



Published in final edited form as:

*IEEE Signal Process Mag.* 2020 January ; 37(1): 128–140. doi:10.1109/MSP.2019.2950640.

## Deep-Learning Methods for Parallel Magnetic Resonance Imaging Reconstruction: A Survey of the Current Approaches, Trends, and Issues

**Florian Knoll [Member, IEEE], Kerstin Hammernik, Chi Zhang [Student Member, IEEE], Steen Moeller, Thomas Pock, Daniel K. Sodickson, Mehmet Akçakaya [Member, IEEE]**

F. Knoll and D. K. Sodickson are with the Center for Biomedical Imaging, Department of Radiology, New York University. K. Hammernik is with the Department of Computing, Imperial College London. T. Pock is with the Institute of Computer Graphics and Vision, Graz University of Technology. C. Zhang and M. Akçakaya are with the Department of Electrical and Computer Engineering, and Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, MN. S. Moeller is with the Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, MN.

### Abstract

Following the success of deep learning in a wide range of applications, neural network-based machine learning techniques have received interest as a means of accelerating magnetic resonance imaging (MRI). A number of ideas inspired by deep learning techniques from computer vision and image processing have been successfully applied to non-linear image reconstruction in the spirit of compressed sensing for both low dose computed tomography and accelerated MRI. The additional integration of multi-coil information to recover missing k-space lines in the MRI reconstruction process, is still studied less frequently, even though it is the de-facto standard for currently used accelerated MR acquisitions. This manuscript provides an overview of the recent machine learning approaches that have been proposed specifically for improving parallel imaging. A general background introduction to parallel MRI is given that is structured around the classical view of image space and k-space based methods. Both linear and non-linear methods are covered, followed by a discussion of recent efforts to further improve parallel imaging using machine learning, and specifically using artificial neural networks. Image-domain based techniques that introduce improved regularizers are covered as well as k-space based methods, where the focus is on better interpolation strategies using neural networks. Issues and open problems are discussed as well as recent efforts for producing open datasets and benchmarks for the community.

---

Personal use is permitted, but republication/redistribution requires IEEE permission. See [http://www.ieee.org/publications\\_standards/publications/rights/index.html](http://www.ieee.org/publications_standards/publications/rights/index.html) for more information.

<sup>11</sup>Contact Information Of Corresponding Author: Mehmet Akçakaya, McKnight-Land Grant Assistant Professor, Department of Electrical and Computer Engineering, and Center for Magnetic Resonance Research, University of Minnesota, 200 Union Street SE, Keller Hall Rm. 5-159, Minneapolis, MN, USA 55455, [akcakaya@umn.edu](mailto:akcakaya@umn.edu).

<sup>4</sup><https://github.com/hkaggarwal/modl>

**Index Terms—**

Accelerated MRI; Parallel Imaging; Iterative Image Reconstruction; Numerical Optimization; Machine Learning; Deep learning

---

**I. Introduction**

During recent years, there has been a substantial increase of research activity in the field of medical image reconstruction. One particular application area is the acceleration of Magnetic Resonance Imaging (MRI) scans. This is an area of high impact, because MRI is the leading diagnostic modality for a wide range of exams, but the physics of its data acquisition process make it inherently slower than modalities like X-Ray, Computed Tomography or Ultrasound. Therefore, the shortening of scan times has been a major driving factor for routine clinical application of MRI.

One of the most important and successful technical developments to decrease MRI scan time in the last 20 years was parallel imaging [1–3]. All state of the art clinical MRI scanners from major vendors are equipped with parallel imaging technology, and it is the default option for a large number of scan protocols. As a consequence, there is a substantial benefit of using multi-coil data for machine learning based image reconstruction. Not only does it provide a complementary source of acceleration that is unavailable when operating on single channel data, or on the level of image enhancement and post-processing, it also is the scenario that ultimately defines the use-case for accelerated clinical MRI, which makes it a requirement for clinical translation of new reconstruction approaches. The drawback is that working with multi-coil data adds a layer of complexity that creates a gap between cutting edge developments in deep learning [4] and computer vision, where the default data type are images. The goal of this manuscript is to bridge this gap by providing both a comprehensive review of the properties of parallel MRI, together with an introduction how current machine learning methods can be used for this particular application.

**A. Background on multi-coil acquisitions in MRI**

The original motivation behind phased array receive coils was to increase the SNR of MR measurements. These arrays consist of  $n_c$  multiple small coil elements, where an individual coil element covers only a part of the imaging field of view. These individual signals are then combined to form a single image of the complete field of view. The central idea of all parallel imaging methods is to complement spatial signal encoding of gradient fields with information about the spatial position of these multiple coil elements. For multiple receiver coils, the MR signal equation can be written as follows

$$f_j(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, y) c_j(x, y) e^{-i(k_x x + k_y y)} dx dy. \quad (1)$$

In Equation (1),  $f_j$  is the MR signal of coil  $j = 1, \dots, n_c$ ,  $u$  is the target image to be reconstructed, and  $c_j$  is the corresponding coil sensitivity. Parallel imaging methods use the

redundancies in these multi-coil acquisitions to reconstruct undersampled k-space data. After discretization, this undersampling is described in matrix-vector notation by

$$\mathbf{f}_j = \mathbf{F}_\Omega \mathbf{C}_j \mathbf{u} + \mathbf{n}_j, \quad (2)$$

where  $\mathbf{u}$  is the image to be reconstructed,  $\mathbf{f}_j$  is the acquired k-space data in the  $j^{\text{th}}$  coil,  $\mathbf{C}_j$  is a diagonal matrix containing the sensitivity profile of the receiver coil [2],  $\mathbf{F}_\Omega$  is a partial Fourier sampling operator that samples locations  $\Omega$ , and  $\mathbf{n}_j$  is measurement noise in the  $j^{\text{th}}$  coil.

Historically, parallel imaging methods were put in two categories: Approaches that operate in image domain, inspired by the sensitivity encoding (SENSE) method [2] and approaches that operated in k-space, inspired by simultaneous acquisition of spatial harmonics (SMASH) [1] and generalized autocalibrating partial parallel acquisition (GRAPPA) [3]. This is conceptually illustrated in Figure 1. While these two schools of thought are closely related, we organized this document according to these classic criteria for historical reasons.

## II. Classical parallel imaging in image space

Classical parallel imaging in image space follows the SENSE method [2], which can be identified by two key features. First, the elimination of the aliasing artifacts is performed in image space after the application of an inverse Fourier transform. Second, information about receive coil sensitivities is obtained via precomputed, explicit coil sensitivity maps from either a separate reference scan or from a fully sampled block of data at the center of k-space (all didactic experiments that are shown in this manuscript follow the latter approach). The reconstruction in image domain in Figure 1 shows three example undersampled coil images, corresponding coil sensitivity maps and the final reconstructed images from a brain MRI dataset. The coil sensitivities were estimated using ESPIRiT [5].

MRI reconstruction in general and parallel imaging in particular can be formulated as an inverse problem. This provides a general framework that allows easy integration of the concepts of regularized and constrained image reconstruction as well as machine learning that are discussed in more detail in later sections. Equation (1) can be discretized and then written in matrix-vector form

$$\mathbf{f} = \mathbf{E} \mathbf{u} + \mathbf{n}, \quad (3)$$

where  $\mathbf{f}$  contains all k-space measurement data points and  $\mathbf{E}$  is the forward encoding operator that includes information about the sampling trajectory and the receive coil sensitivities and  $\mathbf{n}$  is measurement noise. The task of image reconstruction is to recover the image  $\mathbf{u}$ . In classic parallel imaging the number of receive elements is usually larger than the acceleration factor, and Equation (3) is solved in the least-squares-sense via the Pseudo-Inverse  $(\mathbf{E}^* \mathbf{E})^{-1} \mathbf{E}$ , where  $\mathbf{E}^*$  denotes the conjugate transpose of  $\mathbf{E}$ . The reason why the acceleration factor is smaller than the number of coils is that these individual coil elements do not measure completely independent information. This leads to an increase of the condition number  $(\mathbf{E}^* \mathbf{E})^{-1}$  and therefore, an ill-posed problem. This can lead to severe noise amplification in the reconstruction. In the original SENSE formulation [2], this noise

amplification can be described exactly via the g-factor. In practice, Equation (3) is usually solved in an iterative manner, which is the topic of the following sections.

### A. Overview of conjugate gradient SENSE (CG-SENSE)

The original SENSE approach is based on equidistant or uniform Cartesian k-space sampling, where the aliasing pattern is defined by a point spread function that has a small number of sharp equidistant peaks. This property leads to a small number of pixels that are folded on top of each other, which allows a very efficient implementation [2]. When using alternative k-space sampling strategies like non-Cartesian acquisitions or random undersampling, this is no longer possible and image reconstruction requires a full inversion of the encoding matrix in Equation (3). This operation is demanding both in terms of compute and memory requirements (the dimensions of  $\mathbf{E}$  are the total number of acquired k-space points times  $N^2$  where  $N$  is the size of the image matrix that is to be reconstructed), which lead to the development of iterative methods, in particular the CG-SENSE method introduced by Pruessmann et al. as a follow up of the original SENSE paper [6]. In iterative image reconstruction the goal is to find a  $\hat{\mathbf{u}}$  that is a minimizer of the following cost function, which corresponds to the quadratic form of the system in Equation (3)

$$\hat{\mathbf{u}} \in \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{E}\mathbf{u} - \mathbf{f}\|_2^2. \quad (4)$$

In standard parallel imaging,  $\mathbf{E}$  is linear and Equation (4) is a convex optimization problem that can be solved with a large number of numerical algorithms like gradient descent, Landweber iterations or the conjugate gradient method. However, since MR k-space data are corrupted by noise, it is common practice to stop iterating before theoretical convergence is reached, which can be seen as a form of regularization. Regularization can be also incorporated via additional constraints in Equation (4), which will be covered in the next section.

As a didactic example for this manuscript, we will use a single slice of a 2D coronal knee exam to illustrate various reconstruction approaches. This data were acquired on a clinical 3T system (Siemens Skyra) using a 15 channel phased array knee coil. A turbo spin echo sequence was used with the following sequence parameters: TR=2750ms, TE=27ms, echo train length=4, field of view 160mm<sup>2</sup> in-plane resolution 0.5mm<sup>2</sup>, slice thickness 3mm. Readout oversampling with a factor of 2 was used, and all images were cropped in the frequency encoding direction (superior-inferior) for display purposes. In the spirit of reproducible research, data, sampling masks and coil sensitivity estimations that were used for the numerical results in this manuscript are available online<sup>1</sup>. Figure 3 shows an example of a retrospectively undersampled CG-SENSE reconstruction with an acceleration factor of 4. The numeric tolerance with respect to the norm of the normalized residual was set to  $5 \cdot 10^{-5}$ , which resulted in 10 CG iterations.

<sup>1</sup>[https://app.globus.org/file-manager?originid=15c7de28-a76b-11e9-821c-02b7a92d8e58&originpath=%2Fknee%2F%20coronal proton density \(pd\) weighted data, subject 17, slice 25.](https://app.globus.org/file-manager?originid=15c7de28-a76b-11e9-821c-02b7a92d8e58&originpath=%2Fknee%2F%20coronal%20proton%20density%20(pd)%20weighted%20data%20subject%2017%20slice%2025)

## B. Nonlinear regularization and compressed sensing

As mentioned in Section II, the acceleration factor in classic parallel imaging is limited by the number of independent channels of the receive coil array. To push the boundaries of this limit, additional a-priori knowledge can be introduced. This is achieved by extending Equation (4) via additional penalty terms, which results in a constrained optimization problem defined in Equation (5), which forms the cornerstone of almost all modern MRI reconstruction methods

$$\hat{\mathbf{u}} \in \arg \min_{\mathbf{u}} \frac{1}{2} \left\| \mathbf{E}\mathbf{u} - \mathbf{f} \right\|_2^2 + \sum_i \lambda_i \Psi_i(\mathbf{u}). \quad (5)$$

Here,  $\Psi_i$  are dedicated regularization terms and  $\lambda_i$  are regularization parameters that balance the trade-off between data fidelity and prior. Since the introduction of compressed sensing and its adoption for MRI [7, 8], nonlinear regularization terms, in particular  $\ell_1$ -norm based ones, are popular in image reconstruction and are commonly used in parallel imaging [8–14]. The goal of these regularization terms is to provide a separation between the target image that is to be reconstructed from the aliasing artifacts that are introduced due to an undersampled acquisition. Therefore, they are usually designed in conjunction with a particular data sampling strategy. The classic formulation of compressed sensing in MRI [7] is based on sparsity of the images in a transform domain in combination with pseudo-random sampling, which introduces aliasing artifacts that are incoherent in the respective domain. While wavelets are a popular choice for static imaging, sparsity in the Fourier domain is commonly used for dynamic applications, where periodic motion is encountered. Total variation based methods have been used successfully in combination with radial and spiral acquisitions as well as in dynamic imaging. More advanced regularizers based on low-rank properties have also been utilized. In contrast to linear reconstructions, where the quality of a reconstruction can be assessed via SNR and g-factor maps, the evaluation of image quality is not trivial in the context of nonlinear reconstructions. Noise is suppressed at the cost of introducing a bias from the nonlinear regularizer. Therefore, it is generally not recommended to use SNR-based metrics as a measure of image quality. Image quality is therefore usually estimated with metrics like NRMSE, SSIM or PSNR, which compare a reconstruction with a reference gold standard, ideally a fully sampled reconstruction. However, this is generally only possible in a research setting, and image quality evaluation of nonlinear reconstruction methods without a reference is still an open research problem in the field.

Figure 3 shows an example of a nonlinear combined parallel imaging and compressed sensing reconstruction with a Total Generalized Variation [10] constraint. The raw data was scaled such that the maximum magnitude value in k-space is 1 and the regularization parameter  $\lambda$  was set to  $2.5 \cdot 10^{-5}$  and the reconstruction was using 1000 primal-dual iterations. While it is usually recommended to use a pseudo-random acquisition when combining parallel imaging with compressed sensing, we chose equidistant sampling for our experiments here for consistency with classic parallel imaging reconstruction methods. A more detailed discussion of this is provided in Section IV and in [15]. The nonlinear

regularization still provides a superior reduction of aliasing artifacts and noise suppression in comparison to the CG-SENSE reconstruction from the last section.

### III. Classical parallel imaging in K-space

Parallel imaging reconstruction can also be formulated in k-space as an interpolation procedure. The initial connections between the SENSE-type image-domain inverse problem approach and k-space interpolation has been made more than a decade ago [16], where it was noted that the forward model in Equation (3) can be restated in terms of the Fourier transform,  $\boldsymbol{\kappa}$  of the combined image,  $\mathbf{u}$  as

$$\mathbf{f} = \mathbf{E}\mathbf{F} * \boldsymbol{\kappa} \triangleq \mathbf{G}_{\text{acq}}\boldsymbol{\kappa}, \quad (6)$$

where  $\mathbf{E}$  is the forward encoding operator and  $\mathbf{F}$  is the discrete Fourier transform (DFT) matrix as before,  $\mathbf{f}$  corresponds to the acquired k-space lines across all coils, and  $\mathbf{G}_{\text{acq}}$  is a linear operator. Similarly, the unacquired k-space lines across all coils can be formulated using

$$\mathbf{f}_{\text{unacq}} = \mathbf{G}_{\text{unacq}}\boldsymbol{\kappa}. \quad (7)$$

Combining these two equations yield

$$\mathbf{f}_{\text{unacq}} = \mathbf{G}_{\text{unacq}}\mathbf{G}_{\text{acq}}^\dagger\mathbf{f}. \quad (8)$$

Thus, the unacquired k-space lines across all coils can be interpolated based on the acquired lines across all coils, assuming the pseudo-inverse,  $\mathbf{G}_{\text{acq}}^\dagger$ , of  $\mathbf{G}_{\text{acq}}$  exists [16]. Thus, the main difference between the k-space parallel imaging methods and the aforementioned image domain parallel imaging techniques is that the former produces k-space data across all coils at the output, whereas the latter typically produces one image that combines the information from all coils.

#### A. Linear k-space interpolation in GRAPPA

The most clinically used k-space reconstruction method for parallel imaging is GRAPPA, which uses linear shift-invariant convolutional kernels to interpolate missing k-space lines using uniformly-spaced acquired k-space lines [3]. For the  $j^{\text{th}}$  coil k-space data,  $k_j$ , we have

$$\kappa_j(k_x, k_y - m\Delta k_y) = \sum_{c=1}^{n_c} \sum_{b_x=-B_x}^{B_x} \sum_{b_y=-B_y}^{B_y} g_{j,m}(b_x, b_y, c)\kappa_c(k_x - b_x\Delta k_x, k_y - Rb_y\Delta k_y), \quad (9)$$

where  $R$  is the acceleration rate of the uniformly under-sampled acquisition;  $m \in \{1, \dots, R-1\}$ ;  $g_{j,m}(b_x, b_y, c)$  are the linear convolutional kernels for estimating the  $m^{\text{th}}$  spacing location in the  $j^{\text{th}}$  coil;  $n_c$  is the number of coils; and  $B_x, B_y$  are parameters determined from the convolutional kernel size. A high-level overview of such interpolation is shown in the reconstruction in k-space section of Figure 1.

Similar to the coil sensitivity estimation in SENSE-type reconstruction, the convolutional kernels  $g_{j,m}(b_x, b_y, c)$  are estimated for each subject, from either a separate reference scan or from a fully-sampled block of data at the center of k-space, called autocalibrating signal (ACS) [3]. A sliding window approach is used in this calibration region to identify the fully-sampled acquisition locations specified by the kernel size and the corresponding missing entries. The former, taken across all coils, is used as rows of a calibration matrix,  $\mathbf{A}$ ; while the latter, for a specific coil, yields a single entry in the target vector,  $\mathbf{b}$ . Thus for each coil  $j$  and missing location  $m \in \{1, \dots, R-1\}$ , a set of linear equations are formed, from which the vectorized kernel weights  $g_{j,m}(b_x, b_y, c)$ , denoted  $\mathbf{g}_{j,m}$ , are estimated via least squares, as  $\mathbf{g}_{j,m} \in \arg \min_{\mathbf{g}} \|\mathbf{b} - \mathbf{A}\mathbf{g}\|_2^2$ . GRAPPA has shown to have several favorable properties compared to SENSE, including lower g-factors, sometimes even less than unity in parts of the image, and more smoothly varying g-factor maps [17].

## B. Advances in k-space interpolation methods

Though GRAPPA is widely used in clinical practice, it is a linear method that suffers from noise amplification based on the coil geometry and acceleration rate [2]. Therefore, several strategies have been proposed in the literature to reduce the noise in reconstruction.

Iterative self-consistent parallel imaging reconstruction (SPIRiT) is a strategy for enforcing self-consistency among the k-space data in multiple receiver coils by exploiting correlations between neighboring k-space points [9]. Similar to GRAPPA, SPIRiT also estimates a linear shift-invariant convolutional kernel from ACS data. In GRAPPA, this convolutional kernel used information from acquired lines in a neighborhood to estimate a missing k-space point. In SPIRiT, the kernel includes contributions from all points, both acquired and missing, across all coils for a neighborhood around a given k-space point. The self-consistency idea suggests that the full k-space data should remain unchanged under this convolution operation. The SPIRiT objective function also includes a term that enforces consistency with the acquired data, where the undersampling can be performed with arbitrary patterns, including random patterns that are typically employed in compressed sensing [7, 8]. Additionally, this formulation allows for incorporation of regularizers, for instance based on transform-domain sparsity, in the objective function to reduce reconstruction noise via non-linear processing in a method called  $\ell_1$ -SPIRiT [9]. Furthermore, SPIRiT has facilitated the connection between coil sensitivities used in image-domain parallel imaging methods and the convolutional kernels used in k-space methods via a subspace analysis in a method called ESPIRiT [5]. It was shown that the k-space based  $\ell_1$ -SPIRiT and coil sensitivity-based  $\ell_1$ -ESPIRiT perform similarly.

An alternative line of work utilizes non-linear k-space interpolation for estimating missing k-space points for uniformly undersampled parallel imaging acquisitions [18]. It was noted that during GRAPPA calibration, both the regressand and the regressor have errors in them due to measurement noise in the acquisition of calibration data, which leads to a non-linear relationship in the estimation. Thus, the reconstruction method, called non-linear GRAPPA (NL-GRAPPA), uses a kernel approach to map the data to a higher-dimensional feature space, where linear interpolation is performed, which also corresponds to a non-linear



interpolation in the original data space. The interpolation function is estimated from the ACS data, although this approach typically required more ACS data than GRAPPA [18]. This method was shown to reduce reconstruction noise compared to GRAPPA. Note that NL-GRAPPA, through its use of the kernel approach, is a type of machine learning approach, though the non-linear kernel functions were empirically fixed a-priori and not learned from data. In another line of work, GRAPPA regularization during calibration was explored using a sparsity-promoting [19] approach. These approaches use regularization in calibration followed by a one-step reconstruction. However, we note that this way is different than the regularization in  $\ell_1$ -SPIRiT, which uses the regularization during reconstruction in an iterative manner.

### C. Low-rank matrix completion for k-space reconstruction

While k-space interpolation methods remain the prevalent method for k-space parallel imaging reconstruction, there has been recent efforts on recasting this type of reconstruction as a matrix completion problem. Simultaneous autocalibrating and k-space estimation (SAKE) is an early work in this direction, where local neighborhoods in k-space across all coils are restructured into a matrix with block Hankel form [20]. Then low-rank matrix completion is performed on this matrix, subject to consistency with acquired data, enabling k-space parallel imaging reconstruction without additional calibration data acquisition. Low-rank matrix modeling of local k-space neighborhoods (LORAKS) is another method exploiting similar ideas, where the motivation is based on utilizing finite image support and image phase constraints instead of correlations across multiple coils, and which was also extended to parallel imaging to further include the similarities between image supports and phase constraints across coils [21]. A further generalization to LORAKS is annihilating filter-based low rank Hankel matrix approach (ALOHA), which extends the finite support constraint to transform domains [22]. By relating transform domain sparsity to the existence of annihilating filters in a weighted k-space, where the weighting is determined by the choice of transform domain, ALOHA recasts the reconstruction problem as the low-rank recovery of the associated Hankel matrix.

## IV. Machine learning methods for parallel imaging in image space

The use of machine learning for image-based parallel MR imaging evolves naturally from Equation (5) based on the following key insights. First, in classic compressed sensing,  $\Psi$  are a general regularizers like the image gradient or wavelet transforms, which were not designed specifically with undersampled parallel MRI acquisitions in mind. These regularizers can be generalized to models that have a higher computational complexity.  $\Psi$  can be formulated as a convolutional neural network (CNN), where the model parameters can be learned from training data inspired by the concepts of deep learning [4], as illustrated in Figure 2. This was already demonstrated earlier in the context of computer vision with a non-convex regularizer of the following form

$$\Psi(\mathbf{u}) = \sum_{i=1}^{N_k} \langle \rho_i(\mathbf{K}_i \mathbf{u}), \mathbf{1} \rangle. \quad (10)$$



The regularizer in Equation (10) consists of  $N_k$  terms of non-linear potential functions  $\rho_i$ , and  $\mathbf{K}_i$  are convolution operators.  $\mathbf{1}$  indicates a vector of ones. The parameters of the convolution operators and the parametrization of the non-linear potential functions form the free parameters of the model, which are learned from training data.

The second insight is that the iterative algorithm that is used to solve Equation (5) naturally maps to the structure of a neural network, where every layer in the network represents an iteration step of a classic algorithm. This follows naturally from gradient descent for the least squares problem in Equation (4) that leads to the iterative Landweber method. After choosing an initial  $\mathbf{u}^0$ , the iteration scheme is given by Equation (11)

$$\mathbf{u}^s = \mathbf{u}^{s-1} - \alpha^s \mathbf{E}^* (\mathbf{E} \mathbf{u}^{s-1} - \mathbf{f}), \quad s > 0. \quad (11)$$

$\mathbf{E}^*$  is the adjoint of the encoding operator and  $\alpha^s$  is the step size of iteration  $s$ . Using this iteration scheme to solve the reconstruction problem in Equation (5) with the regularizer defined in Equation (10) leads to the update scheme defined in Equation (12), which forms the basis of recently proposed image space based machine learning methods for parallel MRI

$$\mathbf{u}^s = \mathbf{u}^{s-1} - \alpha^s \left( \sum_{i=1}^{N_k} (\mathbf{K}_i)^\top \rho'_i(\mathbf{K}_i \mathbf{u}^{s-1}) + \lambda^s \mathbf{E}^* (\mathbf{E} \mathbf{u}^{s-1} - \mathbf{f}) \right). \quad (12)$$

This update scheme can then be represented as a neural network with  $S$  stages corresponding to  $S$  iteration steps in Equation (12).  $\rho'_i$  are the first derivatives of the nonlinear potential functions  $\rho_i$ , which are represented as activation functions in the neural network. The transposed convolution operations  $\mathbf{K}_i^\top$  correspond to convolutions with filter kernels rotated by 180 degrees. The idea of the variational network [15] follows the structure of classic variational methods and gradient-based optimization, and the network architecture is designed to mimic a classic iterative image reconstruction. Since this convolutional steps in this architecture are shallow, the network has a small receptive field ( $11 \times 11$  convolution kernels were used in [15, 23]) and the network cannot capture global image information. The regularizer can be modified by including elements like pooling layers, upconvolutions and skip-connections, following the popular U-Net model [24]. This extends the computational capacity of the regularizer and given sufficient training data samples, will generally improve the performance of the model. However, it comes at the cost of losing the direct connection to gradient based optimization, which makes the approach less interpretable. It can also lead to overfitting in situations where only a small set of training data is available. The method from Aggarwal et al. [25] is an example for such an approach. It follows a similar design concept as the VN, but uses a deeper and more complex regularizer. In order to limit the total number of model parameters, the same set of weights is used for all stages of the network. The other major difference is that in contrast to the gradient in the VN (see Equation (11)), it realizes the data term via a proximal mapping that is implemented as an unrolled conjugate-gradient scheme.

An experiment that compares the properties of image space based machine learning for parallel MRI to CG-SENSE and constrained reconstructions from the previous sections is shown in Figure 3. The architecture and training of the VN exactly follows the description in [15] and the model consists of 131,050 model parameters that are different for all 10 stages in the network architecture. The MoDL formulation and training is a modification of the approach in [25]. The same learned deep learning regularizer was repeated across all 10 stages. In contrast to the approach described in the paper, the regularizer was modified to a U-Net [24] with a total of 694,021 model parameters. The publicly available multi-channel knee data that was described in Section II was used to train the networks. The source for the approaches that are used in these experiments is available online<sup>2,3,4</sup>. The figure shows a single slice of a test case that was not used during training. The normalized root mean sum of squares error (NRMSE) and structural similarity index (SSIM) to the fully sampled reference is shown next to each reconstruction. The experiments illustrates the improved performance as the complexity of the regularizer model is increased. Equidistant Cartesian sampling from traditional parallel imaging was used in all experiments because this type of sampling is predominantly used in clinical practice. For combined parallel imaging and compressed sensing, pseudo-random sampling is generally recommended to improve performance in the literature [7]. An analysis of the influence of the sampling scheme is not included here due to space constraints. Such a study was performed in [15] and showed that the improvement in image quality with the learning approach was independent of the used sampling scheme. To determine the model parameters of the network that will perform the parallel imaging reconstruction task, an optimization problem needs to be defined that minimizes a training objective. In general, this can be formulated in a supervised or unsupervised manner. Supervised approaches are predominantly used while unsupervised approaches are still a topic of ongoing research. Therefore, we will focus on supervised approaches for the remainder of this section. We define the number of stages, corresponding to gradient steps in the network, as  $S$ .  $d$  is the current training image out of the complete set of training data  $D$ . The variable  $\Theta$  contains all trainable parameters of the reconstruction model. The training objective then takes the following form

$$L(\Theta) = \min_{\Theta} \frac{1}{2D} \sum_{d=1}^D \|\mathbf{u}_d^S(\Theta) - \mathbf{u}_d^{\text{ref}}\|_2^2. \quad (13)$$

As it is common in deep learning, Equation (13) is a non-convex optimization problem that is solved with standard numerical optimizers like stochastic gradient descent. This requires the computation of the gradient of the training objective with respect to the model parameters  $\Theta$ . This gradient can be computed via back-propagation

$$\frac{\partial L(\Theta)}{\partial \Theta^s} = \frac{\partial \mathbf{u}^{s+1}}{\partial \Theta^s} \cdot \frac{\partial \mathbf{u}^{s+2}}{\partial \mathbf{u}^{s+1}} \cdots \frac{\partial \mathbf{u}^S}{\partial \mathbf{u}^{S-1}} \cdot \frac{\partial L(\Theta)}{\partial \mathbf{u}^S}. \quad (14)$$

<sup>2</sup><https://cai2r.net/resources/software>

<sup>3</sup><https://github.com/VLOGroup/mri-variationalnetwork>

The basis of supervised approaches is the availability of a target reference reconstruction  $\mathbf{u}^{\text{ref}}$ . This requires the availability of a fully-sampled set of raw phased array coil k-space data. This data is then retrospectively under-sampled by removing k-space data points as defined by the sampling trajectory in the forward operator  $\mathbf{E}$  and serves as the input of the reconstruction network. For training example  $d$ , the current output of the network  $\mathbf{u}_d^D(\theta)$  is then compared to the reference  $\mathbf{u}_d^{\text{ref}}$  via an error metric. The choice of this error metric has an influence on the properties of the trained network, which is a topic of currently ongoing work. A popular choice is the mean squared error (MSE), which was also used in Equation (13). Other choices are the  $\ell_1$  norm of the difference and the SSIM.

The focus of this section was machine learning based extensions of combined parallel imaging and compressed sensing, where the machine learning was mainly used to learn a model that serves as a more complex regularizer. Another set of developments in image space based parallel imaging is focused on the improvement of the estimation of the coil sensitivity maps via joint estimation of image content and coil sensitivities. The first in a recent set developments in that direction was proposed in [26]. This approach first performs the IDFT of the undersampled multi-channel k-space (see the illustration in the left column of Figure 1). The neural network is then trained to learn the mapping of the aliased individual coil images to the combined un-aliased image. The network thus learns how to use the sensitivity information to perform the de-aliasing without using explicit coil sensitivity maps. The authors used a classic fully connected multi-layer-perceptron for this task. The use of fully connected networks is usually challenging for clinically relevant image sizes due to memory requirements, for this particular application it was possible by performing de-aliasing separately for each 1D line in image space in the phase encoding direction. A more general version of this approach was recently proposed in [27]. The authors use a CNN, which eliminates the memory issue and allows them to use the proposed approach for 3D time-of-flight angiography.

## V. Machine learning methods for parallel imaging in k-space

There has been a recent interest in using neural network to improve the k-space interpolation techniques using non-linear approaches in a data-driven manner. These newer approaches can be divided into two groups based on how the interpolation functions are trained. The first group uses scan-specific ACS lines to train neural networks for interpolation, similar to existing interpolation approaches, such as GRAPPA or non-linear GRAPPA. The second group uses training databases, similar to the machine learning methods discussed in image domain parallel imaging.

Robust artificial-neural-networks for k-space interpolation (RAKI) is a scan-specific machine learning approach for improved k-space interpolation [28]. This approach trains CNNs on ACS data, and uses these for interpolating missing k-space points from acquired ones. The interpolation function can be represented by

$$\kappa_j(k_x, k_y - m\Delta k_y) = f_{j,m}(\{\kappa_c(k_x - b_x\Delta x, k_y - Rb_y\Delta y)\}_{b_x \in [-B_x, B_x], b_y \in [-B_y, B_y], c \in [1, n_c]}) \quad (15)$$

where  $f_{j,m}$  is the interpolation rule implemented via a multi-layer CNN for outputting the k-space of the  $m^{\text{th}}$  set of uniformly spaced missing lines in the  $j^{\text{th}}$  coil,  $R$  is the undersampling rate,  $B_x, B_y$  are parameters specified by the receptive field of the CNN, and  $n_c$  is the number of coils. Thus, the premise of RAKI is similar to GRAPPA, while the interpolation function is implemented using CNNs, whose parameters are learned from ACS data with an MSE loss function. The scan-specific nature of this method is attractive since it requires no training databases, and can be applied to scenarios where a fully-sampled gold reference cannot be acquired, for instance in perfusion or real-time cardiac MRI, or high-resolution brain imaging. Example RAKI, NL-GRAPPA and GRAPPA reconstructions for such high-resolution brain imaging datasets, which were acquired with prospective undersampling are shown in Figure 4. These data were acquired on a 7T system (Siemens Magnex Scientific) with 0.6 mm isotropic resolution.  $R = 5$ , 6 data were acquired with two averages for improved SNR to facilitate visualization of any residual artifacts. Other imaging parameters are available in [28]. For these datasets, RAKI leads to a reduction in noise amplification compared to GRAPPA. Note the noise reduction here is based on exploiting properties of the coil geometry, and not on assumptions about image structure, as in traditional regularized inverse problem approaches, as in Section II-B. However, the scan-specificity also comes with downsides, such as the computational burden of training for each scan, as well as the requirement for typically more calibration data. In this dataset, RAKI and NL-GRAPPA have similar performance for  $R = 4, 5$ . At  $R = 6$ , RAKI preserves sharper details compared to NL-GRAPPA, although the differences are subtle. In Figure 5, reconstructions of the knee dataset from Figure 3 are shown, where all methods, which rely on subject-specific calibration data, exhibit a degree of artifacts, due to the small size of the ACS region, while RAKI has the highest SSIM and lowest NRMSE.

While originally designed for uniform undersampling patterns, this method has been extended to arbitrary sampling, building on the self-consistency approach of SPIRiT [29]. Additionally, recent work has also reformulated this interpolation procedure as a residual CNN, with residual defined based on a GRAPPA interpolation kernel [30]. Thus, in this approach called residual RAKI (rRAKI), the CNN effectively learns to remove the noise amplification and artifacts associated with GRAPPA, giving a physical interpretation to the CNN output, which is similar to the use of residual networks in image denoising. An example application of the rRAKI approach in simultaneous multi-slice (SMS) imaging [31] with an SMS factor of 16, i.e. a 16-fold acceleration in coverage is shown in Figure 6, showing improvement over both GRAPPA and NL-GRAPPA in reducing residual aliasing and noise amplification. We note that for all experiments shown in this section, the same number of ACS lines were used for all methods. Thus, the differences between methods are not due to the size of the calibration data.

A different line of work, called DeepSPIRiT, explores using CNNs trained on large databases for k-space interpolation with a SPIRiT-type approach [32]. Since sensitivity profiles and number of coils vary for different anatomies and hardware configurations, k-

space data in the database were normalized using coil compression to yield the same number of channels [33]. Coil compression methods effectively capture most of the energy across coils in a few virtual channels, with the first virtual channel containing most of the energy, the second being the second most dominant, and so on, in a manner reminiscent of principal component analysis. Following this normalization of the k-space database, CNNs are trained for interpolating different regions of k-space. The method was shown to remove aliasing artifacts, though difficulty with high-resolution content was noted. Since DeepSPIRiT trains interpolation kernels on a database, it does not require calibration data for a given scan, potentially reducing acquisition time further.

Neural networks have also been applied to the Hankel matrix based approaches in k-space [34]. Specifically, the completion of the weighted k-space in ALOHA method has been replaced with a CNN, trained with an MSE loss function. The method was shown to not only improve the computational time, but also the reconstruction quality compared to original ALOHA by exploiting structures beyond low-rankness of Hankel matrices. In another line of work, neural networks have been applied to a Hankel matrix based approach that models signals as piecewise smooth [35]. These methods are described in more detail in another article in this issue [36].

## VI. Discussion

### A. Issues and open problems

Several advantages of neural network based machine learning approaches over classic constrained reconstruction using predefined regularizers have been proposed in the literature. First, the regularizer is tailored to a specific image reconstruction task, which improves the removal of residual artifacts. This becomes particularly relevant in situations where the used sampling trajectory does not fulfill the incoherence requirements of compressed sensing, which is often the case for clinical parallel imaging protocols. Second, machine learning approaches decouple the compute-heavy training step from a lean inference step. In medical image reconstruction, it is critical to have images of diagnostic quality available immediately after the scan so that technologists and radiologists can decide immediately whether a certain sequence needs to be repeated or acquisition parameters need to be changed. In contrast, prolonged training procedures that can be done on specialized computing hardware, are generally acceptable. For example, the training of the VN reconstruction in the experiment in Figure 3 took 40 hours for 150 epochs with 200 slices of training data on a single NVIDIA M40 GPU with 12GB of memory. Training data, model and training parameters exactly follow the training from [23]. Reconstruction of one slice then took 200ms, which is compatible with current image reconstruction times of algorithms like SENSE and GRAPPA on clinical scanners, and therefore compatible with clinical workflow. In comparison, the computation times were 10ms for zero filling, 150ms for CG-SENSE and 10000ms for the PI-CS TGV constrained reconstructions on the same GPU hardware.

The focus in Section IV and Section V were on methods that were developed specifically in the context of parallel imaging. Some architectures for image domain machine learning have been designed specifically towards a target application, for example dynamic imaging [37].

Another approach was recently developed that combines k-space and image space CNN data processing [38]. In their current form, these methods were not yet demonstrated in the context of multi-coil data. The approach recently proposed by Zhu et al. learns the complete mapping from k-space raw data to the reconstructed image [39]. The proposed advantage is that since no information about the acquisition is included in the forward operator  $\mathbf{E}$ , it is more robust against systematic calibration errors during the acquisition. This comes at the price of a significantly higher number of model parameters. The corresponding memory requirements make it challenging to use this model for matrix sizes that are currently used in clinical applications. A systematic comparison of these recent approaches from the literature is still an open question in the field. However, a fair comparison is challenging because their performance also depends on the quality of the training data. Most approaches are also designed with a particular set of target applications in mind, and different research groups usually build their own data sets as part of their developments. Thus, it can be a non-trivial task to modify a particular approach so that the performance is optimal for a new type of data. For example, in approaches that are designed either for dynamic or static imaging, the regularizer models are tailored to the specific properties of these data. We also note that there are fewer works in k-space machine learning methods for MRI reconstruction. This may be due to the different nature of k-space signal that usually has very different intensity characteristics in the center versus the outer k-space, which makes it difficult to generalize the plethora of techniques developed in computer vision and image processing that exploit properties of natural images.

Machine learning reconstruction approaches also come with a number of drawbacks when compared to classic constrained parallel imaging. First, they require the availability of a curated training data set that is representative so that the trained model generalizes to new unseen test data. While recent approaches from the literature [15, 25, 37] have either been trained with hundreds of examples rather than millions of examples as it is common in deep learning for computer vision, or trained on synthetic non-medical data that is publicly available from existing databases. However, this is still a challenge that will potentially limit the use of machine learning to certain applications. Several applications in imaging of moving organs, such as the heart, or in imaging of the brain connectivity, such as diffusion MRI, cannot be acquired with fully-sampled data due to constraints on spatio-temporal resolutions. This hinders the use of fully-sampled training labels for such datasets, highlighting applications for scan-specific approaches or unsupervised training strategies.

These reconstruction methods also require the availability of computing resources during the training stage. This is a less critical issue due to the increased availability and reduced prices of GPUs. The experiments in this paper were made with computing resources that are available for less than 10,000 USD, which are usually available in academic institutions. In addition, the availability of on-demand cloud-based machine learning solutions is constantly increasing.

A more severe issue is that in contrast to conventional parallel imaging and compressed sensing, machine learning models are mostly non-convex. Their properties, especially regarding their failure modes and generalization potential for daily clinical use, are understood less well than conventional iterative approaches based on convex optimization.



For example, it was recently shown that while reconstructions generalize well with respect to changes in image contrast between training and test data, they are susceptible towards systematic deviations in SNR [23]. It is also still an open question how specific trained models have to be. Is it sufficient to train a single model for all types of MR exams, or are separate models required for scans of different anatomical areas, pulse sequences, acquisition trajectories and acceleration factors as well as scanner manufacturers, field strengths and receive coils? While pre-training a large number of separate models for different exams would be feasible in clinical practice, if certain models do not generalize with respect to scan parameter settings that are usually tailored to the specific anatomy of an individual patient by the MR technologist, this will severely impact their translational potential and ultimately their clinical use. More generally speaking, an improved understanding of neural network training and architecture design to optimization theory is a very active research topic in the field of machine learning. We expect that future research in similar directions will further bridge the gap between current experimental results and the underlying theory and lead to a better understanding of generalization properties, failure modes in worst case scenarios and architecture design for specific types of problems.

## B. Availability of training databases and community challenges

As mentioned in the previous section, one open issue in the field of machine learning reconstruction for parallel imaging is the lack of publicly available databases of multi-channel raw k-space data. This restricts the number of researchers who can work in this field to those who are based at large academic medical centers where this data is available, and for the most part excludes the core machine learning community that has the necessary theoretical and algorithmic background to advance the field. In addition, since the used training data becomes an essential part of the performance of a certain model, it is currently almost impossible to compare new approaches that are proposed in the literature with each other if the training data is not shared when publishing the manuscript. While the momentum in initiatives for public releases of raw k-space data is growing [40], the number of available data sets is still on the order of hundreds and limited to very specific types of exams. Examples of publicly available rawdata sets are [mridata.org](https://mridata.org)<sup>5</sup> and the fastMRI dataset<sup>6</sup>.

## VII. Conclusion

Machine learning methods have recently been proposed to improve the reconstruction quality in parallel imaging MRI. These techniques include both image domain approaches for better image regularization and k-space approaches for better k-space completion. While the field is still in its development, there are many open problems and high-impact applications, which are likely to be of interest to the broader signal processing community.

## Acknowledgments

The authors thank Hemant Aggarwal and Mathews Jacob for providing the MODL reconstructions.

---

<sup>5</sup><https://mridata.org>

<sup>6</sup><https://fastmri.med.nyu.edu/>



This work was partially supported by NIH R01EB024532, NIH R00HL111410, NIH P41EB015894, NIH P41EB027061, NIH P41EB017183, NSF CAREER CCF-1651825

## Short Bio For Each Author

I.

• **Florian Knoll** received M.Sc. and Ph.D. degrees in electrical engineering in 2006 and 2011, respectively, both from Graz University of Technology, Graz, Austria. He is currently an Assistant Professor at the Center for Biomedical Imaging, New York University School of Medicine. His research is centered around computational methods for data acquisition and image reconstruction in medical imaging and currently holds an R01 award from NIH on this topic. In particular, he is working on iterative image MR, PET and CT image reconstruction, with special focus on parallel MR imaging, compressed sensing and machine learning. Email: florian.knoll@nyumc.org

• **Kerstin Hammernik** received her M.Sc. degree in Biomedical Engineering and her Ph.D. degree in Computer Science from Graz University of Technology, Graz, Austria in 2015 and 2019, respectively. Currently, she is a research associate in the group of Daniel Rueckert at the Department of Computing, Imperial College London, United Kingdom. During her Ph.D., she spent four months at New York University, School of Medicine, for an internship with a focus on clinical MR imaging, supervised by Daniel Sodickson and Florian Knoll. Her research interests include optimization and learning of variational models, medical image reconstruction and medical computer vision. In her current work, she focuses on machine learning for accelerated MR image reconstruction. Email: k.hammernik@imperial.ac.uk

• **Chi Zhang** received the B. Eng. degree in Electrical Engineering from Beijing University of Technology, Beijing, China, in 2016. He is currently a research student and a Ph.D. student under the supervision of Mehmet Akçakaya in the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN. His research interests include MR image reconstruction and machine learning. Email: zhan4906@umn.edu

• **Steen Moeller** received the Cand. Scient degree in Mathematics from Aalborg University, Aalborg, Denmark, in 1997, and the Ph.D. degree from Aalborg University, Aalborg, Denmark, in 2002. He was the recipient of the Danish Travel award for Mathematicians. He is currently an Assistant Professor in the Department of Radiology, University of Minnesota Minneapolis, MN, USA. His contributions include silent MRI with the SWIFT technology and the use of rapid scanning techniques that are used in the Human Connectome Project. His research centers on optimal encoding and computational methods for image reconstruction in MRI, with special focus on parallel MR imaging, and machine learning. Email: moeller@cmrr.umn.edu

• **Thomas Pock** received his M.Sc. (1998–2004) and his Ph.D. (2005–2008) in Computer Engineering (Telematik) from Graz University of Technology. After a Post-doc position at the University of Bonn, he moved back to Graz University of Technology where he has been an Assistant Professor at the Institute for Computer Graphics and Vision. In 2013 Thomas Pock received the START price of the Austrian Science Fund (FWF) and the German Pattern

recognition award of the German association for pattern recognition (DAGM) and in 2014, Thomas Pock received a starting grant from the European Research Council (ERC). Since June 2014, Thomas Pock is a Professor of Computer Science at Graz University of Technology and a principal scientist at the Center for Vision, Automation & Control at the Austrian Institute of Technology (AIT). The focus of his research is the development of mathematical models for computer vision and image processing, machine learning, and the development of efficient non-smooth optimization algorithms. Email: pock@icg.tugraz.at

• **Daniel K. Sodickson** received a B.S. in Physics and a B.A. in Humanities from Yale College in 1988, followed by a Ph.D. in Medical Physics from the Massachusetts Institute of Technology in 1994, and an M.D. from Harvard Medical School in 1996. He currently serves as Vice-Chair for Research in the Department of Radiology, Director of the Bernard and Irene Schwartz Center for Biomedical Imaging, and Principal Investigator of the Center for Advanced Imaging Innovation and Research at New York University School of Medicine. Dr. Sodickson is credited with founding the field of parallel magnetic resonance imaging, in which distributed arrays of detectors are used to gather magnetic resonance images at previously inaccessible speeds. For this work, he received the Gold Medal of the International Society for Magnetic Resonance in Medicine (ISMRM) in 2006. He served as President of the ISMRM in 2017–2018, and Chair of the NIH Study Section on Biomedical Imaging Technology from 2016–2018, and he is a Distinguished Investigator of the Academy for Radiology and Biomedical Imaging Research. Dr. Sodickson leads a multidisciplinary team that aims to develop a new paradigm of rapid continuous comprehensive imaging, taking advantage of complementary tools in image acquisition and reconstruction, including parallel imaging, compressed sensing, and artificial intelligence. Email: daniel.sodickson@nyumc.org

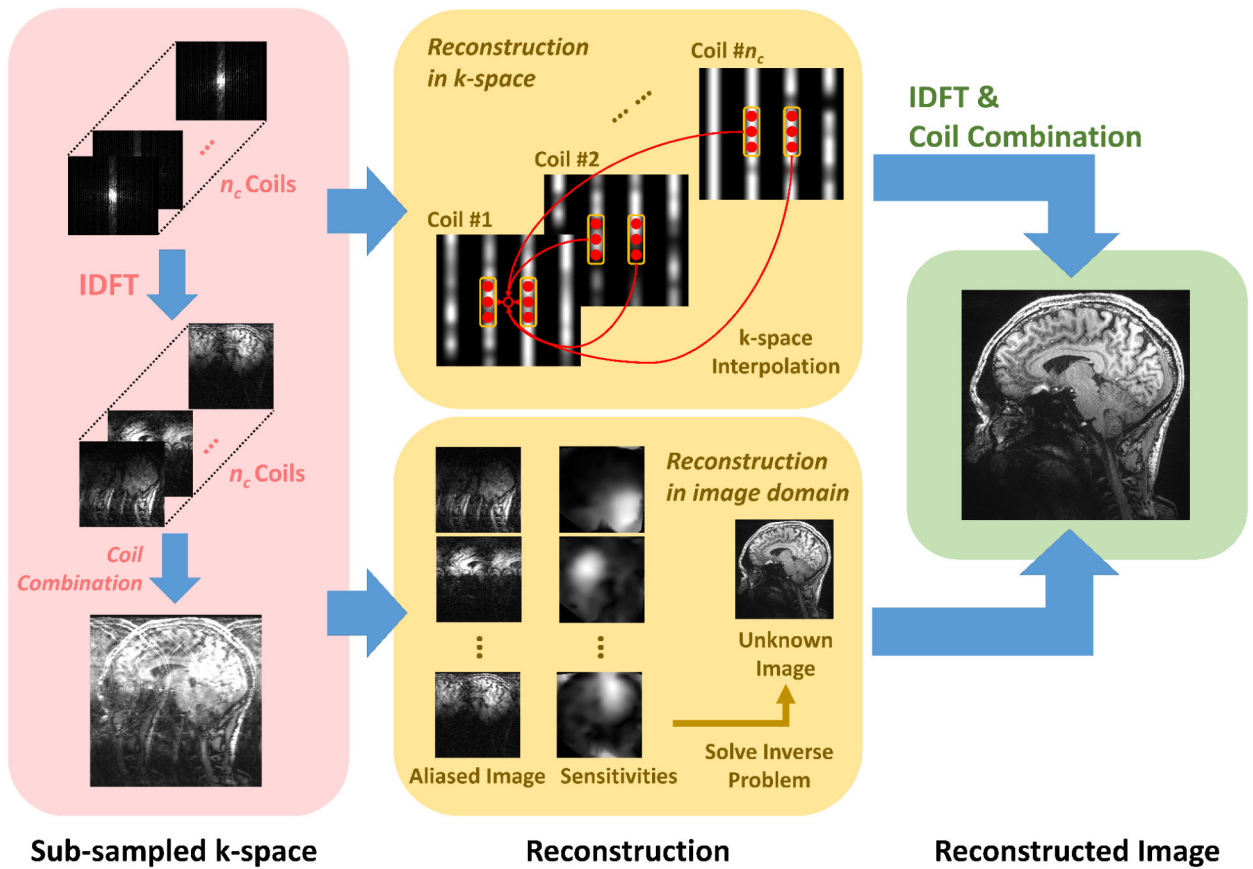
• **Mehmet Akçakaya** received the B. Eng. degree from McGill University, Montreal, QC, Canada, in 2005, and the S.M. and Ph.D. degrees from Harvard University, Cambridge, MA, USA, in 2010. From 2010 to 2015, he was with the Harvard Medical School. He is currently a McKnight-Land Grant Assistant Professor at the University of Minnesota, Minneapolis, MN, USA. His work on accelerated MRI has received a number of international recognitions. He holds an R00 Award from NIH and a CAREER Award from NSF. His research interests include image reconstruction, cardiac MRI, machine learning, inverse problems and signal processing. Email: akcakaya@umn.edu

## References

- [1]. Sodickson DK and Manning WJ, “Simultaneous Acquisition of Spatial Harmonics (SMASH): Fast Imaging with Radiofrequency Coil Arrays,” *Magn Reson Med*, vol. 38, no. 4, pp. 591–603, 1997. [PubMed: 9324327]
- [2]. Pruessmann KP, Weiger M, Scheidegger MB, and Boesiger P, “SENSE: Sensitivity encoding for fast MRI,” *Magn Reson Med*, vol. 42, no. 5, pp. 952–962, 1999. [PubMed: 10542355]
- [3]. Griswold MA, Jakob PM, et al., “Generalized autocalibrating partially parallel acquisitions (GRAPPA),” *Magn Reson Med*, vol. 47, no. 6, pp. 1202–1210, 6 2002. [PubMed: 12111967]
- [4]. LeCun Y, Bengio Y, and Hinton G, “Deep Learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [PubMed: 26017442]

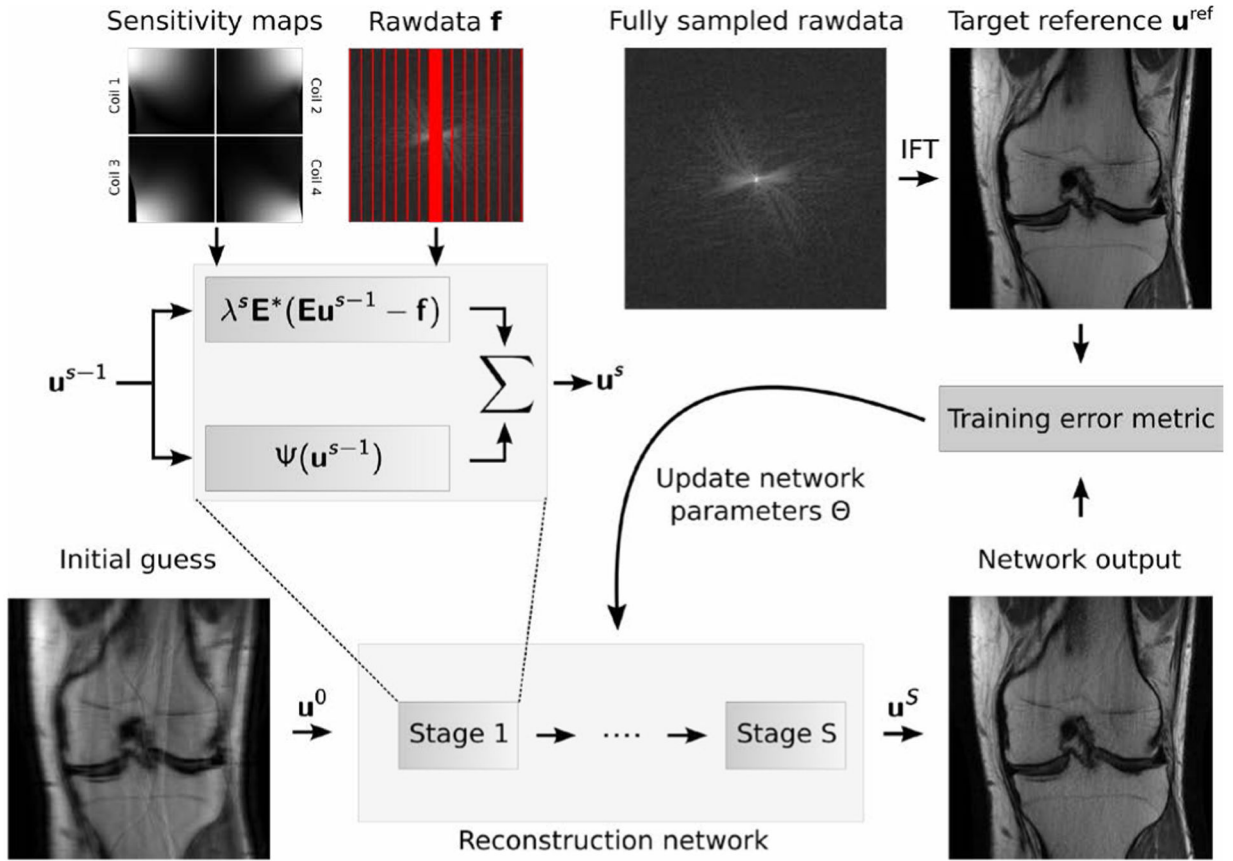
- [5]. Uecker M, Lai P, et al., “ESPIRiT – An Eigen-value Approach to Autocalibrating Parallel MRI: Where SENSE meets GRAPPA,” *Magn Reson Med*, vol. 71, no. 3, pp. 990–1001, 2014. [PubMed: 23649942]
- [6]. Pruessmann KP, Weiger M, Boernert P, and Boesiger P, “Advances in sensitivity encoding with arbitrary k-space trajectories,” *Magn Reson Med*, vol. 46, no. 4, pp. 638–651, 2001. [PubMed: 11590639]
- [7]. Lustig M, Donoho D, and Pauly JM, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” *Magn Reson Med*, vol. 58, no. 6, pp. 1182–1195, 2007. [PubMed: 17969013]
- [8]. Block KT, Uecker M, and Frahm J, “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. [PubMed: 17534903]
- [9]. Lustig M and Pauly JM, “SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space,” *Magn Reson Med*, vol. 64, no. 2, pp. 457–471, 8 2010. [PubMed: 20665790]
- [10]. Knoll F, Bredies K, Pock T, and Stollberger R, “Second order total generalized variation (TGV) for MRI,” *Magn Reson Med*, vol. 65, no. 2, pp. 480–491, 2011. [PubMed: 21264937]
- [11]. Knoll F, Clason C, Bredies K, Uecker M, and Stollberger R, “Parallel imaging with nonlinear reconstruction using variational penalties,” *Magn Reson Med*, vol. 67, no. 1, pp. 34–41, 2012. [PubMed: 21710612]
- [12]. Akcakaya M, Basha TA, et al., “Low-dimensional-structure self-learning and thresholding: Regularization beyond compressed sensing for MRI reconstruction,” *Magn Reson Med*, vol. 66, no. 3, pp. 756–767, 9 2011. [PubMed: 21465542]
- [13]. Akcakaya M, Basha TA, Chan RH, Manning WJ, and Nezafat R, “Accelerated isotropic sub-millimeter whole-heart coronary MRI: compressed sensing versus parallel imaging,” *Magn Reson Med*, vol. 71, no. 2, pp. 815–822, 2 2014. [PubMed: 23440946]
- [14]. Jung H, Sung K, Nayak KS, Kim EY, and Ye JC, “k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI,” *Magn Reson Med*, vol. 61, no. 1, pp. 103–116, 2009. [PubMed: 19097216]
- [15]. Hammernik K, Klatzer T, et al., “Learning a variational network for reconstruction of accelerated MRI data,” *Magn Reson Med*, vol. 79, pp. 3055–71, 2018.
- [16]. Kholmovski EG and Parker DL, “Spatially variant GRAPPA,” in *Proc. 14th Scientific Meeting and Exhibition of ISMRM, Seattle, 2006*, p. 285.
- [17]. Robson PM, Grant AK, et al., “Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions,” *Magn Reson Med*, vol. 60, no. 4, pp. 895–907, 10 2008. [PubMed: 18816810]
- [18]. Chang Y, Liang D, and Ying L, “Nonlinear GRAPPA: A kernel approach to parallel MRI reconstruction,” *Magn Reson Med*, vol. 68, no. 3, pp. 730–740, 9 2012. [PubMed: 22161975]
- [19]. Weller DS, Polimeni JR, et al., “Sparsity-promoting calibration for GRAPPA accelerated parallel MRI reconstruction,” *IEEE Trans Med Imaging*, vol. 32, no. 7, pp. 1325–1335, 7 2013. [PubMed: 23584259]
- [20]. Shin PJ, Larson PE, et al., “Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion,” *Magn Reson Med*, vol. 72, no. 4, pp. 959–970, 10 2014. [PubMed: 24248734]
- [21]. Haldar JP and Zhuo J, “P-LORAKS: Low-rank modeling of local k-space neighborhoods with parallel imaging data,” *Magn Reson Med*, vol. 75, no. 4, pp. 1499–1514, 4 2016. [PubMed: 25952136]
- [22]. Jin KH, Lee D, and Ye JC, “A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank hankel matrix,” *IEEE Trans Comp Imaging*, vol. 2, no. 4, pp. 480–495, 12 2016.
- [23]. Knoll F, Hammernik K, et al., “Assessment of the generalization of learned image reconstruction and the potential for transfer learning,” *Magn Reson Med*, 2019.
- [24]. Ronneberger O, Fischer P, and Brox T, “UNet: Convolutional Networks for Biomedical Image Segmentation,” *Miccai*, pp. 234–241, 2015.

- [25]. Aggarwal HK, Mani MP, and Jacob M, "MoDL: Model-Based Deep Learning Architecture for Inverse Problems," *IEEE Trans Med Imaging*, 2019.
- [26]. Kwon K, Kim D, and Park H, "A Parallel MR Imaging Method using Multilayer Perceptron," *Medical Physics*, 2017.
- [27]. Jun Y, Eo T, et al., "Parallel imaging in time-offlight magnetic resonance angiography using deep multistream convolutional neural networks," *Magn Reson Med*, 2019.
- [28]. Akcakaya M, Moeller S, Weingartner S, and Ugurbil K, "Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging," *Magn Reson Med*, vol. 81, no. 1, pp. 439–453, 1 2019. [PubMed: 30277269]
- [29]. Hosseini SAH, Moeller S, Weingärtner S, Ugurbil K, and Akçakaya M, "Accelerated coronary MRI using 3D SPIRiT-RAKI with sparsity regularization," in *Proc. IEEE International Symposium on Biomedical Imaging*, 2019.
- [30]. Zhang C, Moeller S, Weingärtner S, Ugurbil K, and Akçakaya M, "Accelerated MRI using residual RAKI: Scan-specific learning of reconstruction artifacts," in *Proc. 27th Annual Meeting of the ISMRM, Montreal, Canada*, 2019.
- [31]. Zhang C, Moeller S, Weingartner S, Ugurbil K, and Akçakaya M, "Accelerated simultaneous multi-slice MRI using subject-specific convolutional neural networks," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, 10 2018, pp. 1636–1640.
- [32]. Cheng JY, Mardani M, Alley MT, Pauly JM, and Vasanawala SS, "DeepSPIRiT: Generalized parallel imaging using deep convolutional neural networks," in *Proc. 26th Annual Meeting of the ISMRM, Paris, France*, 2018.
- [33]. Buehrer M, Pruessmann KP, Boesiger P, and Kozerke S, "Array compression for MRI with large coil arrays," *Magn Reson Med*, vol. 57, no. 6, pp. 1131–1139, 6 2007. [PubMed: 17534913]
- [34]. Han Y and Ye JC, "k-space deep learning for accelerated MRI," *IEEE Trans Med Imaging*, in press, 2019.
- [35]. Pramanik A, Aggarwal H, Mani MP, and Jacob M, "Off-the-grid model-based deep learning (OMODL)," in *Proc. IEEE International Symposium on Biomedical Imaging*, 2019.
- [36]. Jacob M, Mani MP, and Ye JC, "Structured low-rank algorithms: Theory, MR applications, and links to machine learning," *IEEE Signal Processing Magazine*, 2020.
- [37]. Schlemper J, Caballero J, Hajnal JV, Price AN, and Rueckert D, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans Med Imaging*, vol. 37, no. 2, pp. 491–503, 2018. [PubMed: 29035212]
- [38]. Eo T, Jun Y, et al., "KIKI-net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images," *Magn Reson Med*, vol. 80, pp. 2188–2201, 2018. [PubMed: 29624729]
- [39]. Zhu B, Liu JZ, Cauley SF, Rosen BR, and Rosen MS, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018. [PubMed: 29565357]
- [40]. Knoll F, Zbontar J, et al., "fastMRI: a publicly available raw k-space and DICOM dataset for accelerated MR image reconstruction using machine learning," *Radiology AI*, in press, 2019.



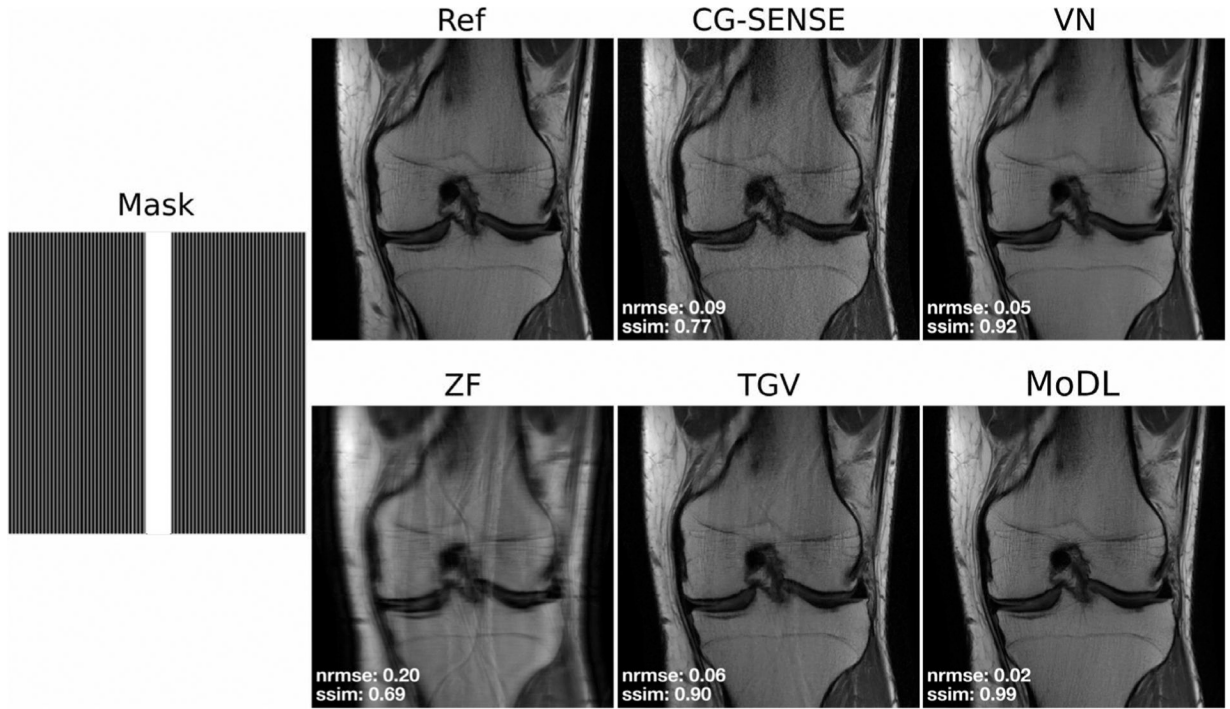
**Fig. 1:**

In k-space based parallel imaging methods, missing data is recovered first in k-space, followed by an inverse discrete Fourier transform (IDFT) and combination of the individual coil elements. In image space based parallel imaging, the IDFT is performed as the first step, followed by coil sensitivity based removal of the aliasing artifacts from the reconstructed image by solving an inverse problem.



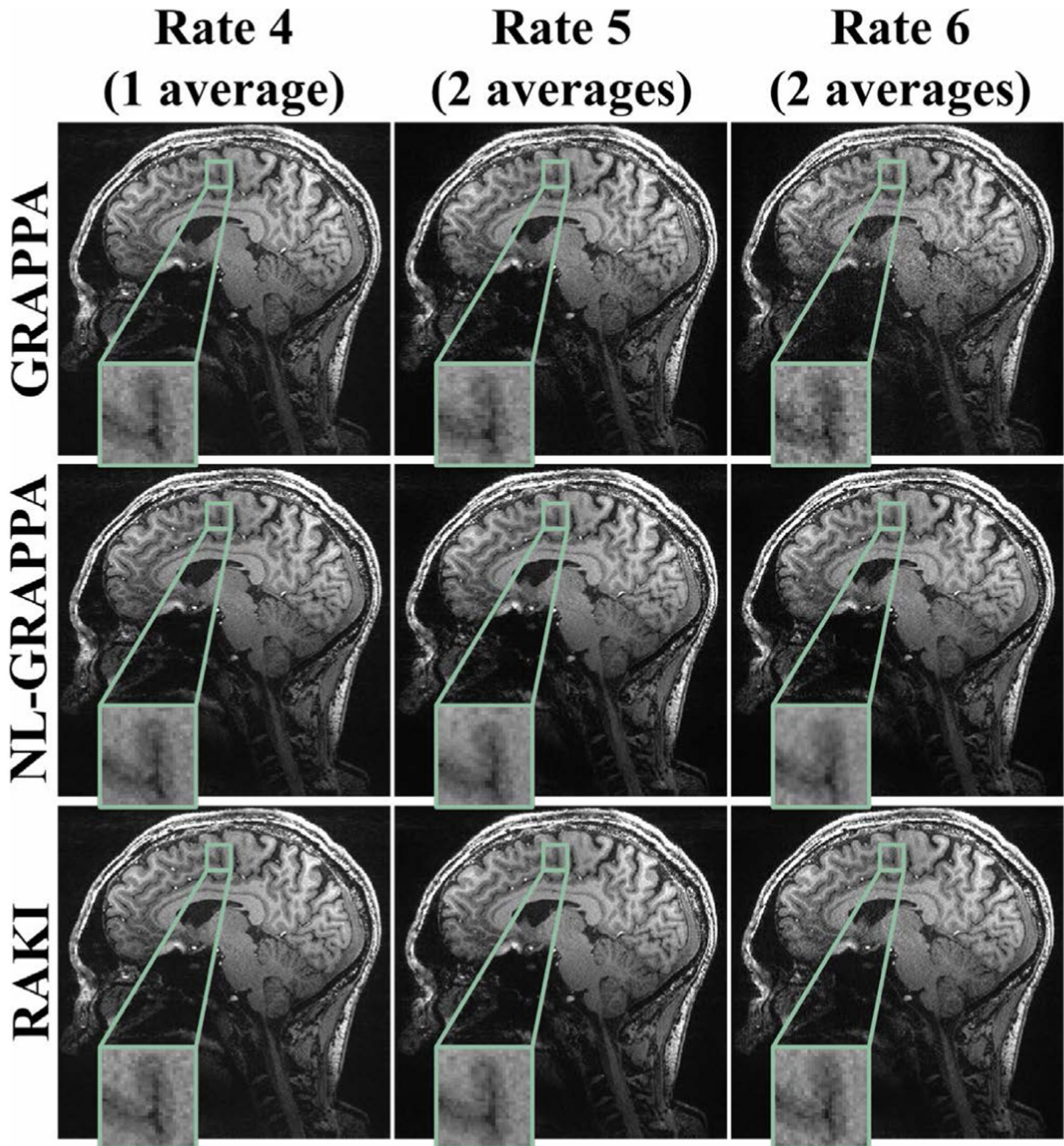
**Fig. 2:** Illustration of machine learning-based image reconstruction. The network architecture consists of  $S$  stages that perform the equivalent of gradient descent steps in a classic iterative algorithm. Each stage consists of a regularizer and a data consistency layer. Training the network parameters  $\Theta$  is performed by retrospectively undersampling fully sampled multi-coil raw k-space data and comparing the output of the network  $\mathbf{u}_d^S(\Theta)$  to a target reference reconstruction  $\mathbf{u}_d^{\text{ref}}$  obtained from the fully sampled data for the current training example  $d$ .





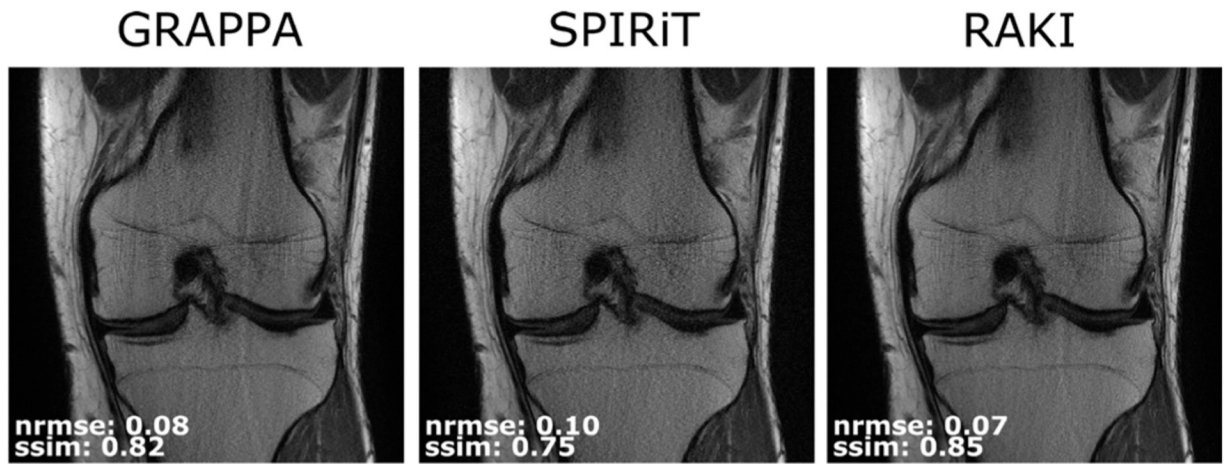
**Fig. 3:** Comparison of image-domain based parallel imaging reconstructions of a retrospectively accelerated coronal knee acquisition. The used sampling pattern, zero-filling, CG-SENSE, combined parallel imaging and compressed sensing with a TGV-constraint, and learned reconstructions using the Variational Network and MoDL architectures are shown, along with their NRMSE and SSIM values to the fully sampled reference. See text in the respective sections for details on the individual experiments.





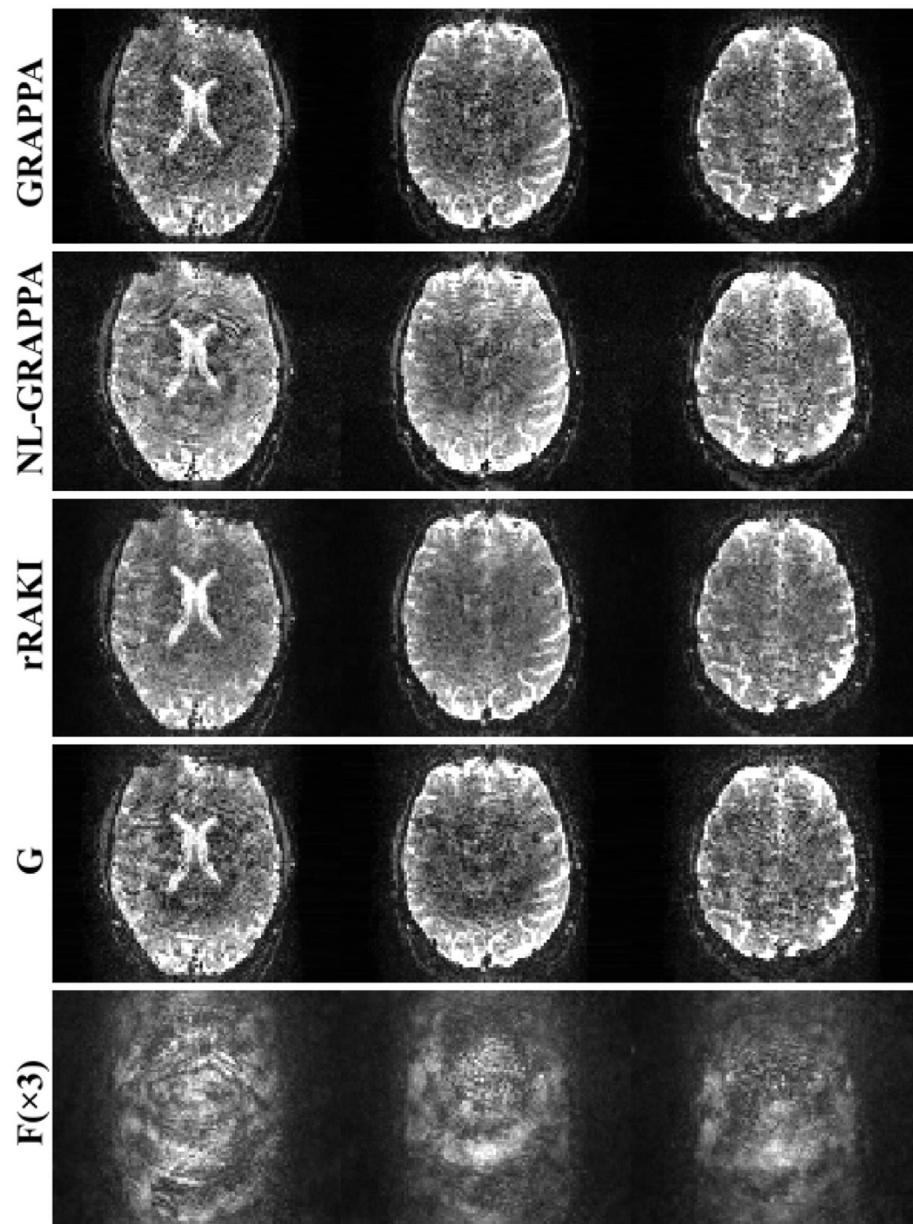
**Fig. 4:**

A slice from a high-resolution (0.6 mm isotropic) 7T brain acquisition, where all acquisitions were performed with prospective acceleration. It is difficult to acquire fully-sampled reference datasets for training for such acquisitions, thus two scan-specific k-space methods were compared. The CNN-based RAKI method visibly reduced noise amplification compared to the linear GRAPPA reconstruction. NL-GRAPPA and RAKI have similar noise properties, while RAKI produces a slightly sharper image at  $R = 6$ .



**Fig. 5:**

Comparison of k-space parallel imaging reconstructions of a retrospectively accelerated coronal knee acquisition, as in Figure 3. Due to the small size of the ACS data relative to the acceleration rate, the methods, none of which utilizes training databases, exhibit artifacts. GRAPPA has residual aliasing, whereas SPIRiT shows noise amplification. These are reduced in RAKI, though the residual artifacts remain. Respective NRMSE and SSIM values reflect these visual assessment.



**Fig. 6:** Reconstruction results of simultaneous multi-slice imaging of 16 slices in fMRI (i.e. 16-fold acceleration in coverage), where a sample of 3 slices are shown. GRAPPA method exhibits noise amplification at this high acceleration rate. NL-GRAPPA reduces noise amplification but suffers from residual aliasing and leakage. The rRAKI method, which consists of a linear convolutional component  $G$ , in parallel with a non-linear CNN component  $F$  that learns the artifacts arising from  $G$ , exhibits reduced noise and reduced aliasing. Due to imperfections in the ACS data for this application, the residual component includes both noise amplification and residual artifacts.