Christopher M. Sandino, Joseph Y. Cheng, Feiyu Chen,
Morteza Mardani, John M. Pauly, and
Shreyas S. Vasanawala

Compressed Sensing: From Research to Clinical Practice With Deep Neural Networks

Shortening scan times for magnetic resonance imaging

©ISTOCKPHOTO.COM/GOODLIFESTUDIO

Digital Object Identifier 10.1109/MSP.2019.2950433
Date of current version: 15 January 2020

Compressed sensing (CS) reconstruction methods leverage sparse structure in underlying signals to recover high-resolution images from highly undersampled measurements. When applied to magnetic resonance imaging (MRI), CS has the potential to dramatically shorten MRI scan times, increase diagnostic value, and improve the overall patient experience. However, CS has several shortcomings that limit its clinical translation. These include 1) artifacts arising from inaccurate sparse modeling assumptions, 2) extensive parameter tuning required for each clinical application, and 3) clinically infeasible reconstruction times. Recently, CS has been extended to incorporate deep neural networks as a way of learning complex image priors from historical exam data. Commonly referred to as *unrolled neural networks*, these techniques have proven to be a compelling and practical approach to address the challenges of sparse CS. In this tutorial, we review the classical CS formulation and outline steps needed to transform this formulation into a deep-learning-based reconstruction framework. Supplementary open-source code in Python is used to demonstrate this approach with open databases. Further, we discuss considerations in applying unrolled neural networks in the clinical setting.

Introduction

MRI enables noninvasive visualization of soft tissue anatomy but is limited by long scan times and sensitivity to motion artifacts. Scan durations are directly related to the number of data samples measured; therefore, collecting fewer measurements enables faster imaging. However, sampling below the Nyquist rate will introduce aliasing artifacts, which may obscure anatomy and reduce diagnostic confidence. In fully sampled imaging, tradeoffs between the resolution and signal-to-noise ratio (SNR) are necessary to achieve shorter scan times and reduce sensitivity to patient motion.

Over the past few decades, constrained image reconstruction methods have been developed to enable rapid MRI techniques by leveraging prior information about the underlying signal to recover high-resolution images from relatively few measurements [1]. CS is one such method. It assumes the underlying

signal to be sparse (or compressible) in some manually chosen transform domain [2]. With this assumption in mind, the signal can be iteratively recovered from highly undersampled data by solving a regularized inverse problem. Further undersampling rates are achieved by combining CS with parallel imaging [3]–[5], which leverages the localized sensitivity profiles of each receiver element in a coil array. The total scan time acceleration achieved by combined parallel imaging and CS algorithms is powerful in enabling a broad range of clinical applications. For example, high-resolution volumetric imaging that would take minutes to acquire can now be achieved in a single breath-hold [6]. This strategy minimizes the sensitivity of patient motion and increases the diagnostic quality of the resulting images.

Over the past decade since CS was introduced to MRI, many developments have been made to extend this idea and bring it into clinical practice [7], [8]. One area in which significant research and clinical activity have been done is in multidimensional imaging. With more dimensions to exploit sparsity, the subsampling factor can be increased substantially (more than 10-fold). As a result, multidimensional scans can be completed in clinically feasible scan times. For example, a volumetric time-resolved blood-flow imaging sequence (4D flow) can be performed in a 5–10-min scan instead of the hour-long scan needed to satisfy the Nyquist criterion [9]. This single 4D flow scan enables a comprehensive cardiac evaluation with flow quantification, functional assessment, and anatomical information. Instead of an hour-long cardiac exam for congenital heart-defect patients with complex cardiac anomalies, the exam can be completed in a simple-to-execute single 4D flow scan [Figure 1(a)]. Other examples of multidimensional CS include dynamic-contrast-enhanced imaging [10] [Figure 1(b)] and "extradimensional" imaging with more than four dimensions [11].

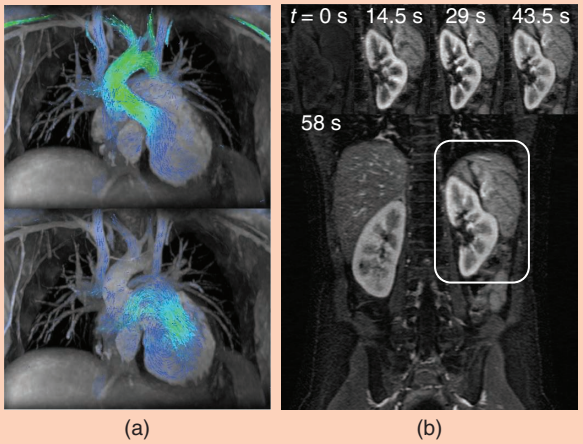

FIGURE 1. Images showing example clinical applications of CS. (a) Two cardiac phases of the cardiac-resolved volumetric velocity MR image (4D flow) for evaluating a congenital heart defect. To enable a clinically feasible scan duration of 5–15 min while maintaining high spatial ($0.9 \times 0.9 \times 1.6\,\text{mm}^3$) and high temporal resolutions (22 ms), a subsampling factor of 15 was used [9]. (b) High spatial resolution ($1 \times 1 \times 2\,\text{mm}^3$) with a 14.5-s temporal resolution was achieved for a dynamic contrast-enhanced MR image using an acceleration factor of 6.5 [10]. High spatiotemporal resolutions are required for capturing the rapid hemodynamics of pediatric patients.

Additionally, rapid imaging has a significant impact on the clinical workflow. Exam times can be significantly shortened to reduce patient burden and discomfort. For pediatric imaging, the shortened exam time enables the reduction of the depth and length of anesthesia [7]. For extremely short scan times (less than 15 min), anesthesia can be entirely eliminated.

Remaining challenges

Much success has been observed by applying CS to specific clinical applications, such as those related to pediatric [7] or volumetric cardiac imaging [8]. However, the potential that CS brings is not fully realized in terms of maximal acceleration and breadth of applications. A number of challenges limit the use of CS for clinical practice.

In CS, the regularization function is typically designed to promote sparsity of the signal in a manually chosen transform domain, such as wavelet or finite differences [2]. However, depending on the choice of transform, this assumption may not accurately model the underlying distribution of the data. For example, finite difference constraints (also known as *total variation*) can be used to dramatically reduce noise-like aliasing artifacts from the reconstructed images [12], but these rely on the assumption that the image consists of piecewise constant regions. For high undersampling rates, this can lead to reconstructions that appear cartoon-like and unrealistic. More complex priors are highly desirable to sufficiently represent the complexity of the target data.

Second, CS is sensitive to tuning various optimization parameters. One example of these parameters is the regularization parameter, which determines the weight of the regularization term with respect to the data consistency term in the optimization's objective. Increasing the regularization parameter will improve the perceived SNR of reconstructed images. However, while the perceived SNR improves, fine structures in the reconstructed images may be oversmoothed, or new image textures may be introduced, resulting in images with an artificial appearance. For high subsampling factors, the optimal value of the regularization parameter can vary among scans and patients, making manual parameter tuning infeasible. The impact of the regularization parameter on the reconstruction performance is illustrated in Figure 2.

Finally, CS usually has long reconstruction times due to the use of iterative optimization algorithms. To achieve clinically acceptable image quality, the number of iterations may vary from 50 to more than 100, depending on acquisition parameters [15]. For this reason, the number of iterations is set conservatively high to ensure convergence for various sets of acquisition parameters. Even with graphics processing unit (GPU) acceleration, reconstruction times can vary from several minutes to more than an hour, depending on the size of the data set, dimensionality, and the spatial encoding scheme. Long reconstruction times further lead to delays and queues in clinical workflows and may require expensive dedicated computational hardware.

Current solutions

Much research has been done in an effort to address these challenges. The performance of CS is highly dependent on the regularization function and regularization parameters used. To reduce textural artifacts and bias, the regularization needs to be carefully chosen and hand-tuned for each application. Another class of constrained reconstruction methods that leverage low-rank structure in the underlying signal was concurrently developed with sparse CS [16]. In contrast to sparsity-based constraints, which are enforced in a preselected transform domain, low-rank constraints are adaptive and depend on the subspace characteristics of each data set. These methods are highly suitable for certain dynamic imaging applications but can produce oversmoothed images when the underlying signals are not rank-deficient. Low-rank constraints have since been combined with sparsity-based constraints to achieve more expressive regularization functions [17]. However, these methods require extensive parameter tuning for each additional constraint.

Typically, the regularization parameter is fixed for all scans based on empirical tuning within a small number of data sets. This approach is shown to be effective for moderate subsampling rates across many patients in large clinical studies [7], [8]. Ideally, the regularization parameters for each scan should be chosen based on the system noise. However, many other factors, such as patient motion, contribute to measurement error. Previous works have shown that automatic parameter selection for each patient can improve the image quality of CS reconstructions [18]. This comes at the cost of additional reconstruction time; as a result, automatic parameter selection is uncommon in practice.

Regarding reconstruction times, preconditioning techniques can reduce the number of iterations necessary for algorithmic convergence, but these often require more computations per iteration [19]. Further acceleration in reconstruction time can be achieved by early truncation of the iterative reconstruction [20].

Most of these measures require manual adjustment of the original CS problem to specific applications. This tuning and redesigning process requires extra time and effort, and it calls for constant attention as hardware and clinical protocols evolve. Previous solutions to automate some of this process involve additional memory and more complex computations.

Overview

To overcome the obstacles with CS, deep learning has recently become a compelling and practical approach [13], [14], [21]–[25]. The purpose of this tutorial is to review an existing framework for extending CS to a deep-learning-based approach using unrolled neural networks. We also discuss a number of new challenges when using this framework, including considerations for clinical deployment. We have released supplementary Python code on GitHub (https://github.com/MRSRL/dl-cs) to demonstrate an example of this deep-learning -based CS framework.

Unrolled CS

Background

The MRI reconstruction problem can be formulated as a minimization problem [2]. The optimization consists of solving the following equation:

$$\hat{m} = \underset{m}{\operatorname{argmin}} \frac{1}{2} \| Am - y \|_2^2 + \lambda R(m), \tag{1}$$

where m is the reconstructed image set, A describes the imaging model, and y is the measured data in the k-space domain. The imaging model for MRI consists of signal modulation by coil sensitivity profile maps S, Fourier transform operation, and data subsampling. These sensitivity profile maps S are specific for each data set y.

The goal of this optimization is to reconstruct an image set m that best matches the measured data y in the least squares sense. For highly subsampled data sets, this problem is ill posed: many solutions satisfy (1); thus, the regularization function $R(m)$ and corresponding regularization parameter λ incorporate image priors to help constrain the problem. In CS, the regularization function is designed to promote transform sparsity in the optimal solution. Many optimization algorithms have been developed to solve the minimization problem in (1).

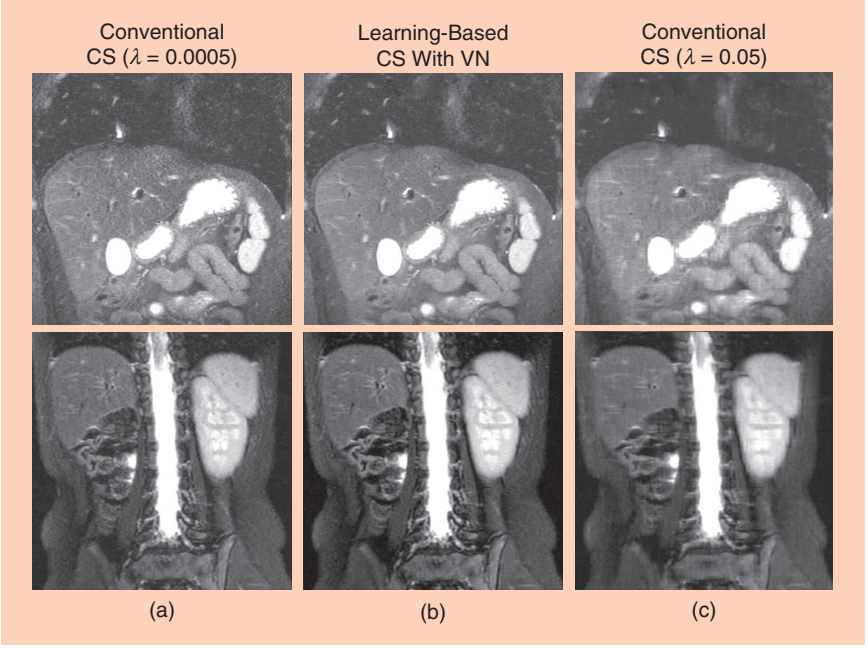

FIGURE 2. Example results of CS and variational network (VN) reconstructions of 3.25 × subsampled data acquired with a 32-channel torso coil [13], [14]. Conventional CS requires tuning of the regularization parameter λ. The optimal value of this parameter may vary with different scans. Compared to conventional CS using L1-ESPIRiT [5] [(a) and (c)], learning a VN (b) achieves proper regularization without the need for manual tuning the regularization parameter.

For simplicity, we base our discussion on the proximal gradient method. We refer the reader to similar approaches based on other optimization algorithms [13], [21], [24].

To solve (1), we split the problem into two alternating steps that are repeated. For the kth iteration, a gradient update is performed as

$$m^{(k+)} = m^{(k)} - 2tA^H(Am^{(k)} - y), \tag{2}$$

where A^H is the transpose of the imaging model, and t is a scalar specifying the size of the gradient step. The current guess of image m is denoted here as $m^{(k+)}$. Afterwards, the proximal problem with regularization function R is solved as follows:

$$m^{(k+1)} = \mathbf{prox}_{\lambda R}(m^{(k+)}) = \underset{u}{\operatorname{argmin}} R(u) + \frac{1}{2t\lambda}\left\| u - m^{(k+)} \right\|_2^2, \tag{3}$$

where u is a helper variable that transforms the regularization into a convex problem that can be more readily solved. The updated guess of image m at the end of this kth iteration is denoted as $m^{(k+1)}$, and $m^{(k+1)}$ is then used for the next iteration in (2). This proximal problem is a simple soft-thresholding step for specific regularization functions, such as $R(m) = \|\Psi m\|_1$, where Ψ is a wavelet transform [20].

Data-driven learning

To develop fast and robust reconstruction algorithms for different MRI sequences and scans, a compelling alternative is to take a data-driven approach to learn the optimal regularization functions and parameters. Some approaches have attempted to directly learn a parameterized form of the regularization function using a field-of-experts model [13], [14]. This tutorial focuses on a different approach, which is to learn the proximal step in (3) and therefore implicitly learn both the regularization function and regularization parameter. In this setup, as described in [22], [23], and [26], the proximal step is replaced with a deep neural network capable of learning a more natural regularization function than transform sparsity. With this in mind, the proximal update step in (3) can be rewritten as follows:

$$m^{(k+1)} = E_{\theta_k}(m^{(k+)}), \tag{4}$$

where E_{θ_k} is the neural network and θ_k is the set of learned parameters.

The steps of (2) and (4) can be unrolled with a fixed number of iterations [27] and be denoted as model $G_\theta(y,A)$ with inputs of measurements y and imaging model A. The training of such a model can then be performed using the following loss function:

$$\min_\theta \sum_i \left\| G_\theta(y_i, A_i) - m_i \right\|_2^2, \tag{5}$$

where m_i is the ith ground-truth example that is retrospectively subsampled by a sampling mask M in the k-space domain to generate $y_i = MA_i^H m_i$. For deployment, new scans are acquired according to the sampling mask M, and the measured data y_i can be used to reconstruct images $\hat{m}_i$ as

$$\hat{m}_i = G_\theta(y_i, A_i), \tag{6}$$

where A_i contains sensitivity profile maps S_i, that can be estimated using algorithms like JSENSE [4] or ESPIRiT [5]. An example network is shown in Figure 3(a) and (b).

Neural network design for unrolled MRI reconstruction

Important factors must be considered when applying deep neural networks to MRI reconstruction. This includes handling of complex data [29], circular convolutions in the image domain for Cartesian data [24], and incorporating the acquisition model [13].

Since the data measured during an MRI scan is complex, complex data are used for the MRI reconstruction process. As a result, the networks used for MRI reconstruction need to handle complex data types. Two approaches can be used to address this issue. The first approach is to convert the complex data into two channels of data. This conversion can be performed as concatenating the magnitude of the data with the phase of the data in the channels dimension or as concatenating the real component with the imaginary component of the data in the channels dimension. The second approach is to build the neural network with operations that directly support and exploit structure in complex data. Several efforts have been made to enable complex operations in the convolutional layers (with complex backpropagation) and the activation functions [29]. For demonstration purposes, we construct the networks in Figure 3(b) and (c) with the first approach because of its simplicity.

Another consideration in building neural networks for MRI reconstruction is handling the convolutional operation at the image edges. Assuming zeros beyond the image edges for image-domain convolutions is sufficient for most cases, especially when the imaging volume is surrounded by air and when no high-intensity signal exists near the edges of the field of view (FOV). However, when these conditions are not satisfied, this assumption may result in residual aliasing artifacts. These artifacts are usually observed when the imaging object is larger than the FOV. In Cartesian imaging, data are measured in the Fourier space and transformed using the fast Fourier transform algorithm to the image domain; as a result, signals are circularly wrapped. Thus, circular convolutions should be performed when applying convolutions in the image domain. For this purpose, we first circularly pad the images before applying the convolutional layers [Figure 4(b)]. The padding width is set such that the "valid" portion of the output has the same image dimensions as the input. Note that for data acquired using non-Cartesian trajectories, such as radial or spiral, signals outside of the FOV will instead manifest as noise-like artifacts across the image. In these cases, circular padding should not be used.

Leveraging a more accurate MRI acquisition model in the neural network will improve the reconstruction performance and simplify the network architecture. More specifically, data are measured using multichannel coil receivers. The sensitivity profile maps of the different coil elements can be leveraged as a strong prior to help constrain the reconstruction problem. Further, by exploiting the (soft) SENSE model [5], the multichannel complex data (y) are reduced to a single-channel complex image (m) before each convolutional network block, denoted as the proximal block in Figure 3(b). As a further benefit, the learned network can be trained and applied to data sets with different numbers of coil channels, because the input to the learned convolutional network block only requires a single-channel complex image.

Training data

Reconstructed image quality can vary depending on the quality, authenticity, and diversity of the training data used to train the network. Therefore, it is important to consider how training data are collected, preprocessed, and augmented.

Training data should be collected at the point in the imaging pipeline where the model will be deployed. For the purpose of image reconstruction of subsampled data sets, we would need to collect the raw measurement *k*-space data. Typically, raw imaging data are not readily available as only magnitude images are saved in the Digital Imaging and Communications in Medicine (DICOM) format and stored in hospital imaging databases. Furthermore, these stored magnitude images are often processed with image filters, which can bias network outputs. Additionally, accurate simulation of raw imaging data from DICOM images is difficult, if not impossible, to perform, especially in simulating realistic phase information. To facilitate development, a number of different open data initiatives have been recently launched, including mridata.org [30] and fastMRI [31]. These resources provide an initial starting point for development, but more data sets from various vendors and of varying acquisition parameters and field strengths are needed.

Data augmentation strategies should be used during training to prevent overfitting to the original training data set. However, care must be taken when applying data augmentation transformations. For example, data interpolation is needed for random rotations that may degrade the quality of the input data. Other image domain operations may introduce unrealistic errors in measuring the *k*-space domain, such as aliasing in the *k*-space domain. Flipping and transposing the data set will preserve the original data quality; thus, these operations can be included in the training.

Finally, data normalization strategies can help improve the training and the final performance of the learned model. For simplicity, the raw measurement data can be normalized according to

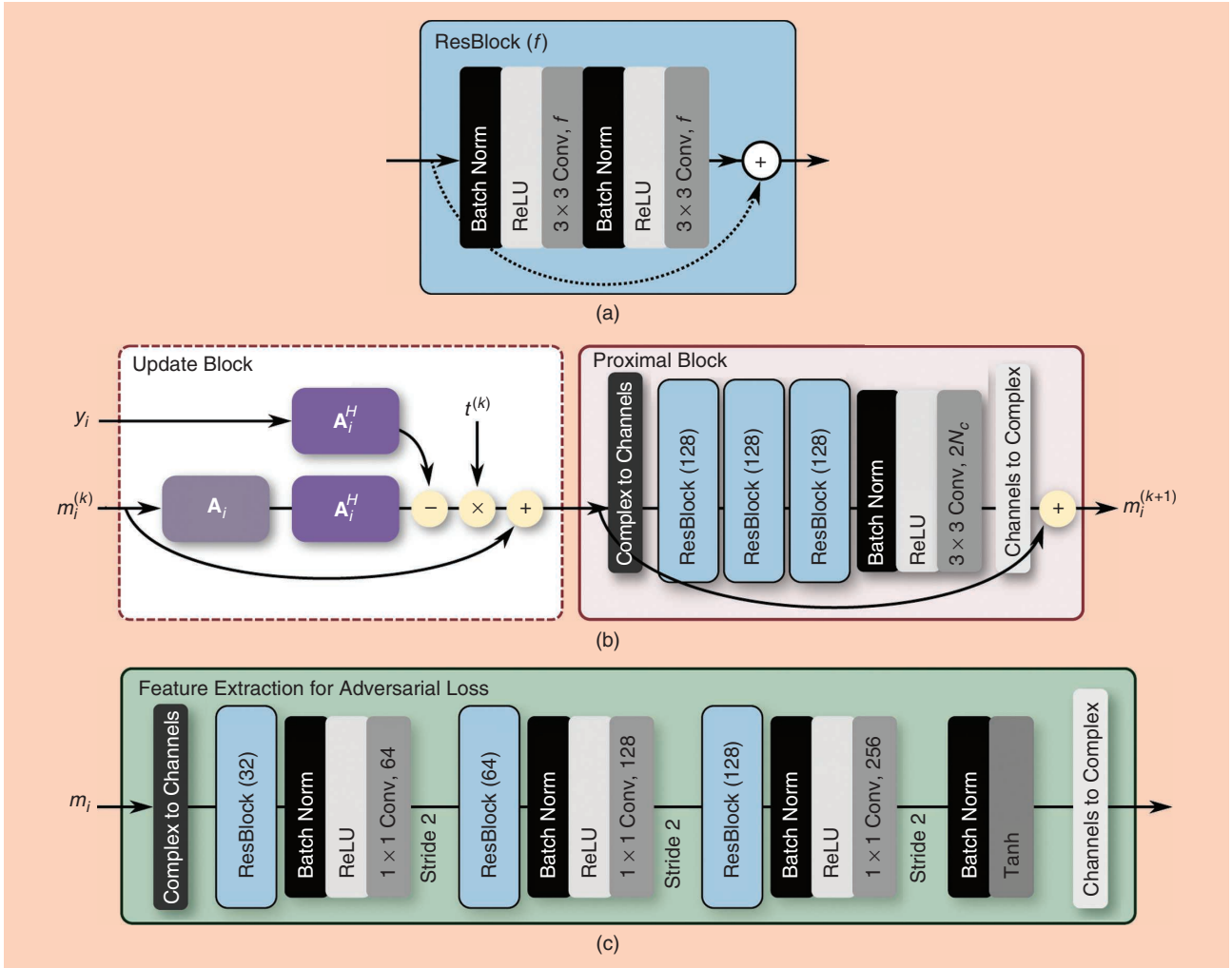

FIGURE 3. Diagrams showing a neural network architecture for MR image reconstruction. One residual block (ResBlock) [28] is illustrated in (a) with *f* feature maps. The ResBlock is used as a building block for the different networks. (b) One iteration of the reconstruction network where the *i*th data set is passed through the *k*th iteration. Matrix A_i represents the imaging model, m_i represents the data set in the image domain, and y_i represents the data set in the *k*-space domain. The final output can be passed through the network in (c) to extract feature maps that can be compared to the feature maps extracted from the ground-truth data using the same network. The tanh activation function in (c) is used to ensure that the values in the outputted feature maps are within ±1. Batch Norm: batch normalization; ReLU: rectified linear unit; Conv: convolutional.

the L2 norm of the central k-space signals. In our demonstration, we normalize the input data by the L2 norm of the central 5×5 region of k-space. Due to similarities to CS algorithms, training convergence and reconstruction accuracy may be improved by data prewhitening and normalization by an estimate of the noise statistics among different coil array receivers. These noise statistics can be measured with a fast calibration scan, during which data are measured with no radio-frequency excitation.

Loss function

The performance of unrolled CS is highly dependent on the loss function used. The easiest loss functions to use for training are L1 and L2 losses. The L2 loss is described in (5) and can be converted to an L1 loss by using the L1 norm instead of the L2 norm.

L1 and L2 losses have been previously used to successfully train supervised reconstruction networks [13], [23]. Alternative loss functions that further capture the idea of structure or perception have also been studied. For example, a perceptual loss function can be constructed using a network pretrained on natural images to extract "perceptual" features [32]. Though general, this feature-extraction network should be trained for the specific problem domain and the specific task at hand. Generative adversarial networks [33] can be used to model the properties of the ground-truth images and to exploit that information for improving the reconstruction quality.

A feature-extraction network $D_\omega(m)$, shown in Figure 3(c), is trained to extract the necessary features to compare the reconstruction output with the ground truth. The training loss function in (5) becomes

$$\min_{\theta} \max_{\omega} \sum_i \left\| D_\omega(G_\theta(y_i, A_i)) - D_\omega(m_i) \right\|_2, \tag{7}$$

where parameters ω are optimized to maximize the difference between the reconstructed image $G_\theta(y_i, A_i)$ and the ground-truth data m_i. At the same time, parameters θ are optimized to minimize the difference between the reconstruction and the ground-truth data after passing both these images through network D_ω to extract feature maps. The optimization of (7) consists of alternating between the training of parameters ω

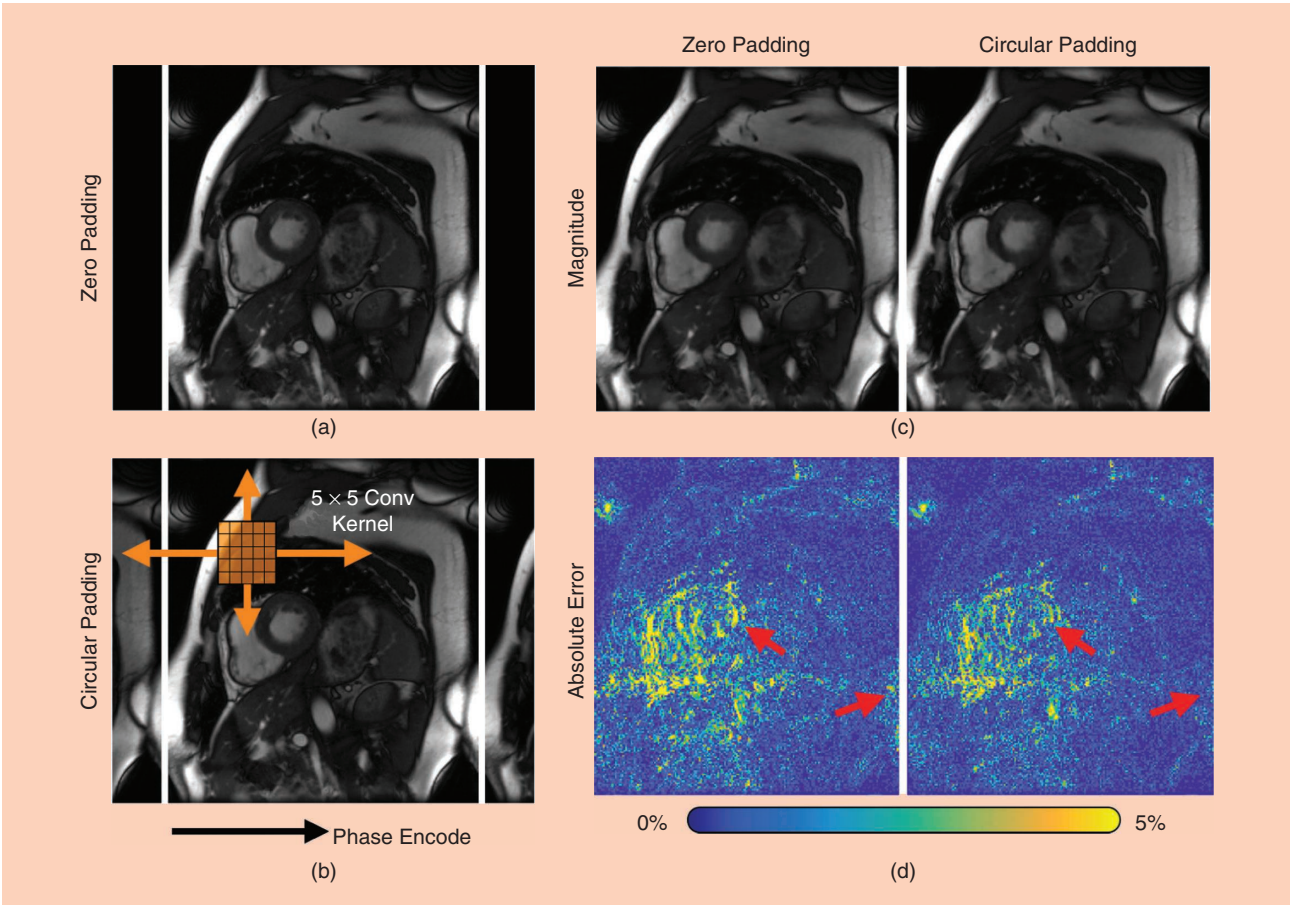

FIGURE 4. Images showing circular convolutional (Conv) layer for MRI reconstruction. (a) Current deep-learning frameworks support zero padding for convolutional layers to maintain the original dimensions. However, data are measured in the frequency space, and the fast Fourier transform algorithm is used to convert this data into the image domain. Thus, for Cartesian imaging, circular convolutions should be performed. (b) This circular convolution can be performed by first circularly padding the input data and cropping the result to the original input dimension. For cardiac cine imaging, images can also be circular-padded in time to take advantage of cyclical dynamics, such as heart motion. (c) Magnitude images of 8×-subsampled data sets reconstructed using a 3D unrolled neural network with and without circular spatiotemporal padding. (d) The absolute error shows that images reconstructed with circular convolutions depict reduced error at image edges and heart tissue at end-systole (red arrows).

with the θ constant and the training of parameters θ with the ω constant. We refer to this training approach as training with an *adversarial loss*.

This min–max loss function in (7) can be unstable and difficult to train. Two main components have been described in [33] to help stabilize the training. First, G_θ is pretrained using either an L1 or L2 loss so that the parameters in this network are properly initialized. The network is then fine-tuned with the adversarial network using a reduced training rate. Second, a pixelwise cost function is added to (7):

$$\min_{\theta}\max_{\omega}\sum_{i}\mu\left\|D_\omega(G_\theta(y_i,A_i))-D_\omega(m_i)\right\|_2^2+\left\|G_\theta(y_i,A_i)-m_i\right\|_2^2. \tag{8}$$

The images before passing through D_ω can be considered as additional feature maps to help stabilize the training process. The hyperparameter μ is used to weigh between the two components in the loss function and must be heuristically tuned during the training phase.

Demonstration

To demonstrate the unrolled CS reconstruction and enable reproducibility of the results, we downloaded 20 fully sampled volumetric knee data sets freely available in the database of MRI raw data (mridata.org) [30]. Each volume was collected with an eight-channel knee coil, proton density weighting, and 320 slices in the readout direction in x. Each of the x-slices were treated as separate examples during training and validation. The data sets were divided by subject: 15 subjects for training (4,800 x-slices), two subjects for validation (640 x-slices), and three subjects for testing. Sensitivity maps were estimated using JSENSE [4]. Poisson-disk sampling masks [2] were generated using an acceleration factor R of 9.4 with corner cutting (effective R of 12) and a fully sampled calibration region of 20 × 20.

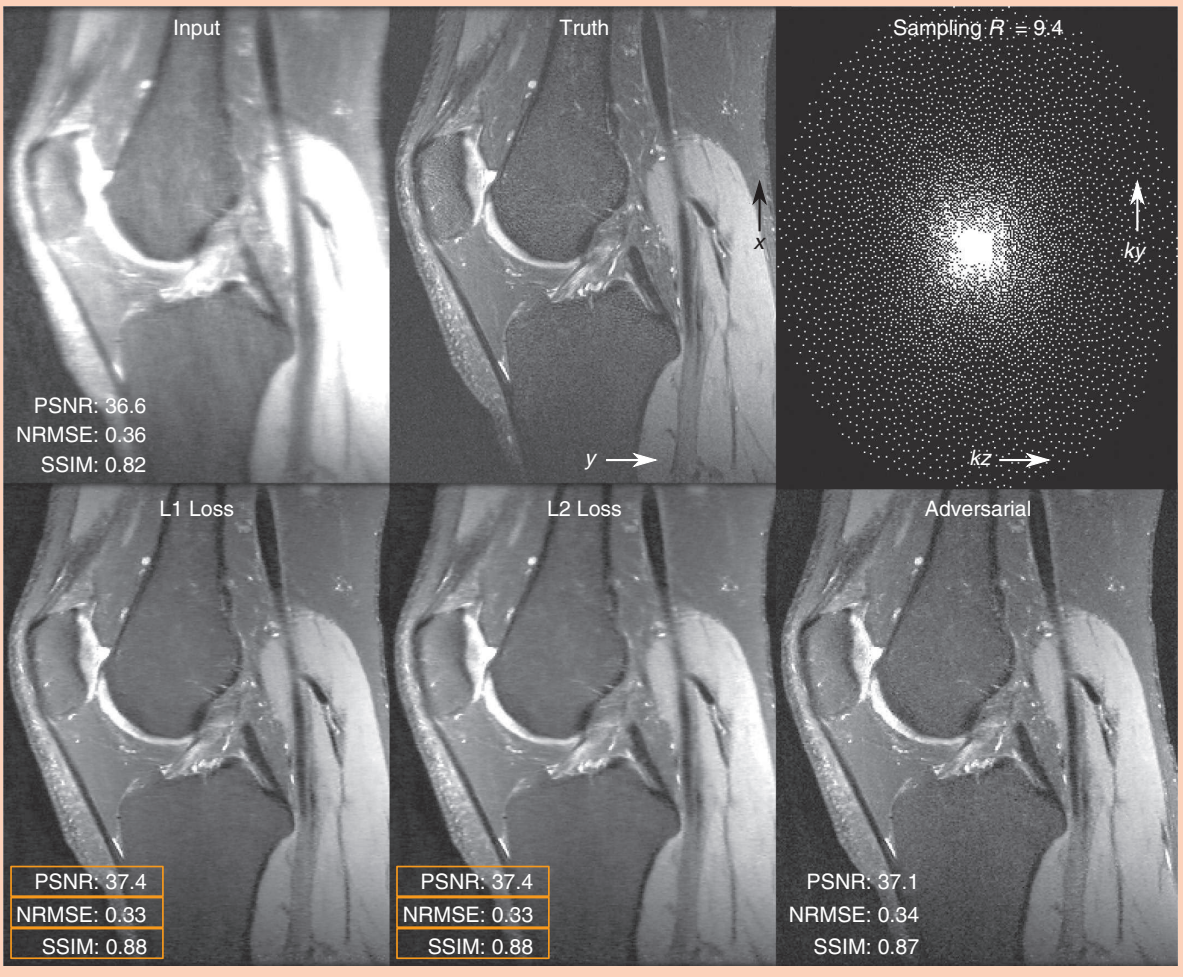

FIGURE 5. Images showing demonstration results of a volumetric knee data set subsampled with an acceleration factor R of 9.4 with corner cutting (effective R of 12). In the bottom row, the volume was reconstructed slice by slice using three networks trained with different loss functions. The reconstruction using the network trained using the L1 and L2 losses yielded the best results in terms of peak SNR (PSNR), normalized root-mean-square error (NMRSE), and structural similarity (SSIM). However, the reconstruction using the network trained with the adversarial loss yielded results with the most realistic texture.

The networks in Figure 3 were implemented in Python using the TensorFlow framework. Additional reconstruction components were performed using the SigPy Python package (https://github.com/mikgroup/sigpy). The reconstruction network was built using four iterations; this setup allowed for faster training for demonstration purposes. Performance can be improved by increasing the number of iterations. Training and experiments were performed on a single Nvidia Titan Xp graphics card with 12 GB of memory, which supported a batch size of 2. The network was trained multiple times using different loss functions: an L1 loss for 20,000 steps and L2 loss for 20,000 steps. For adversarial loss, the network was first pretrained using an L1 loss for 10,000 steps and then jointly trained with the adversarial loss and an L1 loss for 60,000 steps (10,000 steps were for the reconstruction network and 50,000 steps were for the adversarial feature-extraction network). The data preparation, training, validation, and testing Python scripts are available on GitHub. Example results are shown in Figure 5.

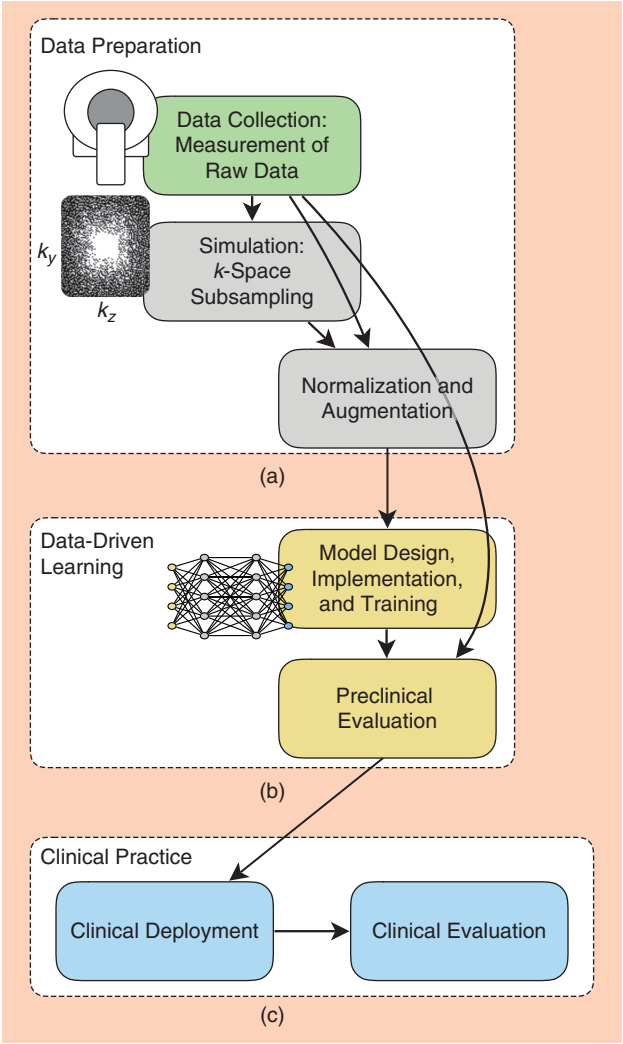

FIGURE 6. Diagrams showing an example development cycle for the data-driven learning framework. (a) Fully sampled raw k-space data sets are first collected and used to perform realistic simulation of subsampling. (b) The reconstruction algorithm is then built, trained, and internally validated. (c) The reconstruction model is deployed and clinically evaluated.

Clinical integration

Example setup

Many of these algorithms are computationally intensive, but these algorithms can leverage off-the-shelf consumer hardware. The bulk of the computational burden is now in the network, and GPUs can be leveraged for this purpose. The compute system only needs the minimal CPU and memory requirements to support the GPUs. Using an Nvidia Titan Xp graphics card, one 320 × 256 slice took on average 0.1 s to reconstruct, 1.8 s for a batch of 16 slices, and a total of 36 s for the entire volume with 320 slices. These benchmarks include the time to transfer the data to and from the CPU to the GPU. The computational speed can be further improved using newer GPU hardware and/or more cards. Alternatively, inference on the reconstruction network can be performed on the CPU to leverage more memory in a more cost-effective manner, but this setup comes with slower inference speed. Figure 6 illustrates an example development cycle for the data-driven learning framework.

Clinical cases

Through initial developments, we have deployed a number of different models at our clinical site. Example results of unrolled CS with variational networks (VNs) [13], [14] are shown in Figure 2. Conventional CS using L1-ESPIRiT [5] achieved high noise when the regularization parameter was too low (0.0005) [Figure 2(a)] and high residual and blurring artifacts when the regularization parameter was too high (0.05) [Figure 2(c)]. The optimal value of the regularization parameter may vary with different scans. Compared to conventional CS, learning a VN [Figure 2(b)] achieved proper regularization without the need to tune the regularization parameter.

The best image prior for multidimensional imaging is more difficult to engineer, but these larger data sets benefit tremendously

Imaging Metrics

Imaging metrics are commonly used to evaluate results. This includes peak SNR (PSNR) and normalized root-mean-square error (NRMSE). For these quantities, we use the following equations:

$$\mathrm{MSE}(x, x_r) = \frac{1}{NM}\sum_{i}^{N}\sum_{j}^{M}|x[i,j] - x_r[i,j]|^2, \quad \text{(S1)}$$

$$\mathrm{PSNR}(x, x_r) = 10\log_{10}(\max(|x_r|^2)/\mathrm{MSE}), \quad \text{(S2)}$$

$$\mathrm{NRMSE}(x, x_r) = \sqrt{\mathrm{MSE}}\,/\sqrt{\frac{1}{NM}\sum_{i}^{N}\sum_{j}^{M}|x_r[i,j]|^2}, \quad \text{(S3)}$$

where x denotes the test image, $x[i, j]$ denotes the value of the test image at pixel (i, j), and x_r denotes the reference ground-truth image.

from subsampling due to long acquisition times. Thus, for multidimensional space, a compelling approach is to use data-driven learning. In Figure 7, a reconstruction network consisting of 3D spatiotemporal convolutions was trained on 12 fully sampled, breath-held, multislice cardiac cine data sets acquired with a balanced steady-state free precession readout at multiple scan orientations. The reconstruction network enabled the acquisition of a cine slice in a single heartbeat and a full stack of cine slices in a single breath-hold [34].

Evaluation

For development and prototyping, image metrics provide a straightforward method to evaluate performance. These metrics include peak SNR, normalized root-mean-square error, and structural similarity. These metrics are described in "Imaging Metrics." However, what is more important in the clinical setting is the ability for clinicians to make an informed decision. Therefore, we recommend performing clinical studies.

For comparing methods, the evaluation of clinical studies can be performed with blinded grading of image quality by multiple radiologists. This evaluation can be based on set criteria, including overall image quality, perceived SNR, image contrast, image sharpness, residual artifacts, and confidence of diagnosis [14]. Scores from 1 to 5 or from –2 to 2 are given for each clinical patient scan with a specific definition for each score for objectivity. An example of scoring for SNR and image sharpness is described in Table 1.

A total number of more than 20 consecutive scans are usually collected to achieve a comprehensive and statistically meaningful evaluation for an initial study. The difference between two reconstruction approaches can be statistically tested with Wilcoxon tests on the null hypothesis to show that there is no significant difference between two approaches. The Wilcoxon test is a nonparametric statistical test that evaluates the null hypothesis that the scores from the two different reconstructions are from the same distribution. Ideally, the new reconstruction algorithm should yield results from a distribution different from that of the original algorithm.

Discussion

This data-driven approach to accelerated imaging has the potential to eliminate many of the challenges associated with CS, including the need to design hand-crafted priors and to hand tune the regularization functions. In addition, the described framework inspired by iterative inference algorithms provides a principled approach for designing reconstruction networks [22]. However, the data-driven reconstruction framework faces new challenges, including data scarcity, generalizability, reliability, and meaningful metrics.

High-quality training labels or fully sampled data sets are scarce, especially in the clinical settings where patient motion

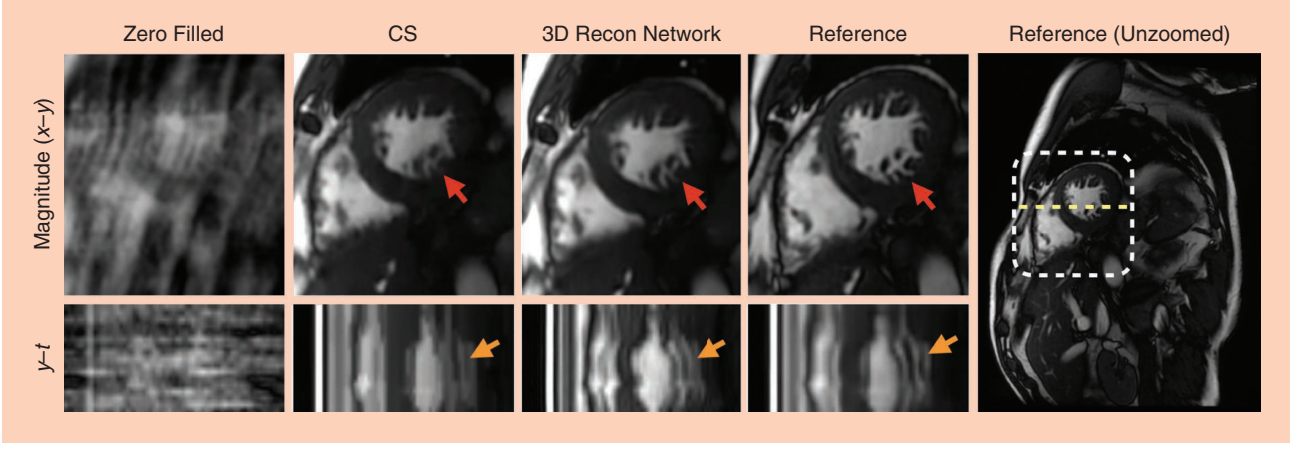

FIGURE 7. Images showing clinical deployment of a 3D spatiotemporal convolutional reconstruction (Recon) network for 2D cardiac-resolved imaging. The top row shows a prospectively acquired short-axis data set acquired with 11.6-fold subsampling using a 32-channel cardiac coil that was reconstructed using CS (L1-ESPIRiT) [5] and a trained 3D reconstruction network. These reconstructions were compared with the fully sampled reference images (acquired in a separate scan). The 3D network reconstructed fine structures, such as papillary muscles and trabeculae, with less blurring than L1-ESPIRiT (red arrows). Furthermore, *y–t* profiles (plotted along yellow dotted line) show that the 3D network images depict more natural cardiac motion compared to L1-ESPIRiT with total variation, which introduced staircasing artifacts (orange arrows).

Table 1. Example scoring criteria to evaluate methods.

Score	Overall Image Quality	SNR	Sharpness
1	Nondiagnostic	All structures appear to be too noisy.	Some structures are not sharp on most images.
2	Limited	Most structures appear to be too noisy.	Most structures are sharp on some images.
3	Diagnostic	Few structures appear to be too noisy on most images.	Most structures are sharp on most images.
4	Good	Few structures appear to be too noisy on a few images.	All structures are sharp on most images.
5	Excellent	There is no noticeable noise on any of the images.	All structures are sharp on all images.

affects image quality and where lengthy fully sampled acquisitions would be impractical to perform if a faster solution existed. To address data scarcity, an important future direction pertains to designing compact network architectures that are effective with small training labels and are possibly trained in an unsupervised fashion. Early attempts were reported in [26], where recurrent neural networks were leveraged for learning proximal operators using only a couple of residual blocks, which performed well for small training-sample sizes.

While neural networks do not need to be manually hand-tuned scan to scan like in conventional CS approaches, they may need to be retrained or fine-tuned for each application to maximize performance. Knoll et al. showed that reconstructed image quality degrades when a network is applied to data acquired with a different SNR from that of the training data [35]. Since the SNR can vary drastically across different applications, the network's ability to generalize will depend on the heterogeneity of the noise statistics in the training data. To improve generalizability, the network can be retrained or fine-tuned on heterogeneous data acquired with the SNR and acquisition parameter range of interest. However, this strategy requires sufficient training data sets for all different settings to be readily available.

Given the scarcity of training examples and the cost to collect these examples, another strategy is to design and train networks that are highly generalizable. This generalizability can be achieved with smaller networks, as discussed, or by training with a larger image manifold, such as natural images. With this approach, a loss in performance may be observed since a larger-than-necessary manifold is learned [36]. In the worst case, the images reconstructed using this data-driven approach should not be worse than images reconstructed using a more conservative approach, such as zero filling or parallel imaging [3], [5].

Regarding generalizability, an important question concerns reconstruction performance for subjects with unseen abnormalities. Patient abnormalities can be quite heterogeneous, and these abnormalities are rare and unlikely to be included in the training data set. If not designed carefully, generative networks might remove critical features or create false features that could result in a misdiagnosis. The optimization-based network architecture uses data consistency, as exemplified by the gradient update step, for the image-recovery problem. However, the inherent ambiguity of ill-posed problems does not guarantee faithful recovery. Therefore, a systematic study is required to analyze the recovery of images. Also, effective regularization techniques (possibly through adversarial training) are needed to avoid missing important diagnostic information. More efforts are needed to develop a holistic quality score that captures the uncertainty in the acquisition scheme and training data.

Developing for medical images a standard unbiased metric that assesses the authenticity of medical images is extremely important. Here, we discussed the use of different loss functions to train the network and common imaging metrics to evaluate the images. Additionally, we discussed an example of a possible clinical evaluation that can be performed to assess the algorithm in the clinical setting. However, the reconstruction task should ideally be optimized and evaluated for the end task, which can consist of detection and quantification. With significant effort in automating image interpretation, this data-driven framework provides an opportunity to pursue the ability to train the reconstruction end to end for the ultimate goal of improving patient care.

In conclusion, deep learning has the potential to increase the accessibility and generalizability of rapid imaging through data subsampling. Previous challenges with CS can be approached with a data-driven framework to create a solution that is more readily translated to clinical practice.

Acknowledgments

This work was supported by in part by GE Healthcare and by the National Institutes of Health under grants NIH R01-EB009690, NIH R01-EB026136, and NIH R01-EB019241.

Authors

Christopher M. Sandino (sandino@stanford.edu) received his B.S. degree in electrical engineering from the University of Southern California, Los Angeles, in 2015. He is a Ph.D. candidate in electrical engineering at Stanford University, California. His research focus is in applying deep-learning methods to dynamic magnetic resonance imaging techniques to reduce scan time and image artifacts. He is a recipient of the National Science Foundation Graduate Research Fellowship.

Joseph Y. Cheng (jycheng@stanford.edu) received his B.S. and M.Eng. degrees from the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, Cambridge, in 2006 and 2007, respectively. He received his Ph.D. degree from the Department of Electrical Engineering at Stanford University, California, in 2013, where he is currently a senior research scientist in the Department of Radiology. His current research interests include the development and application of deep learning to medical imaging. He received the International Society for Magnetic Resonance in Medicine W.S. Moore Young Investigator Award for Original Clinical Research in 2015. He is a Member of the IEEE.

Feiyu Chen (feiyuc@stanford.edu) received his B.S. degree in biomedical engineering from Tsinghua University, Beijing, China, in 2014. He received his Ph.D. degree in electrical engineering from Stanford University, California. He is interested in applying deep-learning methods to magnetic resonance imaging (MRI). His other interests include designing new sampling schemes and developing new pulse sequences for MRI.

Morteza Mardani (morteza@stanford.edu) received his Ph.D. degree in electrical engineering and mathematics from the University of Minnesota, Minneapolis–St. Paul, in 2015. He is a research scientist in the Departments of Electrical Engineering and Radiology at Stanford University, California. His research interests include machine learning and statistical signal processing for data science and artificial intelligence. He is currently working on deep learning and generative

adversarial neural networks for medical imaging. He is a recipient of a Young Author Best Paper Award from the IEEE Signal Processing Society (2017) and a Best Student Paper Award from the IEEE Workshop on Signal Processing Advances in Wireless Communications (2012). He is a Member of the IEEE.

John M. Pauly (pauly@stanford.edu) received his Ph.D. degree in electrical engineering from Stanford University, California, in 1990, where he is currently a professor in the Department of Electrical Engineering and also codirects the Magnetic Resonance Image Laboratory. His research interests include developing wireless magnetic resonance imaging systems, and magnetic resonance-guided interventional procedures. In 2012, he was given the Gold Medal from the International Society of Magnetic Resonance in Medicine. He is a Senior Member of the IEEE.

Shreyas S. Vasanawala (vasanawala@stanford.edu) received his B.S. degree in mathematics from the California Institute of Technology, Pasadena, in 1994. He received his M.D. degree and Ph.D. degree in biophysics with research in magnetic resonance imaging (MRI) techniques from Stanford University, California, where he is currently a professor in the Department of Radiology. His research efforts are focused on developing fast and quantitative MRI methods. He serves as the director of MRI at Stanford Hospital and Clinics and at Stanford Children's Health. He also serves as the division chief of Body MRI.

References

[1] Z.-P. Liang, F. E. Boda, R. T. Constable, E. M. Haacke, P. C. Lauterbur, and M. R. Smith, "Constrained reconstruction methods in MR imaging," *Rev. Magn. Reson. Med.*, vol. 4, pp. 67–185, 1992.

[2] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[3] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, no. 5, pp. 952–962, 1999.

[4] L. Ying and J. Sheng, "Joint image reconstruction and sensitivity estimation in SENSE (JSENSE)," *Magn. Reson. Med.*, vol. 57, no. 6, pp. 1196–1202, 2007.

[5] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, "ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA," *Magn. Reson. Med.*, vol. 71, no. 3, pp. 990–1001, 2014.

[6] T. Zhang, S. Chowdhury, M. Lustig, R. A. Barth, M. T. Alley, T. Grafendorfer, P. D. Calderon, F. J. Robb et al., "Clinical performance of contrast enhanced abdominal pediatric MRI with fast combined parallel imaging compressed sensing reconstruction," *J. Magn. Reson. Imaging*, vol. 40, no. 1, pp. 13–25, 2014.

[7] S. S. Vasanawala, M. T. Alley, B. A. Hargreaves, R. A. Barth, J. M. Pauly, and M. Lustig, "Improved pediatric MR imaging with compressed sensing," *Radiology*, vol. 256, no. 2, pp. 607–616, 2010.

[8] T. A. Basha, M. Akçakaya, C. Liew, C. W. Tsao, F. N. Delling, G. Addae, L. Ngo, W. J. Manning et al., "Clinical performance of high-resolution late gadolinium enhancement imaging with compressed sensing," *J. Magn. Reson. Imaging*, vol. 46, no. 6, pp. 1829–1838, 2017.

[9] J. Y. Cheng, K. Hanneman, T. Zhang, M. T. Alley, P. Lai, J. I. Tamir, M. Uecker, J. M. Pauly et al., "Comprehensive motion-compensated highly accelerated 4D flow MRI with ferumoxytol enhancement for pediatric congenital heart disease," *J. Magn. Reson. Imaging*, vol. 43, no. 6, pp. 1355–1368, 2016.

[10] T. Zhang, J. Y. Cheng, A. G. Potnick, R. A. Barth, M. T. Alley, M. Uecker, M. Lustig, J. M. Pauly et al., "Fast pediatric 3D free-breathing abdominal dynamic contrast enhanced MRI with high spatiotemporal resolution," *J. Magn. Reson. Imaging*, vol. 41, no. 2, pp. 460–473, 2015.

[11] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, "XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing," *Magn. Reson. Med.*, vol. 75, no. 2, pp. 775–788, 2016.

[12] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint," *Magn. Reson. Med.*, vol. 57, no. 6, pp. 1086–1098, 2007.

[13] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, 2018.

[14] F. Chen, V. Taviani, I. Malkiel, J. Y. Cheng, J. I. Tamir, J. Shaikh, S. T. Chang, C. J. Hardy et al., "Variable-density single-shot fast spin-echo MRI with deep learning reconstruction by using variational networks," *Radiology*, vol. 289, no. 2, pp. 366–373, 2018.

[15] M. Murphy, M. Alley, J. Demmel, K. Keutzer, S. Vasanawala, and M. Lustig, "Fast l1-SPIRiT compressed sensing parallel imaging MRI: Scalable parallel implementation and clinically feasible runtime," *IEEE Trans. Med. Imag.*, vol. 31, no. 6, pp. 1250–1262, 2012.

[16] Z. P. Liang, "Spatiotemporal imaging with partially separable functions," in *Proc. 4th IEEE Int. Symp. Biomedical Imaging: From Nano to Macro*, 2007. doi: 10.1109/ISBI.2007.357020.

[17] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1042–1054, 2011.

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

[19] K. Koolstra, J. van Gemert, P. Börnert, A. Webb, and R. Remis, "Accelerating compressed sensing in parallel imaging reconstructions using an efficient circulant preconditioner for Cartesian trajectories," *Magn. Reson. Med.*, vol. 81, no. 1, pp. 670–685, 2019.

[20] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[21] Y. Yang, J. Sun, H. Li, and Z. Xu, ADMM-Net: A deep learning approach for compressive sensing MRI. 2017. [Online]. Available: https:arXiv:org/abs/1705.06869

[22] S. Diamond, V. Sitzmann, F. Heide, and G. Wetzstein, Unrolled optimization with deep priors. 2017. [Online]. Available: https:arXiv:org/abs/1705.08041

[23] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 491–503, 2018.

[24] J. Y. Cheng, F. Chen, M. T. Alley, J. M. Pauly, and S. S. Vasanawala, Highly scalable image reconstruction using deep neural networks with bandpass filtering. 2018. [Online]. Available: https:arXiv:org/abs/1805.03300

[25] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, 2019.

[26] M. Mardani, Q. Sun, S. Vasawanala, V. Papyan, H. Monajemi, J. Pauly, and D. Donoho, Neural proximal gradient descent for compressive imaging. 2018. [Online]. Available: https:arXiv:org/abs/1806.03963

[27] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Machine Learning*, 2010, pp. 399–406.

[28] K. He, X. Zhang, S. Ren, and J. Sun, Identity mappings in deep residual networks. 2016. [Online]. Available: https:arXiv:org/abs/1603.05027

[29] P. Virtue, S. X. Xu, and M. Lustig, "Better than real: Complex-valued neural nets for MRI fingerprinting," in *IEEE Int. Conf. Image Processing*, Beijing, 2017, pp. WA–PF.12.

[30] F. Ong, S. Amin, S. S. Vasanawala, and M. Lustig, "An open archive for sharing MRI raw data," in *ISMRM & ESMRMB Joint Annu. Meeting*, Paris, France, 2018, p. 3425.

[31] J. Zbontar, F. Knoll, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras et al., fastMRI: An open dataset and benchmarks for accelerated MRI. 2018. [Online]. Available: https:arXiv:11811.08839

[32] J. Johnson, A. Alahi, and L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution. 2016. [Online]. Available: https:arXiv:org/abs/1603.08155

[33] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 167–179, 2019.

[34] C. Sandino, P. Lai, M. A. Janich, A. Brau, S. S. Vasanawala, and J. Y. Cheng, "ESPIRiT with deep priors: Accelerating 2D cardiac CINE MRI beyond compressed sensing," in *Proc. SCMR/ISMRM Workshop*, Seattle, Washington, 2019.

[35] F. Knoll, K. Hammernik, E. Kobler, T. Pock, M. P. Recht, and D. K. Sodickson, "Assessment of the generalization of learned image reconstruction and the potential for transfer learning," *Magn. Reson. Med.*, vol. 81, no. 1, pp. 116–128, 2019.

[36] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.

SP