.com/byaman14/SSDU).

2.7 | Image evaluation

Quantitative assessment of the experimental results was performed using normalized mean square error (NMSE) and structural similarity index (SSIM) when fully sampled data were available as reference. Moreover, qualitative assessment of the image quality from different reconstruction methods was performed by an experienced radiologist. For knee MRI, the proposed multi-mask SSDU was compared with ground truth obtained from fully sampled data, SSDU, and the parallel imaging method CG-SENSE, all at R = 8 with 2D uniform undersampling. For brain MRI, the proposed multi-mask SSDU was compared with SSDU at R = 8. Additionally, the CG-SENSE approach at the acquisition acceleration R = 2 was evaluated to serve as the clinical baseline. The reader, with 15 years of experience of musculoskeletal and neuro imaging, was blinded to the reconstruction method, which were shown in a randomized order to avoid bias, except for the knowledge of the reference image in the knee MRI dataset. Evaluations were based on a four-point ordinal scale, adopted from[14] 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 less than 0.05.

3 | RESULTS

3.1 | Number of partitions for multi-mask SSDU

Figure 2 shows the effect of the proposed multi-mask self-supervised network training at R = 8 using 2D uniform sheared undersampling with different numbers of masks ($K = 3, 5, 6, 7, 8,$ and 10), as well as the ground truth reference and the zero-filled undersampled data. The multi-mask SSDU approach suppresses residual artifacts as K increases from 3 to 6. At $K = 7$, the visible residual artifacts are removed completely. When K is further increased to 8 and 10, residual artifacts reappear. The quantitative assessment of the validation dataset further confirms this qualitative assessment. The median and interquartile range of SSIM values on the validation set were 0.8256 [0.7980, 0.8507], 0.8260 [0.8002, 0.8516], 0.8264 [0.8016, 0.8527], 0.8267 [0.8027, 0.8537], 0.8263 [0.8007, 0.8519], and 0.8257 [0.7989, 0.8511], and the NMSE values were 0.0138 [0.0121, 0.0158], 0.0135 [0.0119, 0.0158], 0.0135 [0.0119, 0.0157], 0.0134 [0.0118, 0.0156], 0.0135 [0.0119, 0.0158], and 0.0137 [0.0121, 0.0159] using Gaussian selection for $K \in 3, 5, 6, 7, 8,$ and 10, respectively. Hence, $K = 7$ was used for the remainder of the study.

3.2 | 3D imaging datasets

Figure 3 depicts representative reference and reconstruction results of the 3D knee dataset using CG-SENSE, Blind-CS, DIP, supervised DL-MRI, SSDU, and the proposed multi-mask SSDU, as well as the difference images of these methods with respect to the reference for the 2D uniform
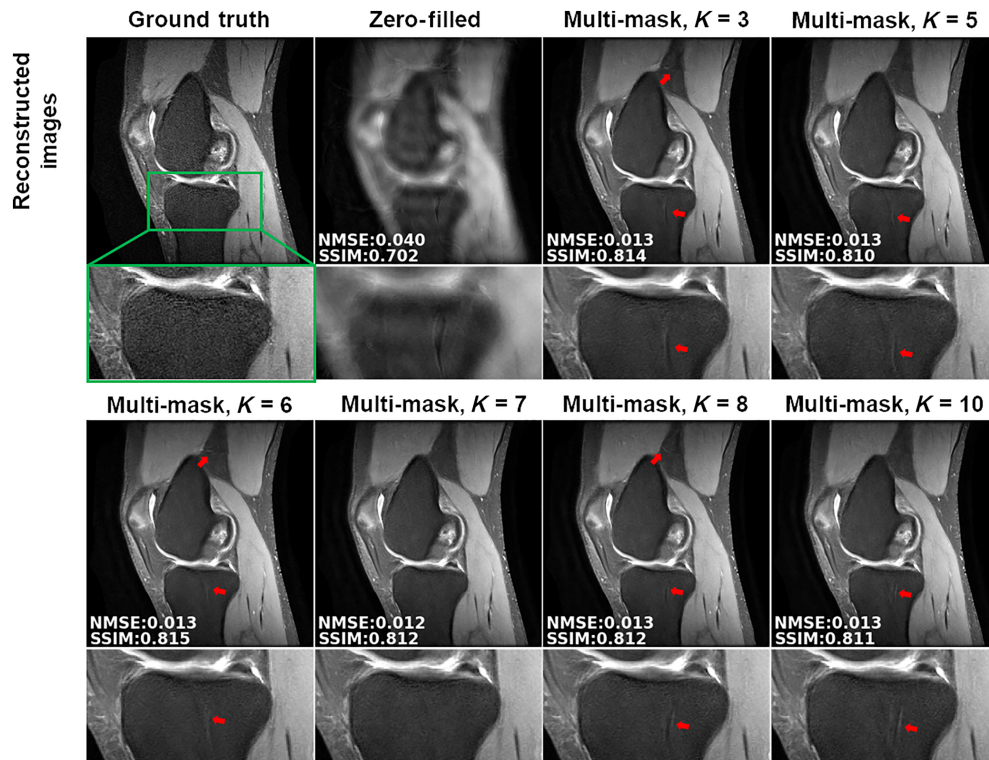
FIGURE 2 A representative test slice showing the reconstruction results for different numbers of partitions K using a 2D uniform sheared $R = 8$ undersampling mask. Red arrows mark residual artifacts for $K \leq 6$ and $K \geq 8$. These artifacts are suppressed at $K = 7$, which is used for the remainder of the study. NMSE, normalized mean square error; SSIM, structural similarity index

sheared $R = 8$ undersampling mask. Red arrows indicate that CG-SENSE suffers from highly visible residual artifacts and noise amplification. Blind-CS and DIP severely suffer from the blurring artifacts. SSDU alleviates these artifacts substantially, although residual artifacts remain. The proposed multi-mask SSDU further removes these artifacts for both slices shown, while achieving similar reconstruction quality to supervised DL-MRI for the first slice, and further reducing the residual aliasing artifacts visible in the supervised DL-MRI approach for the second slice. Quantitative metrics and difference images in the figure further confirm that multi-mask SSDU outperforms SSDU, while performing similarly to supervised DL-MRI. Additional experimental results on the 3D knee dataset using different undersampling patterns at $R = 8$ and $R = 12$ are provided in Figures S4, S5, and S6. In all of these experiments, multi-mask SSDU visibly and quantitatively outperformed SSDU, further reducing the gap between self-supervised learning and supervised DL-MRI.

Table 1 summarizes the median and interquartile ranges for NMSE and SSIM value metrics for different undersampling masks and acceleration rates across the whole knee MRI test datasets. In all cases, CG-SENSE reconstruction is significantly outperformed by all the DL approaches. Among the DL approaches, supervised DL-MRI outperforms self-supervised learning methods, while multi-mask SSDU quantitatively improves upon SSDU.

Figure 4 demonstrates CG-SENSE reconstruction of a slice of the 3D-MPRAGE dataset at prospective acceleration $R = 2$, as well as CG-SENSE, Blind-CS, DIP, SSDU, and the proposed multi-mask SSDU approach at retrospective acceleration $R = 8$ using the 2D uniform sheared undersampling mask. Blind-CS and DIP reconstructions significantly suffer from blurring artifacts. SSDU at high acceleration $R = 8$ achieves a similar reconstruction quality to CG-SENSE at acquisition acceleration $R = 2$. Multi-mask SSDU further improves the reconstruction quality by suppressing the noise evident in SSDU and CG-SENSE.

3.3 | Image evaluation scores

Figure 5A,B summarize the reader study results for knee and brain datasets using the 2D uniform sheared $R = 8$ undersampling mask, respectively. For knee MRI, the proposed multi-mask SSDU was rated highest in terms of SNR, with a statistically significant improvement over all other methods, except for supervised DL-MRI. For blurring, ground truth data was rated better than all the other methods. In terms of aliasing artifacts and overall image quality, the proposed multi-mask SSDU approach was rated best compared with the other methods and ground truth. In terms of these two evaluation criteria, all DL-MRI approaches and the reference showed similar statistical behavior, except SSDU was statistically worse
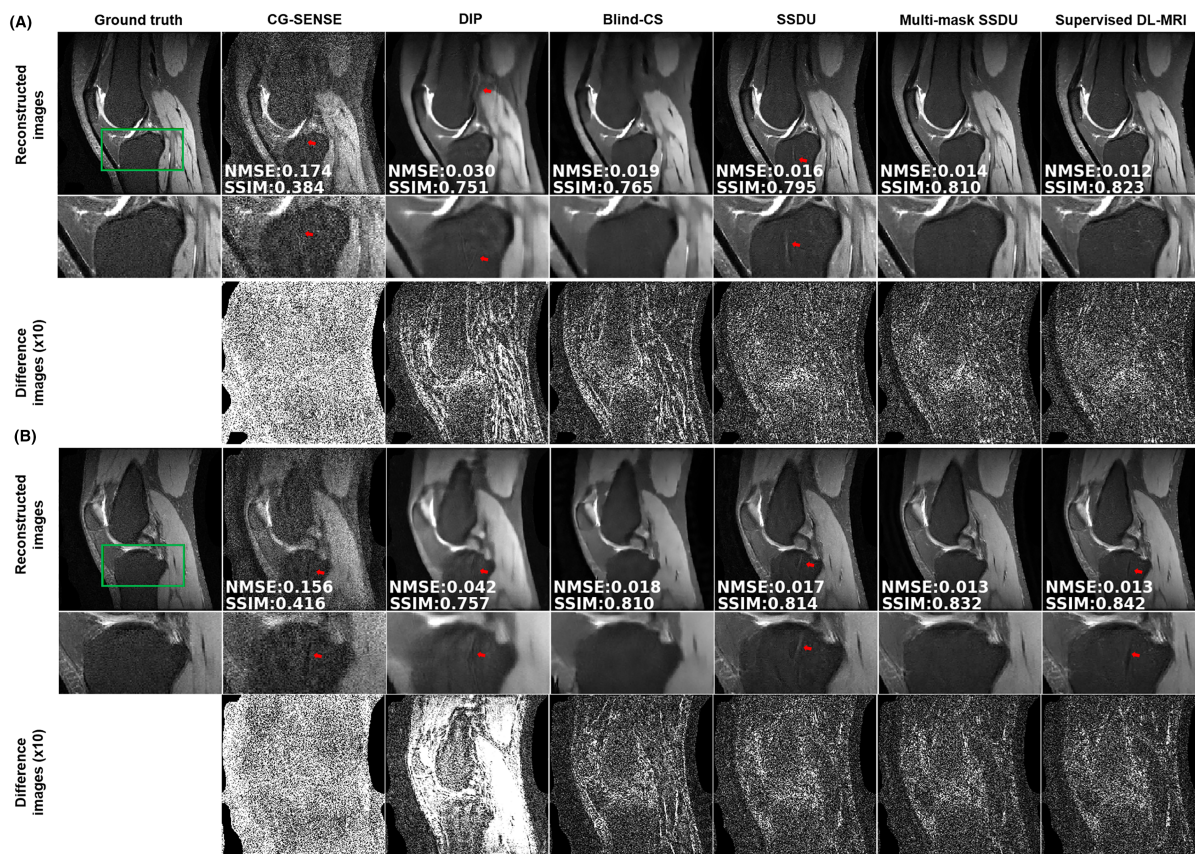
FIGURE 3 (A, B) Representative test slices from the 3D fast spin-echo knee magnetic resonance imaging (MRI) dataset showing the reconstruction results for the proposed multi-mask self-supervised deep learning (DL)-MRI (multi-mask SSDU), self-supervised DL-MRI (SSDU), supervised DL-MRI, DIP, Blind-CS, and CG-SENSE approaches for retrospective 2D uniform undersampling R = 8, as well as the error images with respect to the fully sampled reference. CG-SENSE suffers from substantial residual artifacts that are shown with red arrows for both slices. Blind-CS and DIP suffer severely from blurring artifacts. DL-MRI with SSDU learning suppresses a large portion of these artifacts, but still exhibits visible residual artifacts in both scenarios. The proposed multi-mask SSDU successfully suppresses these artifacts further for both slices; in (A) it closely matches the performance of supervised DL-MRI, and in (B) it reduces residual aliasing further compared with supervised DL-MRI. Blind-CS, the blind compressed sensing method; CG-SENSE, the parallel imaging method; DIP, deep image prior; NMSE, normalized mean square error; SSDU, self-supervision via data undersampling; SSIM, structural similarity index

than the proposed multi-mask SSDU and supervised approach in terms of aliasing artifacts. A more comprehensive comparison also containing reader scores for CG-SENSE is presented in Figure S7.

For brain MRI, DL-MRI reconstructions trained using the proposed multi-mask SSDU and SSDU approach at an acceleration rate of 8 performed similar to CG-SENSE at acquisition R = 2 in terms of SNR and blurring. However, in terms of aliasing artifacts, the proposed multi-mask SSDU significantly outperformed its counterparts. In terms of overall image quality, both SSDU methods at R = 8 showed a statistically significant improvement over CG-SENSE at R = 2, while the proposed multi-mask SSDU achieved the best performance.

4 | DISCUSSION

In this work, we extended our earlier work on self-supervision via data undersampling, which trains a physics-guided neural network without fully sampled data, to a multi-mask setting where multiple pairs of disjoint sets were used for each training slice in the dataset. The training of physics-guided DL-MRI reconstruction without ground truth data remains an important topic, because the acquisition of fully sampled data is either impossible or challenging in a number of scenarios.[23–27] Among multiple methods proposed for this goal,[28–30,42,43] self-supervision directly uses the acquired data without relying on generative models or intermediate estimates. Our work makes several contributions to these approaches. The main contribution of the proposed multi-mask self-supervised learning approach is to use the available undersampled data more efficiently to enable physics-guided DL training, by retrospectively splitting these data into multiple 2-tuple of sets for the DC units during training and for defining loss. Another resulting contribution of the proposed multi-mask SSDU is an alternative data augmentation strategy for DL-MRI

TABLE 1 The median and interquartile ranges for NMSE and SSIM metrics for different undersampling masks and acceleration rates. Note that due to the different size of the ACS data, 1D masks correspond to an effective acceleration rate of 5.2, while the 2D masks yield an effective acceleration rate of 7.7

		CG-SENSE	DIP	Blind-CS	Supervised DL-MRI	SSDU	Multi-mask SSDU
Uniform 1D R = 8	NMSE	0.0994 [0.0949, 0.1419]	0.0457 [0.0316, 0.0627]	0.0192 [0.0172, 0.2274]	0.0122 [0.0108, 0.0139]	0.0166 [0.0147, 0.0191]	0.0140 [0.0124, 0.0161]
	SSIM	0.4698 [0.4193, 0.5353]	0.7224 [0.6819, 0.7435]	0.7945 [0.7648, 0.8332]	0.8505 [0.8306, 0.8751]	0.8211 [0.7947, 0.8527]	0.8315 [0.8084, 0.8600]
Uniform 2D R = 8	NMSE	0.1475 [0.1291, 0.1779]	0.0438 [0.0297, 0.0622]	0.0187 [0.0168, 0.0218]	0.0124 [0.0112, 0.0143]	0.0164 [0.0148, 0.0189]	0.0135 [0.0123, 0.0155]
	SSIM	0.4411 [0.3797, 0.4976]	0.7124 [0.6704, 0.7452]	0.7884 [0.7526, 0.8126]	0.8421 [0.8201, 0.8662]	0.8150 [0.7877, 0.8426]	0.8298 [0.8067, 0.8560]
Random 1D R = 8	NMSE	0.0994 [0.0805, 0.1236]	0.0392 [0.0216, 0.0524]	0.0179 [0.0162, 0.0205]	0.0121 [0.0107, 0.0139]	0.0156 [0.0135, 0.0177]	0.0137 [0.0121, 0.0155]
	SSIM	0.4886 [0.4305, 0.5571]	0.7326 [0.7005, 0.7715]	0.8032 [0.7789, 0.8407]	0.8524 [0.8314, 0.8756]	0.8328 [0.8089, 0.8615]	0.8367 [0.8144, 0.8633]
Random 2D R = 8	NMSE	0.1473 [0.1301, 0.1759]	0.0412 [0.0256, 0.0567]	0.0197 [0.0178, 0.0224]	0.0130 [0.0117, 0.0149]	0.0173 [0.0155, 0.0199]	0.0145 [0.0131, 0.0165]
	SSIM	0.4239 [0.3631, 0.4766]	0.7286 [0.6957, 0.7592]	0.7914 [0.7648, 0.8229]	0.8379 [0.8164, 0.8637]	0.8123 [0.7853, 0.8417]	0.8224 [0.8002, 0.8509]
Poisson R = 8	NMSE	0.1035 [0.0937, 0.1206]	0.0358 [0.0182, 0.0497]	0.0158 [0.0134, 0.0178]	0.0101 [0.0091, 0.0112]	0.0131 [0.0118, 0.0149]	0.0108 [0.0098, 0.0121]
	SSIM	0.4885 [0.4394, 0.5397]	0.7325 [0.6914, 0.7629]	0.8075 [0.7800, 0.8370]	0.8554 [0.8365, 0.8793]	0.8312 [0.8066, 0.8585]	0.8421 [0.8212, 0.8679]
Random 2D R = 12	NMSE	0.1331 [0.1207, 0.1556]	0.0542 [0.0386, 0.0785]	0.0247 [0.0218, 0.0287]	0.0157 [0.0141, 0.0179]	0.0221 [0.0198, 0.0254]	0.0185 [0.0167, 0.0208]
	SSIM	0.4325 [0.3756, 0.4796]	0.6834 [0.6576, 0.7182]	0.7614 [0.7264, 0.7916]	0.8148 [0.7916, 0.8431]	0.7809 [0.7517, 0.8151]	0.7982 [0.7722, 0.8288]
Poisson R = 12	NMSE	0.0876 [0.0795, 0.1012]	0.0491 [0.0342, 0.0673]	0.0173 [0.0151, 0.0192]	0.0119 [0.0107, 0.0133]	0.0151 [0.0136, 0.0169]	0.0129 [0.0117, 0.0145]
	SSIM	0.5119 [0.4638, 0.5597]	0.7014 [0.6632, 0.7376]	0.7753 [0.7385, 0.8164]	0.8362 [0.8156, 0.8625]	0.8133 [0.7862, 0.8442]	0.8326 [0.8098, 0.8609]

Abbreviations: ACS, autocalibration signal; Blind-CS, the blind compressed sensing method; CG-SENSE, the parallel imaging method; DIP, deep image prior; DL-MRI, deep learning magnetic resonance imaging; NMSE, normalized mean square error; SSDU, self-supervision via data undersampling; SSIM, structural similarity index.
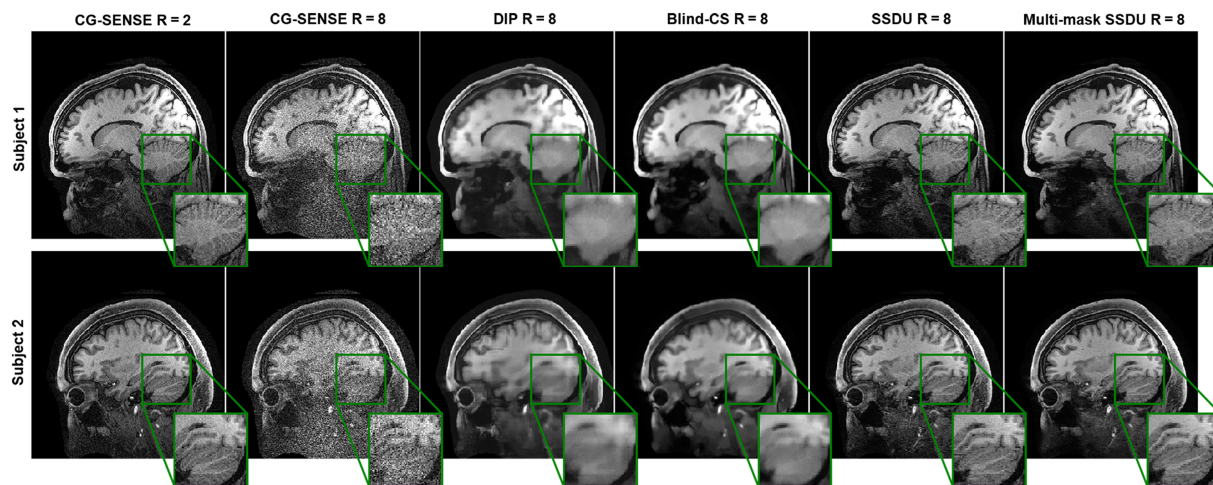
FIGURE 4 Reconstruction results from prospectively two-fold equispaced undersampled brain magnetic resonance imaging (MRI). SSDU, multi-mask SSDU, DIP, Blind-CS, and CG-SENSE are applied at further retrospective acceleration rates of 8 with equispaced sheared k_y-k_z undersampling patterns, while CG-SENSE is also used at the acquisition rate of 2, which serves as the clinical baseline. CG-SENSE suffers from visibly higher noise amplification at R = 8. Blind-CS and DIP reconstructions significantly suffer from blurring artifacts. SSDU DL-MRI performs successful reconstruction at R = 8, while achieving similar image quality to CG-SENSE at R = 2. The proposed multi-mask SSDU DL-MRI further enhances the SSDU DL-MRI performance by achieving a lower noise level in reconstruction results. Blind-CS, the blind compressed sensing method; CG-SENSE, the parallel imaging method; DIP, deep image prior; DL, deep learning; SSDU, self-supervision via data undersampling
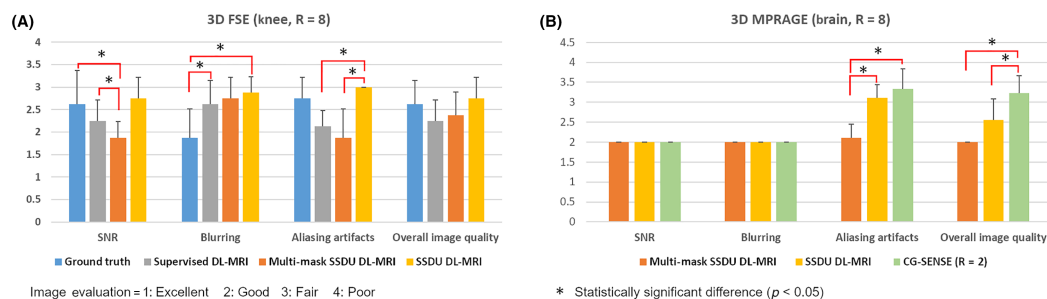

FIGURE 5 (A) Reader study for knee magnetic resonance imaging (MRI) for reconstructions using uniform 2D undersampling masks at R = 8. Bar-plots show average reader scores and their standard deviation across the test subjects. Statistical testing was performed by one-sided Wilcoxon single-rank test, with * showing a significant statistical difference with $p < 0.05$. In terms of SNR, the proposed multi-mask SSDU was rated highest, and statistically better than all approaches, except for supervised DL-MRI. For blurring, ground truth data was rated statistically better than all the other methods, except for the proposed multi-mask SSDU. In terms of aliasing artifacts and overall image quality, the proposed multi-mask SSDU approach was rated best compared with the other methods and ground truth. In terms of these two evaluation criteria, all DL-MRI approaches and the reference showed similar statistical behavior, except SSDU was statistically worse than the proposed multi-mask SSDU and supervised approach in terms of aliasing artifacts. (B) Reader study for brain MRI. CG-SENSE at R = 2, and the proposed multi-mask SSDU and SSDU at R = 8 using a uniform 2D undersampling mask, were in good agreement in terms of SNR and blurring. In terms of aliasing artifacts and overall image quality, the proposed multi-mask SSDU approach received the best scores, while CG-SENSE at R = 2 was rated lowest and showed a significant statistical difference with the proposed multi-mask SSDU in terms of both evaluation criteria and SSDU in terms of overall image quality. The proposed multi-mask SSDU was also rated statistically better than SSDU in terms of aliasing artifacts. CG-SENSE, the parallel imaging method; DL, deep learning; FSE, fast spin-echo; MPRAGE, magnetization-prepared rapid gradient echo imaging; SNR, signal-to-noise ratio; SSDU, self-supervision via data undersampling

reconstruction, via the retrospective holdout masking of the acquired measurements multiple times, with potential applications even beyond self-supervised learning.[44] Finally, we applied the proposed multi-mask approach on knee and brain MRI datasets using different undersampling and acceleration rates, showing its improved reconstruction performance compared with a single-mask SSDU approach. Specifically, the extensive experimental results using different subsampling patterns on the retrospectively subsampled 3D knee dataset at R = 8 and R = 12 show that the proposed multi-mask SSDU consistently outperforms SSDU, while performing similarly to the supervised DL-MRI approach. Similarly, on

prospectively subsampled brain MRI, multi-mask SSDU at R = 8 enhances the reconstruction quality of SSDU, while achieving lower noise level compared with SSDU at R = 8 and CG-SENSE at the acquisition R = 2.

As mentioned earlier, the proposed multi-mask SSDU approach can be interpreted as an alternative technique for data augmentation in DL-MRI reconstruction. With the proposed multi-mask data augmentation, self-supervised training was rated higher than supervised training in the reader study for knee imaging by a musculoskeletal expert reader in terms of noise and aliasing artifacts. Furthermore, Figure 3 shows example slices, where multi-mask self-supervised learning displays a better performance in handling artifacts compared with supervised DL-MRI, despite having lower SSIM and NMSE values. While these observations may seem surprising at first, they are consistent with recent studies that show quantitative metrics may not always align with the reconstruction performance.[45] Additionally, there are recent studies that show self-supervised DL approaches outperforming their supervised counterparts in various applications.[46,47] These and other studies suggest that supervised learning may preclude discovery, hence it may not generalize well on unseen data or may not be as robust as self-supervised learning techniques.[48] Another interesting finding from the reader study on knee data was the worse scores given to the fully sampled ground truth compared with DL-MRI methods. The expert reader noted the low SNR of the fully sampled acquisition, due to the high acquisition resolution compared with conventional clinical scans, which was substantially improved visually using the inherent noise reduction of DL-MRI reconstruction.

While the proposed multi-mask SSDU approach enhances the SSDU performance, it also has a longer training time by a factor of K compared with SSDU due to the increased size of the training dataset. Due to these lengthy training times, holdout cross-validation was used for the hyperparameter selection substudy for optimizing K instead of n-fold cross-validation. Furthermore, while the proposed multi-mask approach enables data augmentation, helping to overcome data scarcity and enhance reconstruction quality, it also bears the risk of overfitting. In a broader context, it is understood that data augmentation can lead to massive datasets, but when this idea is applied to augment initially limited datasets, it may result in overfitting.[49] This phenomenon was also observed in our study as the reconstruction quality does not monotonically improve with increasing K, and residual artifacts reappear for K of 8 or higher. The problem of choosing the optimal size of the postaugmented dataset, which corresponds to K in our setup, remains an open problem in the broader machine learning community and this challenge has been highlighted in a recent survey on data augmentation as "There is no consensus as to which ratio of original to final dataset size will result in the best performing model".[49] Hence, while we have optimized K, this was achieved using the same experimental settings of.[31] We note that the optimal value for K may differ based on the selection of hyperparameters, such as the number of epochs or the learning rate. Nonetheless, our results readily show that multi-mask SSDU improves upon the single-mask SSDU in terms of quantitative metrics for any choice of K more than 1, while also suggesting that it is not advantageous to choose a very high value of K, both from a performance perspective, and from a practical viewpoint, because of the increased training time.

In this study, we have split measurements into training and loss sets in a disjoint manner as SSDU has shown that overlaps between these sets lead to degraded reconstruction performance.[31] A uniformly random selection of masks was used in the multi-mask SSDU. This was motivated by the issue that splitting Ω based on a Gaussian random selection leads to selecting mostly low-frequency components from scarce data, especially at high acceleration rates. With a Gaussian selection of Λ, a multi-mask approach still tends to select low-frequency components for each mask. Figure S2 shows that using uniformly random selection in combination with multi-mask selection may circumvent this issue, because this will ensure that both low and high frequency are contained in the loss masks of each scan.

Although we concentrated on random selection of masks, another special type of multi-mask SSDU may be based on using a cyclic selection that ensures all acquired measurements are used for both training and loss. Figure S8 shows a comparison between the cyclic multi-mask and multi-mask SSDU approach, with multi-mask SSDU showing better reconstruction quality. Multi-mask SSDU does not impose any bounds on K, allowing $K \times |\Lambda| \geq |\Omega|$, while cyclic multi-mask SSDU strictly imposes $K \times |\Lambda| = |\Omega|$. Hence, although cyclic multi-mask SSDU ensures every point in Ω is eventually used in both DC units and for defining the k-space loss, it inherently limits the number of masks for training, which in turn hinders the amount of improvement in reconstruction quality.

The multi-masking approach proposed in this work may also be adapted to the supervised learning setting by using multiple random Θ_j in the DC units of the unrolled network, while calculating the loss on the fully sampled k-space. This extension to supervised training, which introduces an additional degree of randomness to the training process, was recently shown to improve performance over a conventional supervised DL-MRI approach.[50] However, we note that because this is a new extension arising from this work, comparisons in this study were made with the conventional DL-MRI approach that is used in the literature.[20]

The proposed multi-mask SSDU approach also shares similarities with bootstrap aggregation mechanisms. In bootstrap approaches, multiple subdatasets are generated by randomly sampling from a main dataset. The final prediction is performed by averaging outputs from each of these subdatasets to reduce the variance among trained models. However, in multi-mask SSDU, each sample in the main dataset is subsampled multiple times by retrospectively splitting its measurements into disjoint sets. As a result, an aggregated large dataset that contains the measurements of each sample in the main dataset multiple times is obtained and used for training. Unlike bootstrapping approaches, the proposed approach performs final prediction by directly using the model trained on the aggregated large dataset.

The current study has some limitations. In the proposed multi-mask SSDU, we optimized the hyperparameters ρ and K independently for two main reasons: (1) if the joint optimization led to a different ρ value, then this would create a confounding variable for the direct comparison with

the single-mask scenario; and (2) optimizing over the two hyperparameters jointly leads to a large number of trainings for marginal gain. Such a large number of trainings, which does not have a substantial implication in terms of the perspective of the study, may also come at an environmental cost, as the training of DL models has been shown to lead to a considerable amount of carbon emissions.[51] Thus, in the current study, we concentrated on the individual optimization of the K parameter for the fixed ρ value that works best for single-mask SSDU,[31] as this provides a fairer comparison. Additionally, this study focused on methodological development to improve the performance on single-mask SSDU without a specific application that may leverage large stores of existing undersampled data. With the methodology in place, such applications are being pursued in subsequent studies, both in late gadolinium enhancement (LGE) cardiac and brain MRI applications.[52-55] Moreover, the proposed multi-mask SSDU may also be synergistically combined with dynamic cardiac MRI applications, in which spatio-temporal information can be leveraged to boost the performance at high acceleration rates.[56-58]

More recently, zero-shot learning approaches have attracted interest as they enable training using only a single slice.[59-61] Thus, zero-shot learning approaches do not require any external dataset for training, unlike database DL approaches such as SSDU and the proposed multi-mask SSDU. We note that Blind-CS methods,[9,38,62] which predate DL methods, are also learning-based techniques trained on a single slice, requiring no external database. Furthermore, recent Blind-CS methods, which simultaneously combine the sparsity with low rankness, have shown great promise for solving inverse problems,[62] which may further improve upon the performance of FRIST reported in this manuscript. Blind-CS methods typically include hand-tuned parameters for weighing data consistency and regularization terms in their objective functions,[38,62] which are also learnable in some of the more contemporary zero-shot DL approaches.[60] DIP is the pioneering zero-shot DL approach, but it requires a heuristic early stopping criterion. Recent work on zero-shot learning has built on the multi-mask concept of the proposed approach to develop a rigorous stopping criterion.[60] While zero-shot learning enables training from a single slice, the reconstruction times are much longer compared with database learning approaches.

5 | CONCLUSION

The proposed multi-mask SSDU approach enables training of physics-guided neural networks without fully sampled data, while significantly outperforming single-mask SSDU at high acceleration rates through the efficient use of the acquired undersampled data with multiple masks.

ORCID

Burhaneddin Yaman https://orcid.org/0000-0003-0791-5900
Seyed Amir Hossein Hosseini https://orcid.org/0000-0002-0687-0782
Steen Moeller https://orcid.org/0000-0003-1698-7260
Mehmet Akçakaya https://orcid.org/0000-0001-6400-7736

REFERENCES

1. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med.* 2002;47(6): 1202-1210. doi:10.1002/mrm.10171
2. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med.* 1999;42(5):952-962. doi:10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S
3. Lustig M, Pauly JM. SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space. *Magn Reson Med.* 2010;64(2):457-471. doi:10.1002/mrm.22428
4. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med.* 2007;58(6):1182-1195. doi:10.1002/mrm.21391
5. 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(3):756-767. doi:10.1002/mrm.22841
6. 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(5):1090-1099. doi:10.1109/TMI.2010.2089519
7. 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(1):103-116. doi:10.1002/mrm.21757
8. Haldar JP, Hernando D, Liang ZP. Compressed-sensing MRI with random encoding. *IEEE Trans Med Imaging.* 2011;30(4):893-903. doi:10.1109/TMI.2010.2085084
9. Lingala SG, Jacob M. Blind compressive sensing dynamic MRI. *IEEE Trans Med Imaging.* 2013;32(6):1132-1145. doi:10.1109/TMI.2013.2255133
10. 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(4):895-907. doi:10.1002/mrm.21728
11. Sung K, Hargreaves BA. High-frequency subband compressed sensing MRI using quadruplet sampling. *Magn Reson Med.* 2013;70(5):1306-1318. doi:10.1002/mrm.24592
12. Sharma SD, Fong CL, Tzung BS, Law M, Nayak KS. Clinical image quality assessment of accelerated magnetic resonance neuroimaging using compressed sensing. *Invest Radiol.* 2013;48(9):638-645. doi:10.1097/RLI.0b013e31828a012d

13. Wang S, Su Z, Ying L, et al. Accelerating magnetic resonance imaging via deep learning. IEEE 13th International Symposium on Biomedical Imaging (ISBI); 2016:514-517.

14. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med.* 2018;79(6): 3055-3071. doi:10.1002/mrm.26977

15. 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(1):439-453. doi:10.1002/mrm.27420

16. Dar SUH, Özbey M, Çatlı AB, Çukur T. A transfer-learning approach for accelerated MRI using deep neural networks. *Magn Reson Med.* 2020;84(2): 663-685. doi:10.1002/mrm.28148

17. Dar SU, Yurt M, Shahdloo M, Ildız ME, Tınaz B, Çukur T. Prior-guided image reconstruction for accelerated multi-contrast MRI via generative adversarial networks. *IEEE J Sel Top Signal Process.* 2020;14(6):1072-1087. doi:10.1109/JSTSP.2020.3001737

18. Liang D, Cheng J, Ke Z, Ying L. Deep magnetic resonance image reconstruction: inverse problems meet neural networks. *IEEE Signal Process Mag.* 2020;37(1):141-151. doi:10.1109/MSP.2019.2950557

19. Monga V, Li Y, Eldar YC. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE sig Process Mag.* 2021;38(2): 18-44. doi:10.1109/MSP.2020.3016905

20. Aggarwal HK, Mani MP, Jacob M. MoDL: model-based deep learning architecture for inverse problems. *IEEE Trans Med Imaging.* 2019;38(2):394-405. doi:10.1109/TMI.2018.2865356

21. 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(1):280-290. doi:10.1109/TMI.2018.2863670

22. Yang Y, Sun J, Li H, Xu Z. Deep ADMM-Net for compressive sensing MRI. Advances in neural information processing systems; 2016:10-18.

23. Haji-Valizadeh H, Rahsepar AA, Collins JD, et al. Validation of highly accelerated real-time cardiac cine MRI with radial k-space sampling and compressed sensing in patients at 1.5T and 3T. *Magn Reson Med.* 2018;79(5):2745-2751.

24. Coelho-Filho OR, Rickers C, Kwong RY, Jerosch-Herold M. MR myocardial perfusion imaging. *Radiology.* 2013;266(3):701-715. doi:10.1148/radiol. 12110918

25. Kellman P, Hansen MS. T1-mapping in the heart: accuracy and precision. *J Cardiovasc Magn Reason.* 2014;16(2):1-20. doi:10.1186/1532-429X-16-2

26. Uğurbil 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. doi:10.1016/j.neuroimage.2013.05.012

27. Setsompop K, Kimmlingen R, Eberlein E, et al. Pushing the limits of in vivo diffusion MRI for the Human Connectome Project. *NeuroImage.* 2013;80: 220-233. doi:10.1016/j.neuroimage.2013.05.078

28. Tamir JI, Yu SX, Lustig M. Unsupervised Deep Basis Pursuit: Learning inverse problems without ground-truth data. Advances in Neural Information Processing Systems Workshops; 2019.

29. Senouf O, Vedula S, Weiss T, Bronstein A, Michailovich O, Zibulevsky M. Self-supervised learning of inverse problem solvers in medical imaging. arXiv:1905.09325; 2019. doi:10.1007/978-3-030-33391-1_13

30. Sim B, Oh G, Lim S, Ye JC. Optimal Transport, CycleGAN, and Penalized LS for Unsupervised Learning in Inverse Problems. arXiv:1909.12116; 2019.

31. 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(6):3172-3191. doi:10.1002/mrm.28378

32. Yaman B, Hosseini SAH, Moeller S, Ellermann J, Ugurbil K, Akcakaya M. Self-Supervised Physics-Based Deep Learning MRI Reconstruction Without Fully-Sampled Data. IEEE 17th International Symposium on Biomedical Imaging (ISBI); 2020:921-925.

33. Wang AQ, Dalca AV, Sabuncu MR. Neural Network-based Reconstruction in Compressed Sensing MRI Without Fully-sampled Training Data. arXiv: 2007.14979; 2020. doi:10.1007/978-3-030-61598-7_3

34. Ong F, Amin S, Vasanawala S, Lustig M. Mridata.org: An open archive for sharing MRI raw data. *Proceedings of the 26th Annual Meeting of ISMRM.* 2018;26(1).

35. Fessler JA. Optimization methods for magnetic resonance image reconstruction: key models and optimization algorithms. *IEEE Signal Process Mag.* 2020;37(1):33-40.

36. Hosseini SAH, Yaman B, Moeller S, Hong M, Akçakaya M. Dense recurrent neural networks for accelerated MRI: History-cognizant unrolling of optimization algorithms. *IEEE J Selected Topics sig Process.* 2020;14(6):1280-1291. doi:10.1109/JSTSP.2020.3003170

37. Breuer FA, Blaimer M, Mueller MF, et al. Controlled aliasing in volumetric parallel imaging (2D CAIPIRINHA). *Magn Reson Med.* 2006;55(3):549-556. doi:10.1002/mrm.20787

38. Wen B, Ravishankar S, Bresler Y. FRIST—Flipping and rotation invariant sparsifying transform learning and applications. *Inverse Problems.* 2017;33(7): 074007 doi:10.1088/1361-6420/aa6c6e

39. Ulyanov D, Vedaldi A, Lempitsky V. *Deep image prior.* In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018:9446-9454.

40. Darestani MZ, Heckel R. Accelerated MRI with un-trained neural networks. *IEEE Trans Comput Imaging.* 2021;7:724-733. doi:10.1109/TCI.2021. 3097596

41. 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(3):990-1001.

42. Tamir JI, Stella XY, Lustig M. Unsupervised Deep Basis Pursuit: Learning Reconstruction without Ground-Truth Data. Proceedings of the 27th Annual Meeting of ISMRM; 2019.

43. Lei K, Mardani M, Pauly JM, Vasanawala SS. Wasserstein GANs for MR imaging: from paired to unpaired training. *IEEE Trans Med Imaging.* 2020;40(1): 105-115. doi:10.1109/TMI.2020.3022968

44. Yaman B, Hosseini SAH, Akcakaya M. Improved Supervised Training of Physics-Guided Deep Learning Image Reconstruction with Multi-Masking. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2021.

45. Knoll F, Murrell T, Sriram A, et al. Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge. *Magn Reson Med.* 2020;84(6):3054-3070. doi:10.1002/mrm.28338

46. He K, Fan H, Wu Y, Xie S, Girshick R. Momentum contrast for unsupervised visual representation learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020:9729-9738.

47. Misra I, Maaten LVD. Self-supervised learning of pretext-invariant representations. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020:6707-6717.

48. Belthangady C, Royer LA. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat Methods.* 2019;16(12): 1215-1225. doi:10.1038/s41592-019-0458-z

49. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data.* 2019;6(1):1-48. doi:10.1186/s40537-019-0197-0

50. Yaman B, Hosseini SAH, Moeller S, Akçakaya M. Improved Supervised Training of Physics-Guided Deep Learning Image Reconstruction with Multi-Masking. *IEEE Int Conf Acoust, Speech Signal Process (ICASSP).* 2021;1150-1154.

51. Henderson P, Hu J, Romoff J, Brunskill E, Jurafsky D, Pineau J. Towards the systematic reporting of the energy and carbon footprints of machine learning. *J Mach Learn Res.* 2020;21(248):1-43.

52. Yaman B, Shenoy C, Deng Z, et al. Self-Supervised Physics-Guided Deep Learning Reconstruction For High-Resolution 3D LGE CMR. IEEE 18th International Symposium on Biomedical Imaging (ISBI); 2021.

53. Yaman B, Shenoy C, Deng Z, et al. Physics-Guided Self-Supervised Deep Learning Reconstruction without Fully-Sampled Data for Improved High-Resolution 3D LGE CMR. Proceedings of the Society for Cardiovascular Magnetic Resonance; 2021.

54. Demirel OB, Yaman B, Moeller S, et al. Improved Accelerated fMRI Reconstruction using Self-supervised Deep Learning. Proceedings of 29th Annual Meeting of the International Society of Magnetic Resonance in Medicine; 2021.

55. Demirel OB, Yaman B, Moeller S, et al. 20-fold Accelerated 7T fMRI Using Referenceless Self-Supervised Deep Learning Reconstruction. IEEE Engineering in Medicine & Biology Society (EMBC); 2021.

56. Sandino CM, Lai P, Vasanawala SS, Cheng JY. Accelerating cardiac cine MRI using a deep learning-based ESPIRiT reconstruction. *Magn Reson Med.* 2021;85(1):152-167. doi:10.1002/mrm.28420

57. Küstner T, Fuin N, Hammernik K, et al. CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions. *Sci Rep.* 2020;10(1):13710.

58. Qin C, Schlemper J, Duan J, et al. k-t NEXT: dynamic MR image reconstruction exploiting spatio-temporal correlations. International Conference on Medical Image Computing and Computer-Assisted Intervention; 2019:505-513. doi:10.1007/978-3-030-32245-8_56

59. Korkmaz Y, Dar SU, Yurt M, Özbey M, Çukur T. Unsupervised MRI Reconstruction via Zero-Shot Learned Adversarial Transformers. arXiv preprint arXiv:210508059 2021:41(7);1763. doi:10.1109/TMI.2022.3147426

60. Yaman B, Hosseini SAH, Akcakaya M. Zero-Shot Self-Supervised Learning for MRI Reconstruction. International Conference on Learning Representations (ICLR); 2022.

61. Akçakaya M, Yaman B, Chung H, Ye JC. Unsupervised deep learning methods for biological image reconstruction and enhancement. *IEEE Signal Process Mag.* 2021;39(2):28-44. doi:10.1109/MSP.2021.3119273

62. Wen B, Li Y, Bresler Y. The power of complementary regularizers: Image recovery via transform learning and low-rank modeling. arXiv preprint arXiv: 180801316; 2018.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.